Modeling Second Language Preposition Learning

Libby Barak (libby.berk@gmail.com),
Scott Cheng-Hsin Yang (scottchenghsinyang@gmail.com),
Chirag Rank (chiiragrank@gmail.com),
Patrick Shafto (patrick.shafto@gmail.com)
Department of Math and Computer Science, Rutgers University,
Newark, NJ 07102

Abstract

Hundreds of millions of people learn a second language (L2).[1] When learning a specific L2, there are common errors for native speakers of a given L1 language, suggesting specific effects of L1 on L2 learning. Nevertheless, language instruction materials are designed based only on L2. We develop a computational model that mimics the behavior of a non-native speaker of a specific language to provide a deeper understanding of the problem of learning a second language. We use a Naive Bayes to model prepositional choices in English (L2) by native Mandarin (L1) speakers. Our results show that both correct and incorrect responses can be explained by the learner's L1 information. Moreover, our model predicts incorrect choices with no explicit training data of non-native mistakes. Our results thus provide a new medium to analyze and develop tools for L2 teaching.

Keywords: Computational model, Second language learning, Preposition learning, N-gram model, Bayesian model

Introduction

Statistics show that for the three most spoken languages in the world, English, Mandarin-Chinese, and Hindi, a significant portion of speakers are second language (L2) learners (65%, 45%, and 18% respectively).[1] L2 speakers rarely reach competency level similar to native-like performance (e.g. Robenalt & Goldberg, 2016; Bley-Vroman & Joo, 2001), raising questions about why their performance is limited and what methods might improve their learning. Second language teaching often uses material that goes through sections of topic-specific words or lists of all possible syntactic extensions, which are based on the structure of the L2 to be learned. Importantly, learners from a given L1 exhibit common patterns in all levels of their L2 use starting from choice of words, reading patterns, and grammatical mistakes (Goldin, Rabinovich, & Wintner, 2018; Berzak, Nakamura, Flynn, & Katz, 2017). These group patterns are hypothesized to result from transfer or interference of L1 (first language) knowledge. This paper sets out to better understand such L1-based characteristics of L2 use as a step toward better automated teaching curricula.

Computational research has investigated use of L1 to improve grammatical error detection and correction in L2 (Chodorow, Tetreault, & Han, 2007; Dale, Anisimoff, & Narroway, 2012; Ng et al., 2014; Rozovskaya, Roth, & Sammons, 2017). For example, Rozovskaya et al. (2017) incorporate L1 ID into the features used by a Naive Bayes algorithm to predict the mistakes made by non-native users of English. Their method is effective when trained on small scale L2 learner annotated mistakes or large native English data of billions of words. This method can offer support for users who have already gained some knowledge of the language by correcting mistakes in their L2 productions. However, identifying a potential lack of grammaticality alone does not clearly provide an explanation as to why a learner produced such mistake or how to better teach human learners.

Previous research has used recent advances in computational linguistics to automatically generate L2 teaching material optimized for learners' needs. For example, (Xiao, Wang, Zhang, Tan, & Chen, 2018) use a small set of multiple-choice preposition selection questions from textbooks as seed data, which range in difficulty level to fit learners at different proficiency level. They propose a computational approach based on word embedding to generate additional questions in line with the seed textbook data. Whereas they aimed to automatically replicate existing questions, our goal is to develop methods that are tailored to learners from different L1 languages, and their signature errors.

Proposed Model of Preposition Learning

Our vision draws inspiration from pedagogical models of ideal teaching which create optimal teaching data based on a simulated learner (Shafto, Goodman, & Griffiths, 2014; Eaves Jr, Feldman, Griffiths, & Shafto, 2016). Ideal teaching models predict the optimal data sequence for a learner by modeling the learning and the data generation as a coupled process, similar to teachers providing the most useful examples to explain a topic. The core of an effective teaching model is a good learner model that simulates the learner's knowledge state over the learning process. Once a learner model is specified, and not before, one may generate optimal teaching data in systematic ways (Eaves JR & Shafto, 2016). Thus, in this paper we focus on the core first step of developing learner models that can capture L2-learner behavior, for L1 native speakers, on the tasks of interest.

Most computational models for L2 acquisition aim to capture learner's mistakes after they are made, rather than analyzing potential causal factors (Settles, Brust, Gustafson, Hagiwara, & Madnani, 2018). These methods are based on training on large amounts of L2 data and considerable data

[1] According to https://www.ethnologue.com/

of L2 learner mistakes (7M words). Successful approaches gain performance from choice of method and size of data, rather than cognitively-inspired features (Settles et al., 2018). While powerful, given enough data, such approaches risk limited transfer across contexts by mistaking merely correlated variables for causally relevant factors.

