Chang, M. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

Predicting the level of linguistic knowledge from appropriately chosen learning data: A pilot study of English prepositional acquisition for Japanese EFL learners

Yuichi ONO[a*]
[a]*Faculty of Humanities and Social Sciences, University of Tsukuba, Japan*
*ono.yuichi.ga@u.tsukuba.ac.jp

Abstract: In institutions like foreign language education center, it is highly possible, given a proper online learning environment, or Computer-Assisted Language Learning (CALL) environment, that daily learning log data will be stored automatically at an institutional level. However, predicting learners' level of overall linguistic knowledge, or performance proficiency level, is a real challenge, due to the difficulty to set up an appropriate predicting model under multiple complex factors affecting linguistic knowledge or learner's performance. Especially, it is very difficult to predict proficiency level from regular tasks in the classroom or online learning at home under the context of flipped-classroom model. In this paper, I attempt to demonstrate that the prepositional knowledge can lead to the prediction of overall linguistic knowledge including actors affecting linguistic knowledge or learner's performance. Especially, it is very diff grammatical knowledge, reading comprehension, structures, and so on. This study conducted a survey of testing Japanese EFL learner's knowledge of English prepositions by asking 80 questions on *in, on, over, above, under,* and *below*. The analysis involves correlation analysis and implementation of Random Forest Algorithm to detect the key conceptual constructs to divide proficient-less proficient learners. The result is that a strong correlation between overall linguistic knowledge and prepositional knowledge that we collect during the classroom, and the acquisition of "special" or "metaphorical" concepts accompanied with English prepositions seems to be the key to predict overall knowledge of prepositions. Lastly, this paper concludes that this finding is interesting because it provides promising implications for collaborative data-driven or knowledge-driven research between learning analytics and theoretical linguistics, especially, the field of second language acquisition.

Keywords: Foreign Language Center, Micro-level Learning Analytics, Decision Tree, Japanese EFL Learners

1. Introduction

The increasing amount of data generated in digital learning contexts provides opportunities to benefit from learning analytics. As is frequently stated, even in the call for paper to this workshop, new methodologies and technological tools are necessary to analyze and make sense of these data and provide personalized scaffolding and services to stakeholders including students, faculty/teachers and administrators. The curriculum or everyday syllabus must be properly incorporated on the basis of newly-devised methodology which is connected with the needs of institutions with specific purposes. In the case of foreign language education center, instructors and researchers have been familiar with the use of media or technology to improve the instructional design since the age of structural linguistics or behavioral psychology paradigm (Ono & Ishihara, 2012; Rüschoff & Ritter, 2001; Warschauer & Healey, 1998). In the recent paradigms of communicative approach and Computer-Assisted Language Learning (CALL), an increasing number of students are learning with mobile devices at any time even outside the classroom. In addition, instructors are able to collect every log data from the students under the online learning environment. It looks as if language instructors and researchers were ready for learning analytics to provide prediction, clustering, and personalization to improve the quality of the foreign language courses. However, although we have an amazing number of the techniques for

analyzing big data, the use of datamining in education, particularly in language learning, has only recently emerged (Mark, Soobin, Hansol & Bindin, 2019).

Firstly, linguistics knowledge is a purely abstract concept and is not measurable at a deep level. Conceptually, we may posit some constructs which affect proficiency for better prediction, but most of the constructs are not easily measurable due to their invisibility. Secondly, the concept of linguistic proficiency is not operationalized easily in analysis, since language acquisition is associated with linguistic competence, as well as superficial memory such as memorization of words and phrases. For example, acquisition of prepositions involves understanding its core meaning and its degree of extension to peripheral meanings (Tyler & Evans, 2003). The acquisition of these abstract ideas does not originate from human experience to hear and learn prepositions, but possibly from the more universal competence that might be installed unconsciously in the human brain as a linguistic knowledge. Whether or not a student does understand and operationalize this knowledge is highly crucial when they are/are not able to understand and use the prepositions correctly. It is generally assumed that the learners with such knowledge will do better in other linguistic performances in writing or speaking.

