Knowledge tracing and cue contrast: Second language English grammar instruction

Helen Zhao (helenz@cuhk.edu.hk)
Department of English, The Chinese University of Hong Kong
Shatin, Hong Kong

Kenneth R. Koedinger (koedinger@cmu.edu)
Human-Computer Interaction Institute and Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

John Kowalski (jkau@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract

This paper introduces a cognitive tutor designed for second language grammar instruction. The tutor adopted Corbett and Anderson's (1995) Bayesian knowledge tracing model and provided adaptive training on the English article system. We followed the Competition Model (MacWhinney, 1997) and understood the article system as a galaxy of cues determining article usage on the basis of form-function mapping. Cues are in competition during language acquisition; hence cue contrast is predicted to be an effective instructional method. Seventy-eight students were randomly assigned to four article training conditions (to learn 33 cues) and a control condition (to write essays). We found that article-training groups significantly outperformed the control group in an immediate posttest and a delayed posttest. Specifically, our result also suggested that there was a significant interaction between cue contrast and cue type (definite vs. indefinite). Cue contrast promoted more learning on the indefinite cues (more difficult for learners). Knowledge tracing did not demonstrate such an interactional effect with cue types. Instead, it boosted the instructional effect promoted by cue contrast.

Keywords: knowledge tracing; cue contrast; cognitive tutor; second language acquisition; English article.

Introduction

Since the mid-1990s Corbett & Anderson's (1995) Bayesian knowledge tracing model has been widely used to model student knowledge in learning systems of various domains, including tutors for mathematics, computer programming, and reading skills (Baker et al., 2010). In recent years, there has been an emergence of tutoring systems designed to facilitate second language learning (MacWhinney, 1995; Pavlik & Anderson, 2005). Among them we rarely find learning systems adopting Bayesian knowledge tracing to promote robust language learning (Koedinger, Corbett & Perfetti, 2012).

This paper introduces a Bayesian tutorial system of grammar instruction applied in an English as a Foreign Language (EFL) context. The primary goal of this research has been to develop an adaptive vehicle for testing the efficacy of Bayesian Knowledge Tracing in this domain. Another feature of the tutor, which presents grammatical cues in contrast, is informed by the cognitive linguistic theories of the Competition Model (MacWhinney, 1997). This paper discusses how these two areas of thought are blended to shape the design of the tutor and how they interact to influence learning effects. Specifically, the tutor targets the English article system, a difficult grammatical category for second language learners (Butler, 2002; Celce-Murcia & Larsen-Freeman, 1999).

Bayesian Knowledge Tracing

Corbett and Anderson's (1995) Bayesian knowledge tracing assumes that at any given opportunity to use a rule within the software, there exists a probability that a student knows the rule and may either give a correct or incorrect response. A student who does not know a skill generally will give an incorrect response, but there is a certain probability (called $p(G)$, the Guess parameter) that the student will give a correct response. Correspondingly, a student who does know a skill generally will give a correct response, but there is a certain probability (called $p(S)$, the Slip parameter) that the student will give an incorrect response. Each rule has an initial probability ($p(L_0)$) of being known by the student, and at each opportunity to practice a skill, the student has a certain probability ($p(T)$) of learning the skill. Once these four parameters are set, the model can be used to predict student performance. Figure 1 illustrates the relationship between the four parameters in the Bayesian Knowledge Tracing Model.
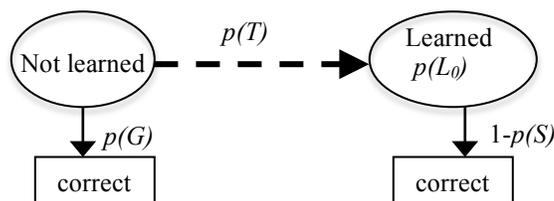

Figure 1: Bayesian Knowledge Tracing Model (Corbett and Anderson, 1995)