We aim to explain both successes and mistakes of human learners as a function of features of their L1 to clarify the strengths and weaknesses of current teaching approaches. Unlike previous work on error detection, our goal is to predict learner's mistake using L1 and L2 data only without (1) training data of learner's errors, or (2) extensive L2 data non-plausible for an L2 learner. We focus on the task of preposition selection with Mandarin as L1 and English as L2 using the experimental data of Xiao et al. (2018) as gold standard. We follow the Naive Bayes framework proposed by Rozovskaya et al. (2017), a simple yet flexible approach that can be adapted to any language pairing with minimal cost and that has been shown to perform better with identification of the L1 grouping. We train monolingual learners using dataset from a single language and model L2 learners by combining data from two languages. Psycholinguistic evidence suggests that native speakers of a specific Language, e.g., Mandarin, present similar mistake pattern when learning a common L2, e.g., English. We verify such group-specific characteristics by comparing the results from our Mandarin-English model to baseline models of English as a second language using other possible L1 choices (Hebrew and Japanese).

Our results will show that L2 learners are better at identifying the correct choice when it correlates with the frequency of the preposition in their L1. Additionally, mistakes correlate with predictions of our L2 simulated learner, indicating that mistakes are driven by both frequency of the preposition and contextual information of words around the preposition in L1 and L2. We conclude that learners require teaching data that addresses potential biases in their L2 knowledge created by discrepancies in regularities between the two languages. While simple frequency information fails to replicate this behavior, our model offers a computational leaner that can be used to develop L1-specific teaching material.

Grammatical Errors in Preposition Selection

Preposition selection errors represent the largest category (29%) of grammatical mistakes made by non-native speakers (Bitchener, Young, & Cameron, 2005). Prepositions also comprise a large portion of verbal communication. Analysis of multiple corpora indicate that 10% of top 100 most frequent words used in English are prepositions (Xiao et al., 2018; Schneider et al., 2016). Moreover, the highly polysemous meaning of prepositions heighten the difficulty of forming a cohesive representation of the meaning of each preposition (Schneider et al., 2016).

To facilitate modeling of preposition learning, Xiao et al. (2018) expand on multiple choice questions of preposition selection from a textbook using large data set by finding question and word embeddings and selecting distractors. Xiao et al. (2018) recruited 491 participants to complete 100 questions spanning over 4 difficulty levels (see Table 1). Participants were all native speakers of Mandarin with English Proficiency ranging from beginner to advanced language users (high-school, undergraduate, and graduate students). Each question was completed by 88 to 163 participants. Their experimental results showed that average score drops as difficulty increases, confirming the plausibility of their method in matching the hypothesized difficulty level for each group. The data therefore includes both proportions of correct as well as incorrect answers for all questions. It covers a wide range of prepositions over a rich selection of context words.

小	框子	用	来	做	什么	?
small	frame	with	come	make	what	?
what	is	the	small	frame	used	for?

(a) Mandarin translation

אני	בכלל	לא	הסתכלתי	על	ה	שעון
I	at_all	no	look	on	the	clock
I	didn't	look	at	the	clock	at_all

(b) Hebrew translation

Figure 1: Examples taken from the L2 training data including L1 sentence, word-by-word translation, and parallel meaning translation.

Experiment

Our goal is to model the decisions of an L2 learner in selecting a preposition for a given context. Following theories of interference and transfer, we train the learner using a Naive Bayes model on n-grams representing the context of preposition in each language. The use of n-grams as training data represents a learner that uses language-specific distribution is predicting a preposition for a context. For example, for the first question in Table 1, the English 3-gram, "opinion of herself" will suggest *of* as the likely preposition based on English distribution. While, in Mandarin, a more common 2-gram "by herself", may mislead the learner resulting in the selection of *by*.

To capture learning over language-specific n-grams, we use training data translated word-by-word to English (see Figure 1). This method captures the meaning of the word in the original language, as opposed to full-sentence translation that aims to align the words to their intended meaning in the target translation. We investigate the effects of L1 on L2 by comparing a model trained on different L1 languages to training to the learner's native L1. We predict that the participants' choices will correlate with the properties of their native language more so than the second language. That is, that a model trained on Mandarin and English data (**L1:Man+L2:Eng**) will match the choices made by Mandarin native-speakers more closely than a model trained on English alone (**L1:Eng**), Mandarin alone (**L1:Man**) or on English combined with other languages. Moreover, we in-

Difficulty Level	Question	Candidate Answers			
1	She has a low opinion ___ herself	(a) with	(b) by*	(c) inside	**(d) of**
3	After all, it hadn't really been her fault that she became mixed ___ in Jack 's business affairs	**(a) up***	(b) round	(c) after	(d) down

Table 1: Examples of test items included in the experimental design of (Xiao et al., 2018). Difficulty levels range from 1 (easiest) to 4 (hardest). Numbers in brackets indicate the percentage of participants selecting this answer. Correct choice is given in bold and participants' top choice marked by an asterisk.

vestigated whether our model predicts mistakes, despite not being trained directly on errors, by capturing the interference and transfer specific for the precise L1-L2 pairing.