Under the assumption that the accurate knowledge of preposition at a deeper level is highly related with the overall linguistic knowledge like reading, vocabulary, grammar and structures, this paper makes a pilot attempt to demonstrate the validity of this assumptions by using the data collected in regular classroom online tasks that are conducted in the classroom.

2. Previous Studies

2.1 Big Data and Little Data in Learning Analytics

Ono (2018) claims that, in order to avoid the so-called "click-to-construct issue" in learning analytics, we need to pick up the "right" data, instead of "big" data, especially in the case of language learning issues, since a lot of factors are not sometimes reflected as the frequency of log data, citing the statement by Borgman (2014):

> "Big Data" offers today's scholars vast opportunities for discovery and insight, but having the right data is often better than having more data. (p. 1)

Ono (2018) further suggests that the page-flipping might be the key to predict overall reading comprehension, among other indices often suggested in the learning analytics literature.

In the current research, main focus is placed on the acquisition English prepositions of *in, on, over, above, under,* and *below*. It is needless to say that these prepositions involve several uses and meanings, making it difficult for Japanese EFL learners to learn and use. We set up multiple-choice questions of diverse uses for each preposition to explore an acquisition order model for Japanese EFL learners.

2.2 Cognitive Linguistics and Instruction in the Classroom

A lot of studies of prepositional acquisition assume that the so-called "unidirectionality hypothesis" holds for Japanese EFL learners as to the acquisition order of English prepositions. This hypothesis originates from cognitive linguistics and states that the direction of semantic extensions is from "Core" meaning, called "Prototypical meaning", "Temporal meaning" and to "Abstract meaning". The metaphorical extension in meaning from Prototypical to Abstract meanings is described as a semantic network. These image-based instructions are very popular in prepositional instructions in the classroom. The example of "Core image" of *over* is given below in Figure 1.

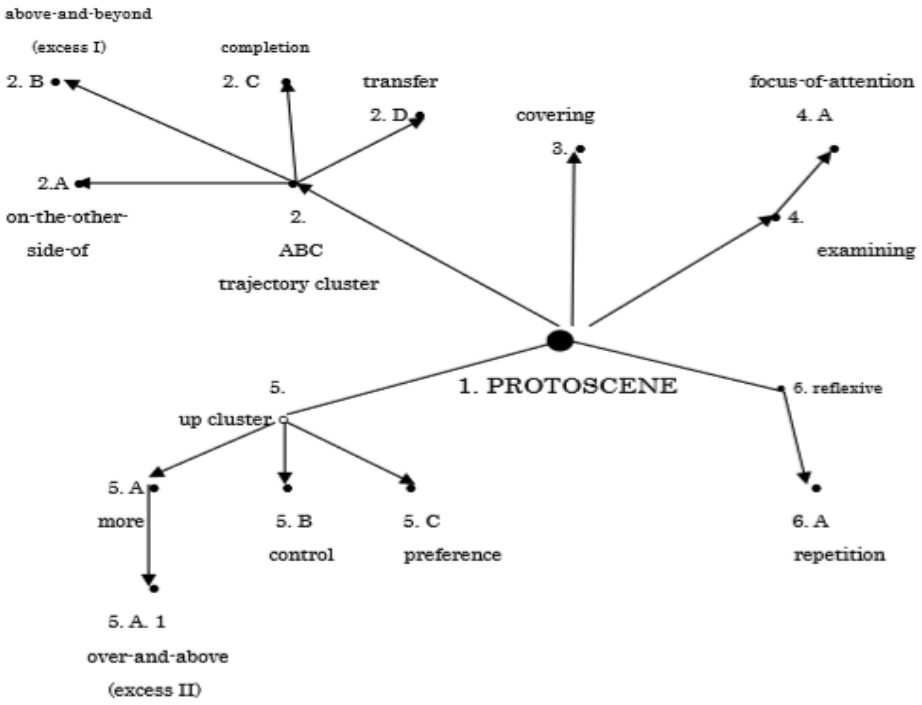

Figure 1. Semantic network of *over* (Tyler and Evans, 2003)