The system's estimate that a student knows a rule at time n ($P(L_n)$) is continually updated every time the student responds (correctly or incorrectly) to a problem step. First,

the system calculates the probability that the student knew the rule before making the attempt, using the evidence from the current step. Then, taking this into account, it computes the probability that the student learned the rule during the problem step. The equations for these calculations are:

$$P(L_{n-1}/Correct_n) = \frac{P(L_{n-1})*(1-P(S))}{P(L_{n-1})*(1-P(S))+(1-P(L_{n-1}))*(P(G))}$$

$$P(L_{n-1}/Incorrect_n) = \frac{P(L_{n-1})*P(S)}{P(L_{n-1})*P(S)+(1-P(L_{n-1}))*(1-P(G))}$$

$$P(L_n/Action_n) = P(L_{n-1}/Action_n)+((1-P(L_{n-1}/Action_n))*P(T))$$

To set the initial P(S), P(G), P(T), and P(L_0) parameters for each skill, we used data from a previous English article study to train each model (Zhao, 2012). During this previous study, we collected 10,523 student attempts at choosing the correct article, with each attempt labeled with the article rule applied. We used the "brute force" method (Baker et al, 2010) to utilize this data and arrive at the most likely parameter values. This method tries every possible combination of the four parameters in the grain size of 0.01 and for each combination, computes the sum of squared residuals (SSR). The parameter value combination that gives the best SSR for that rule are the ones we use to model the rule in the tutor.

The Competition Model

An important feature of the tutorial system in this paper is its theoretical ground in cognitive linguistics. We adopted the Competition Model (MacWhinney, 1997) and integrated one of its fundamental principles (cue competition) into the design of grammar instruction.

The model presents a functionalist rather than nativist view of language acquisition and understands the linguistic sign as a set of mappings between forms and functions. *Forms* are the external phonological and word order patterns that are used in words and syntactic constructions. *Functions* are the communicative intentions or meanings that underlie language usage. Each lexical item or syntactic construction can be understood as a form-to-function mapping.

In the context of discussion of the English article system, there are four forms: *the*, *a*, *an*, and the zero article (*0*). The form *the* specifies the definite article; the forms *a* and *an* encode the indefinite article (*a* is followed by noun phrases starting with consonants; *an* is followed by noun phrases starting with vowels); the form *0* is commonly known as the zero article or null article.

How about functions in the English article system? Adopting the model, we carried out a functional linguistic analysis and found that the four article forms are mapped with approximately 90 different functions or usages (more information see Zhao, 2012). Some functions are syntactic and semantic properties (e.g., countability, singularity, plurality); some functions are discourse-based properties (e.g., first mention, second mention, immediate situation); many functions are idiosyncratic surface features whose usage is highly conventional (e.g., names of rivers, lakes, malls, parks, bridges, theatres); and some functions combine both syntactic and idiosyncratic properties (e.g., names of singular mountains or plural mountains). So many functions are mapped with only four forms. This complex form-function mapping is one of the critical reasons why English articles are difficult to acquire.

The Competition Model understands one form-function mapping as a unit or a *cue*. E.g., the tag "*the – river names*" represents a cue since it maps the association between the form *the* and the semantic property of names of rivers. But the tag "*the – river names and second mention in the discourse*" is not a valid cue because one form is mapped with two functions. In this case, the form-function mapping needs to be broken down to the smallest unit.

The basic claim of the Competition Model is cue competition. It considers cue competition as vital for language acquisition. Sentences (1-2) illustrate how cues compete and one gains dominance over another. In sentence (1), the zero article is required because "wealth" is a noncountable mass noun and is used alone with no modifiers. When this noncountable noun is modified by the prepositional phrase (PP) "of her parents" in sentence (2), the noun "wealth" becomes concrete and identifiable. Hence, the PP structure (strongly associated with *the*) overrides noncountability (strongly associated with *0*) in sentence (2).

(1) Alice is interested in **_0_** wealth.
(2) Alice is interested in **_the_** wealth of her parents.

Early stages of language acquisition focus on obtaining input on individual cues from the learning environments. Learners may not know when a cue can or cannot override another cue. As their language proficiencies increase, learners develop their skills of interpreting cue conflict. Some cue competitions are easier to interpret than others. For example, the competition between the cue "*a – first mention*" and the cue "*the – second mention*" is relatively easier to interpret and are among the earliest acquired cues in the article system. But the competition between the two cues in Sentence (1-2) is relatively harder to interpret. Learners need to know the grammatical properties of a PP structure, its strong association with the definite article, and how a PP structure typically interacts with mass nouns.

This paper proposes an instructional invention of cue contrast that originates from the basic claim of the Competition Model. There are two theoretical justifications for the proposal. First, regularities, and heuristics are always good for learning a complex problem space (Reber et al., 1980; Ellis, 2011). Instead of understanding the article system as a space with almost 90 unrelated usages, learners formulate a more organized mental space with more than 40 contrasting pairs of usages. That leads to the most important advantage of cue contrast: it significantly reduces learners'

memory load and storage cost, and consequently increases their learning capacity. One major theoretical commitment made in the Competition Model is to a capacity-limited model of language processing. This account treats sentence interpretation as a constraint satisfaction process that balances the limitations imposed by verbal memory against the requirements of conceptual interpretation. Our raw memory for strings of nonsense words is not more than about four. However, when words come in meaningful groups, we can remember dozens of words, even when the message is unfamiliar. The most likely candidate for this additional storage is some form of conceptual representation. By presenting article usages to learners as meaningful groups, we help learners form the conceptual representation of the article system. In turn, storage cost is reduced.

Method

Participants

The participants of the current experiment were 78 (31 males, 47 females) Chinese intermediate-advanced learners at a public university in Beijing that specialized in foreign language education and research. Their average years spent learning English was 7.8 years.

Materials and Design

A 2 (Contrast: yes vs. no) × 2 (Knowledge tracing: yes vs. no) × 2 (Cue type: *the* vs. *0/a/an*)[1] mixed model design was used in this experiment. Cue contrast and knowledge tracing are two between-subject variables, whereas cue type is a within-subject variable.

There were five conditions in the experiment (Table 1): four experimental groups who received article training; one control group who did not receive article training. The four experimental conditions were manipulated based on the two between-subject variables. The control group spent roughly the same amount of training time as the four experimental groups. Instead of learning articles, they were asked to write four English essays during the training sessions and weren't given feedback.

Table 1: Experiment conditions.

Cue Contrast (CC)	Knowledge Tracing (KT)		Control group
	KT-CC	noKT-CC	
	KT-noCC	noKT-noCC	

Article training was provided as a sentence-level cloze task. Figure 2 exemplifies training received by the two cue contrast groups. Students were given a prompt question and two sentence items that belonged to a pair of contrasting cues. They used a pull down menu to make choices and

---

[1] We grouped indefinite article cues with zero article cues due to the small number of *a/an* cues.

were given immediate feedback. The feedback included correct/wrong, cue name, explicit cue explanations, and examples. Training pages in the two no-cue-contrast conditions also included two sentences. But they cannot belong to a contrasting pair.
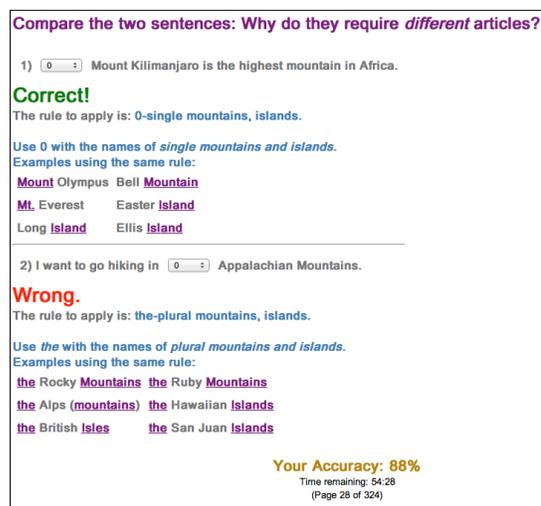

Figure 2: A training page of cue contrast conditions.

This experiment trained 33 article cues that were grouped into 18 pairs of contrasting cues. Definite article cues were paired with indefinite article or zero article cues primarily based on <u>structural</u> distinction. Table 2 exemplifies five representative cue pairs. In the first and second pairs, article choices are distinguished based on the existence of a relative clause as a post-modifier. The distinction in the third pair is due to the "of …" prepositional phrase as a post-modifier. The contrast in the four pair originates from the singular vs. plural distinction of mountain names. When we could not rely on structural distinction to create cue pairs, we relied on <u>semantic</u> distinction. The 'hall' and 'building' cues in the fifth pair, for example, are idiosyncratic cues whose article choices can only be explained by historical conventions. Hence, we relied on their semantic distinction to manipulate the contrast.