On the basis of the above network model, examples involving *over* such as the following are classified as in Table 1 below:

Table 1
Examples of *Over* and Their Types according to Tyler and Evans (2003)

Example	Type	Notes
He looked at himself in the mirror *over* the table.	Proto	
He was wearing a light-grey suit *over* the shirt.	Spatial	Covering
Dave, a pianist, played it *over* a couple of times.	Time	Repetition
T I'm glad that you're *over* the flu.	Abstract	Completion
He's never had any influence *over* her.	Abstract	Control

As to the order of acquisition, Cho (2002) suggests that learners acquire prepositions of prototypical usage first, and the acquisition order is Spatial usage, Temporal usage and Abstract usage. On the basis of this unidirectionality hypothesis, Japanese SLA research focuses on the benefits of "Image-use" instructions, instead of traditional translation-based ones.

However, Kano (2018) challenges this assumption and reviews the results obtained from previous studies, and suggests that more study is required to investigate what kind of knowledge is in the foreign language learner's brain and whether the student really makes use of "image" to understand and produce the correct prepositional use. The research shows that in some cases the unidirectionality hypothesis does not hold and his qualitative analysis demonstrated that irregularity of acquisition order is observed.

In order to solve the issue of whether the unidirectionality hypothesis holds or not, data-based validation is necessary on the basis of foreign language learner's knowledge of English prepositions, which is obtained online in the regular classroom tasks. Thus, the research questions of the current research are set up as follows:

RQ1: What type of specific prepositional knowledge predicts learner's knowledge of English preposition.
RQ2: What type of specific prepositional knowledge predicts learner's overall proficiency level?

3. Methodology

3.1 Participants

A total of 88 national university in Japan participated in this study. Their CEFR level is A2–C1.

Table 2
Participants' Overall Proficiency Level

CEFR Level	*Number of Participants*
C1	*1*
B2	*28*
B1	*30*
A2	*27*

3.2 Dataset

We constructed the dataset by collecting online quizzes that are held regularly in the classroom. The total number of questions is 80. All the questions are multiple-choice questions, where the participants are required to answer the best one among four choices. All the questions are supported by Japanese translations in order to make sure that all the students understand the situation described by the question.

3.3 Random Forest Algorithm

The analysis was conducted by using the statistical programming language R (R Core Team, 2019). We constructed a predictive model where 80 questions are treated as independent variables and TOEFL-ITP score and total score of preposition test as dependent variables. In learning, the number of trees was set to 500.

4. Result and Discussion

Table 3 shows a descriptive statistics of the test scores and TOEFL-ITP scores.

Table 3
Descriptive Statistics

	N	*Mean*	*SD*
Total Score of Prep Test	*86*	*48.4*	*11.4*
Score of TOEFL-ITP	*86*	*507.2*	*50.4*

Note. The maximum score of preposition test is 80.

The correlation between total score of Preposition Test and TOEFL-ITP score is r = .696, which is interpreted as a strong correlation. (95% CI [lower, upper] = 0.568 0.791)

Then, the result of Decision Tree analyses is given in Figures 2 and 3 below. Figure 2 is for TOEFL as a dependent variable, and Figure 3 is for total score for Preposition Test as a dependent variable.

In Figure 2, the total variance explained is 51.44% for this model.