Table 2: Exemplar Cue Pairs (With Examples)

1	*a* –singular countable (*a* store)	*the* – singular countable+clause (*the* store she bought the dress)
2	*0* – noncountable (*0* wealth)	*the* – noncountable+clause (*the* wealth of her parents)
3	*0* – XX University (*0* Yale University)	*the* – the University of XX (*the* University of Chicago)
4	*0* – single mountains (*0* Bell Mountain)	*the* – plural mountains (*the* Rocky Mountains)
5	*0* – hall (*0* Baker Hall)	*the* – building (*the* Tepper Building)

Implementing Bayesian Knowledge Tracing

We adopted Corbett and Anderson's (1995) Bayesian knowledge tracing algorithm and used Baker et al.'s (2010) approach to train the model parameters using learner data from a previous but similar article tutor study (Zhao, 2012).

With the models for each article skill trained this way, the tutor updates $P(L_n)$ after observing correct/incorrect attempts at each skill and uses $P(L_n)$ for the item selection criteria. If we simply chose the next item to present with the lowest $P(L_n)$, the tutor would frequently show the same item back-to-back. To introduce some variety while still giving more practice on the items least learned, we use $P(L_n)$ to set the selection criteria of training items as: random selection in proportion to percent unlearned. Percent unlearned for a rule is calculated by taking the probability the rule is "unlearned": $(1-P(L_n))$, divided by the sum of the probabilities each other rule is "unlearned". A difficult cue will have a higher percent unlearned compared to other cues and will thus be more likely to be chosen next. But it is still possible (but less likely) that a better-acquired cue will be chosen next. This selection criterion avoids over-training of unlearned cues and under-training of better-acquired cues.

Procedure

The study was composed of three sessions. Session I included a pretest (25-min) and the first training session (1 hour). Session II (2 days later) included the second training session (1 hour) and an immediate posttest (25-min). Session III (2 weeks after Session 2) included a delayed posttest (25-min). All sessions were administered online. The tests were also in the format of a sentence-level cloze.

Results

Instructional Effects

A univariate analysis indicates that there was no significant difference of mean accuracy between the article training groups and the control group in the pretest ($F= .207$, $p= .651$). These two groups showed a significant difference of mean accuracy in the immediate posttest ($F= 37.836$, $p< .001$). The article training groups gained a mean accuracy of .154 (SD= .071), whereas the control group only gained a mean accuracy of .026 (SD= .076). Figure 2 illustrates the learning trajectories of the above conditions.

A paired samples t-test shows that the article training groups significantly improved mean accuracy from pretest to immediate posttest ($T= 17.156$, $p< .001$). Though they had a significant drop of accuracy from immediate-posttest to delayed-posttest ($T= -3.774$, $p< .001$), their mean accuracy in the delayed posttest was significantly higher than their pretest level ($T= 13.374$, p< .001). This suggested that the article training groups retained learning two weeks after training.

A paired samples t-test suggests that the control group did not show improvement from pretest to immediate posttest ($T= 1.359$, $p= .196$). Neither did they improve mean

accuracy from immediate-posttest to delayed-posttest ($T= 1.361$, $p= .195$).
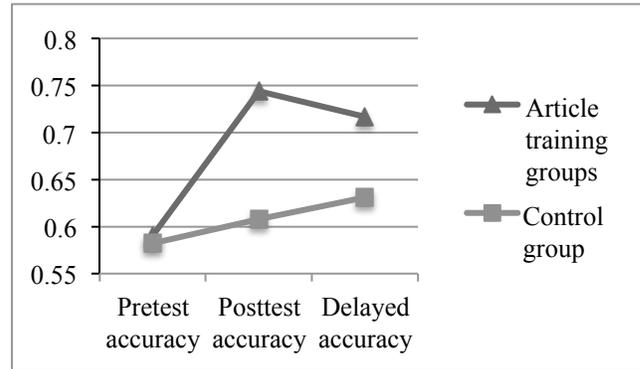

Figure 2: Mean test accuracies of conditions.