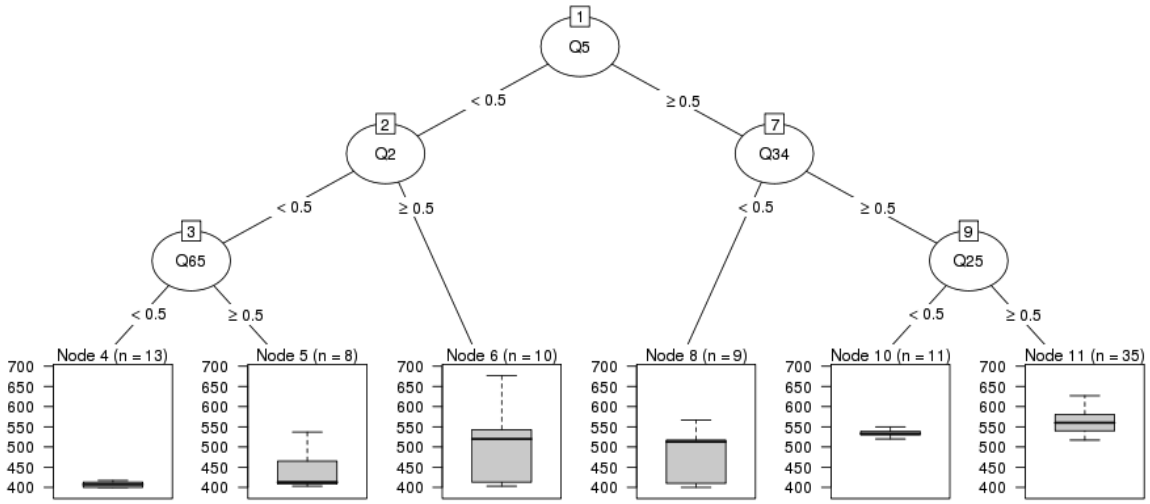

Figure 2. Decision tree and distribution of participants in each node (TOEFL-ITP)

The questions for each node and its value of significance (in this paper, "IncNodePurity" is employed) is described in Table 4 below.

Table 4
Questions and IncNodePurity for TOEFL-ITP

	Question	*Usage*	*IncNodePurity*
Q5	*Several conclusions could be drawn from the results described ( ). [over, up, to, <u>above</u>]*	*Prototypical*	*8.516806e+04*
Q2	*There was a little food left ( ) from the party. [on, in, to, <u>over</u>]*	*Abstract*	*2.432890e+04*
Q34	*She has never got ( ) the shock of her mother's death. [<u>over</u>, from, into, to]*	*Abstract*	*2.446524e+04*
Q65	*I don't want to talk about it ( ) the telephone. [<u>over</u>, in, above, under]*	*Abstract*	*2.175742e+03*
Q25	*I am the new manager and you will be working ( ) me. [below, at, in, <u>under</u>]*	*Abstract*	*3.613272e+03*

Note. The choices are given in square brackets, where the correct answer is underlined.

Now turn to Figure 3 for Preposition Test as a dependent variable. The total variance explained is 80.99% for this model.