Cue Contrast, Knowledge Tracing, Cue Type

Figure 3 compares the four article training conditions (KT-CC, KT-noCC, noKT-CC, noKT-noCC) in terms of mean accuracies of all the cues in the pretest, posttest, and delayed posttest. In the pretest, univariate analysis indicates no significant difference of mean accuracy among the four conditions when all the cues are examined ($F = .468$, $p = .759$, $\eta 2 = .025$). Meanwhile, univariate analyses of pretest accuracy also show no significant difference among the four conditions regarding the acquisition of *the*-cues ($F = 1.083$, $p = .371$) and *0/a/an*-cues ($F = .441$, $p = .778$).
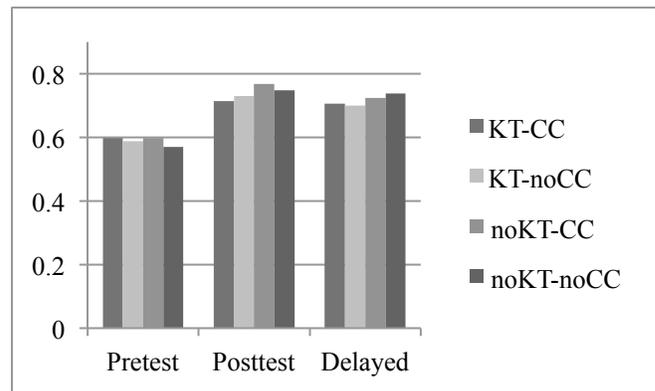

Figure 3: Mean accuracies of three tests of four conditions

In the immediate posttest, univariate analyses suggest no significant difference between contrast and no-contrast ($F = .110$, $p = .741$, $\eta 2 = .002$) and between knowledge-tracing and no-knowledge-tracing ($F= 2.435$, $p= .124$, $\eta 2= .038$), when all the article cues are examined. The same pattern was found in the delayed posttest. We then delved into four conditions and looked for variations. Again, when all article cues were considered, we could not find significant differences among the four conditions in the immediate

posttest ($F= 2.165$, $p= .102$, $\eta 2= .099$) and in the delayed posttest ($F= 1.464$, $p= .234$, $\eta 2= .069$).

Our next step was to explore interaction between the two primary variables (KT, CC) and cue type (*the* vs. *0/a/an*). A repeated measure ANOVA suggested that there was a significant interaction (Figure 4) between cue contrast and cue type ($F= 9.744$, $p< .001$, $\eta 2= .138$). Bonferroni pairwise comparisons indicated that within each cue type, the differences between cue contrast and non-cue contrast were significant: *the*-cues ($p< .001$) and *0/a/an*-cues ($p< .001$). Cue contrast promoted significantly more learning ($p< .001$) of *0/a/an* cues than of *the* cues. In contrast, non-cue contrast created a more balanced instructional effect. Its instructional effect on *the* cues and *0/a/an* cues were not significantly different ($p= .844$).

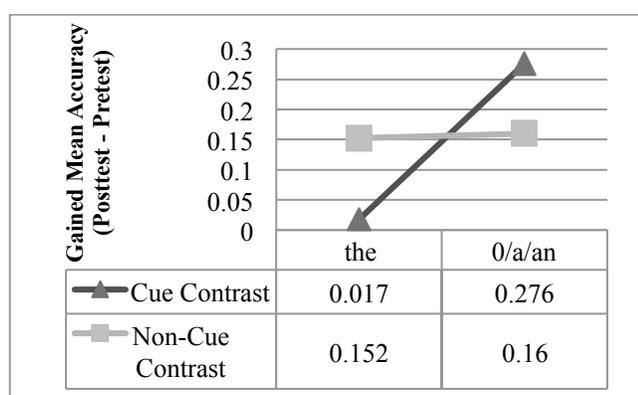|	the	0/a/an
Cue Contrast	0.017	0.276
Non-Cue Contrast	0.152	0.16

Figure 4: Contrast versus non-contrast

Yet, we did not find significant interaction between knowledge tracing and cue type. Bonferroni pairwise comparisons suggested no significant differences between knowledge tracing and non-knowledge tracing within *the* cues ($p= .557$) or *0/a/an* cues ($p= .385$).

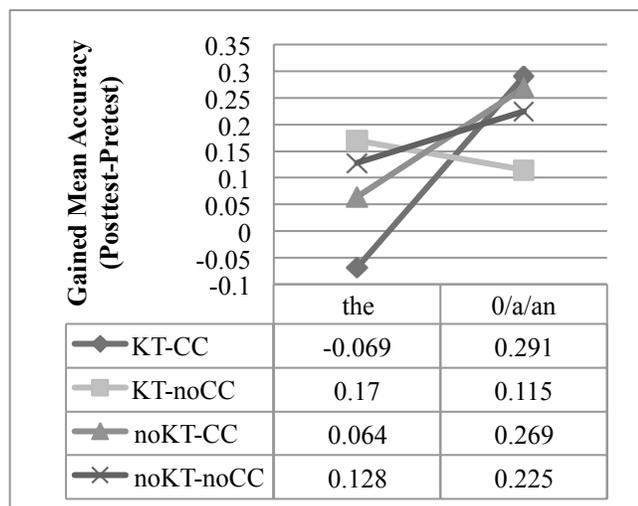|	the	0/a/an
KT-CC	-0.069	0.291
KT-noCC	0.17	0.115
noKT-CC	0.064	0.269
noKT-noCC	0.128	0.225

Figure 5: Gained accuracy (immediate posttest – pretest) of *the*-cues and *0/a/an*-cues in four training conditions

Figure 5 illustrates an interaction between cue types (*the* vs. *0/a/an*) and the four training conditions. The Y-axis is gained mean accuracy (immediate posttest–pretest). As we can see, the cue contrast conditions (KT-CC, noKT-CC) pushed for more learning of the *0/a/an* cues. In particular, KT-CC had a more tilted slope leaning towards more learning on *0/a/an*, which sharpened the interactional effect between cue contrast and article forms. Meanwhile, the non-cue contrast conditions (KT-noCC, noKT-noCC) were suggested to promote more learning on *the* cues. This trend appears to be more obvious in the KT-noCC condition than in the noKT-noCC condition.

The above findings indicated that cue contrast played the primary role of determining patterns of results. Knowledge tracing did not change the interactional effects between cue contrast and article type. It gave the interactional effects a boost and made the patterns sharper.

Discussion and Conclusion

The intermediate-advanced students in this study were given a difficult task of learning 33 article cues within two hours of training. The article training groups managed to show significant learning in the immediate posttest and retained learning two weeks later. This positive instructional effect confirms two principles that MacWhinney (1995) suggested to be important in designing and evaluating foreign language tutoring systems: 1) practice makes perfect and 2) feedback promotes learning (p. 318-319). Our study clearly demonstrated that given a difficult learning task, students would learn when they are given enough practice trials, accurate and digestible feedback, and an effective instructional method that helped to reduce memory and learning loads. The study implications for a third principle of MacWhinney (1995), that 3) cue conflicts are crucial for learning, are less clear.

As indicated by the interaction with type of item ('the' vs. 0/a/an), the cue contrast manipulation had a clear impact, raising performance on *0/a/an*-cues but lowering performance on *the*-cues. By analyzing contrasting cue pairs, it was hypothesized that learners in the cue contrast condition would formulate a new understanding of the article system and develop a more systematic knowledge space. Knowledge tracing effects, in comparison, would be less visible. Indeed, KT did not change how learners conceptualized the article system. It mainly functioned to escalate learning.

Why did only cue contrast show interaction with cue type (*the* vs. *0/a/an*)? Because that was the way the tutor was designed. Contrasting pairs were created because they shared similar features (e.g., *mountain names*) but required different article forms (e.g., *0-single mountain names*, *the-plural mountain names*). But knowledge tracing was not manipulated based on article type.