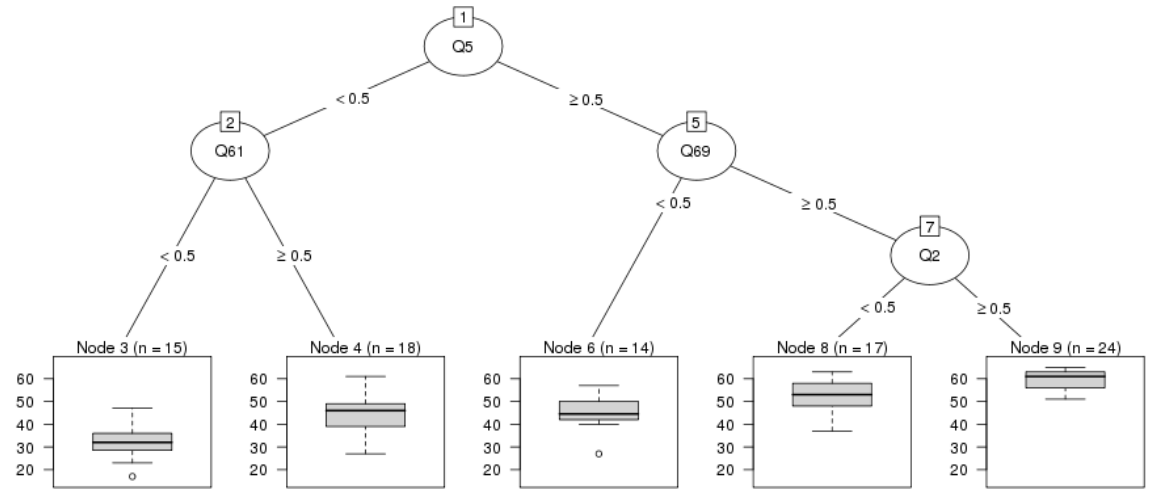

Figure2. Decision tree and distribution of participants in each node (Preposition Test)

The questions for each node and its importance (IncNodePurity) is described in Table 5 below.

Table 5

Questions and IncNodePurity for Preposition Test

	Question	*Usage*	*IncNodePurity*
Q5	*Several conclusions could be drawn from the results described ( ). [over, up, to, <u>above</u>]*	*Spatial*	*1374.1549910*
Q61	*Such people often experience less stress than those in the rank ( ) them. [<u>below</u>, on, around, under]*	*Spatial*	*105.2607250*
Q69	*I was wearing two sweaters ( ) the green jacket. [below, in, by, <u>under</u>]*	*Spatial*	*206.9395434*
Q2	*There was a little food left ( ) from the party. [on, in, to, <u>over</u>]*	*Abstract*	*144.2854017*

Note. The choices are given in square brackets, where the correct answer is underlined.

5. Discussion and Conclusion

From the result above, Q5 is a key question to divide upper from lower in both cases. This is a question on "Spatial expression" using above. The implication here is that the understanding and using Spatial Expressions seems to be an important and basic approach. However, the top group in Figure 1 seem to understand a "Metaphorical Expression" of *under*, which is a very difficult kind of expression. As to the knowledge of preposition shown in Table 2, understanding and using the typical "Spatial" uses seems to be the key. So far, it is safe to say that focus on Prototypical (or Spatial) use should be stressed in the instruction to beginners, which seems to be the key to divide whether the student gets upper or lower.

However, on the other hand, all the spatial expressions among 80 questions behave similarly; that is, some spatial expressions seem to be difficult to answer for some reasons. It is thus necessary to investigate what is happening in the students learning process in more details to explore a more decisive concept to explain the acquisition order of prepositions or the validity of unidirectionality hypothesis. It is needless to say that the ideas from linguistics or second language acquisition is also necessary for future learning analytics research.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Number 19K00903.

References

Cho, K. (2002). A cognitive linguistic approach to the acquisition of English prepositions. *JACET Bulletin, 35*, 63-78.

Kano, T. (2018). Issues on teaching English preposition to EFL learners: Focusing on the image schema of English preposition. In Y. Ono & M. Shimada (Eds.) *Data Science in Collaboration, Volume* 2. Tsukuba: General Affairs Supporting Center.

Mark W, Soobin, Y., Hansol, L., & Binbin, Z. (2019). Recent Contributions of Data Mining to Language Learning. *Research Annual Review of Applied Linguistics, 39*, 93–112.

Rüschoff, B., & Ritter, M. (2001). Technology-Enhanced Language Learning: Construction of Knowledge and Template-Based Learning in the Foreign Language Classroom. *Computer Assisted Language Learning, 14*(3) 219–232.

Ono, Y., & Ishihara, M. (2012). Integrating mobile-based individual activities into the Japanese EFL classroom. *International Journal of Mobile Learning and Organisation*. *6*. 116-137.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning and cognition.* Cambridge: Cambridge University Press.

Warschauer, M., & Healey, D. (1998). Computers and language learning: an overview. *Language Teaching, 31*, 57–71.