And why did cue contrast promote more learning on *0/a/an* cues than *the* cues? At first we suspected an ordering effect behind this interactional effect, i.e., cue contrast made learners pay more attention to the first item on a training

screen and consequently paid less attention to the second item. But in fact there were more *the* items than *0/a/an* items at the top of the training screen. Meanwhile, a repeated measure ANOVA also suggested no significant interaction between cue contrast and item sequence (F= 2.427, p= .124) and no main effects of item sequence (F= 1.311, p= .991) or of cue contrast (F= .049, p= .825). Also we suspected that cue contrast groups got a higher frequency of exposure to 0/a/an items. Yet this was not confirmed either. Frequency of exposure to *the*-items and *0/a/*an-items was balanced between contrast and no-contrast conditions.

Then the most plausible account was that cue contrast illuminated weak areas of learning. *0/a/an* cues were poorly acquired by learners. Their pretest mean accuracy of *0/a/an* cues (M= .472) was significantly lower (p< .001) than *the* cues (M= .761). We found that a particular problem associated with the zero article acquisition was due to a misunderstanding made by students[2]. They thought that all the so-called "proper nouns" (e.g., Lake Michigan, the Colorado River, Baker Hall, the Tepper Building) were unique and therefore had to be used with the definite article. They did not know that some of such noun phrases required the zero article. Therefore, it became enlightening moments for students to see two proper nouns being contrasted in one screen. They allocated more time and attention to the zero article proper noun. As we can see, the mechanism behind cue contrast was cue focusing that directed students' attention to the right areas.

Due to the interaction between contrast and cue type, there was a trade-off that cancelled the overall instructional effect of cue contrast when all cues were considered.

The last question to discuss is the reason why knowledge-tracing groups did not outperform non-knowledge-tracing groups. The most plausible reason was the relatively short duration of instruction. Two hours might not be enough for knowledge tracing to demonstrate its full advantages. Learners in the knowledge tracing conditions were in the middle of tackling the most difficult cues when training ended. They did not have enough time to work on the less difficult cues. The posttest mean accuracy (.744) of the article training groups was far from the mastery level (.950). This sent us a stronger signal that a longer training time was needed for knowledge tracing to be more effective.

In short, this study demonstrated a successful application of cognitive psychology and human-computer interaction theories in second language grammar instruction. We found that cue contrast was an effective method in teaching English article usages to adult second language learners. In particular, contrast allowed learners to become aware of and shift focus to problematic knowledge domains. Knowledge tracing boosted instructional effects of cue contrast. More research is needed to further specify duration of instruction so that we can make the best use of knowledge tracing in second language grammar instruction.

[2] We carried out semi-structured interviews with selected students. Due to space limit, the interview data is not reported here.

References

Baker, R. et al. (2010). Contextual Slip and Prediction of Student Performance After Use of an Intelligent Tutor. *Proceedings of the 18th Annual Conference on User Modeling, Adaptation, and Personalization*, 52-63.

Butler, Y. G. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*, *24*, 451-480.

Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The Grammar Book: An ESL/EFL Teacher's Course* (2nd Edition). U.S.: Heinle & Heinle Publishers.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*, 253-278.

Ellis, N. (2011). Implicit and explicit SLA and their interface. In Sanz, C. & Leow, R. P. (Eds.), *Implicit and Explicit Language Learning: Conditions, Processes, and Knowledge in SLA and Bilingualism*. Washington, DC: Georgetown University Press.

Koedinger, K. R., Corbett, A. C., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*, 757-798.

MacWhinney, B. (1995). Evaluating foreign language tutoring systems. In V. M. Holland, J. D. Kaplan & M. R. Sams (Eds.), *Intelligent Language Tutors: Theory Shaping Technology*. Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (1997). Second language acquisition and the Competition Model. In J. Kroll & A. De Groot (Eds.), *Tutorials in Bilingualism*. Mahwah, NJ: Lawrence Erlbaum.

Pavlik Jr., P. I. & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*, 559-586.

Reber, A. S. et al. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 492-502.

Zhao, H. (2012). *Explicitness, Cue Competition, and Knowledge Tracing: A Tutorial System for Second Language Learning of English Article Usage*. Unpublished PhD Dissertation, Carnegie Mellon University.