Naive Bayes Model

We describe a Bayesian model of the participants' choice in the preposition selection task. Let c denote a particular choice of preposition from the available set of prepositions (the Candidate Answers in Table 1) and q denotes the question (the Question in Table 1). The probability that a participant chooses c given the question q after observing the data D, can be expressed as:

$$P(c|q,D) = \frac{P(c,q|D)}{\sum_c P(c,q|D)}.$$

D can be purely L1 training data (L1 speaker) or a mixture of L1 and L2 training data (L2 speaker). The former aims to capture the influence of the participant's L1 knowledge on the task, while the latter represents the partial L2 knowledge. Following the settings of the experimental design, the probability of each choice, c is measured over the four possible choices for each question as denoted by the sum in the denominator over those four choices.

For each choice c we extract all n-grams from the question q, resulting in m question-based n-grams. The joint probability is calculated using the multinomial Naive Bayes model over the n-gram features extracted from the data, D:

$$P(c,q|D) = \frac{m!}{k_{1c}!\cdots k_{Vc}!}\theta_{1c}^{k_{1c}}\cdots\theta_{Vc}^{k_{Vc}},$$

m is the number of n-grams gathered from the question sentence after filling in the blank with the preposition choice c of interest; k_{jc} is the number of times the j^{th} n-gram in the data appears in the question sentence; V is the number of types of n-grams extracted from the data, D (listed in the N-grams column in Table 2). The probability of c being the correct choice according to the n-gram θ_{jc} is measured by the relative frequency of this n-gram for the preposition:

$$\theta_{jc} = \frac{n_{jc}+1}{n_c+V}.$$

Where, n_{jc} is the frequency of the j^{th} n-gram extracted from all the sentences in D that contain the preposition c; and $n_c = \sum_j n_{jc}$ is the frequency of all the n-grams extracted from the sentences that contain the preposition c. Intuitively, this model assigns a high probability to a choice c in a context q if c is observed with high frequency among sentences in the data that are similar to q in terms of n-gram composition. Mechanistically, $P(c,q|d)$ is high if large θ_{jc}'s are paired with non-zero k_{jc}'s. A k_{jc} typically takes on value 1 or 0, depending on whether or not it is found in the question sentence, and a large θ_{jc} means that the the n-gram indexed by jc is observed with high frequency in the data.

Training data

We used CHILDES corpora for all 4 languages as it provides us with the advantage of part-of-speech tagging, lemmatization, and, uniquely, word-to-word translation for several corpora across multiple languages. We extract our n-Grams from datasets corresponding to each of the 4 languages: English, Mandarin, Hebrew, and Japanese. We choose Hebrew and Japanese as two languages with similar number of n-grams for the prepositional data compared with Mandarin. We used 7 English corpora, the 11 Chinese-Mandarin corpora with word-by-word English translation, the 2 translated Japanese corpora, and 2 translated Hebrew corpora (Brown, 1973; Masur & Gleason, 1980; Kuczaj II, 1977; Sachs, 1983; MacWhinney & Snow, 1990; Suppes, 1974; Tardif, 1995; Hemphill et al., 1994; Chang, 2004; L. Li & Zhou, 2008; Xiangjun & Yip, 2018; Zhou, 2001; H. Li & Zhou, 2015; L. Li & Zhou, 2011; X. Li & Zhou, 2004; Miyata, 1992, 2012; Armon-Lotem, 1996).[2] Sentences that included a preposition were selected for n-gram extraction. To increase the quality of our n-grams, we removed words corresponding to the part-of-speech tags,[3] such as determiners, to approximate the relatedness denoted by the parse tree rather than raw ordering. This step removed 247 word types including words such as *bye*, and *gosh*. Each sentence in the data is labeled by the prepositions that it contains, and and all possible 2- to 5-grams (no skip-grams) are extracted from each sentence. The final number of sentences and n-grams for each language are presented in Table 2.

Methods

Evaluation The test data includes 100 multiple-choice questions with 4 candidate preposition for each question (see Table 1). For each candidate preposition, we calculate the

[2]No citation provided for `https://childes.talkbank.org/access/Chinese/Mandarin/Xinjiang.html`, `https://childes.talkbank.org/access/Chinese/Mandarin/Zhou3.html`, and `https://childes.talkbank.org/access/Other/Hebrew/Ravid.html`

[3]Tags: co, det:art, det:poss, neg, aux, mod, cop, cl, and cm.

percentage of people selecting this as the preferred answer. To obtain the corresponding probability for each candidate from the model, we first remove from the questions any words removed in the pre-processing step of the training data based on the part-of-speech tags.

Baseline models We compare our results to baseline models used to understand to what degree the preposition frequency alone can explain the results. If learners are better at prepositions they frequently heard while learning the L2, the English frequencies will predict their correct choices for these prepositions. But, interference from L1 would entail that the frequency of the preposition in L1 will predict learner's behavior. We constructed the baseline models for each of the languages listed in Table 2 with the form of:

$$P_B(c|s,D) = \frac{f_c + 1}{\sum_c f_c + 4},$$

where f_c is the frequency of the preposition choice c in the training data D. The addition of 1 to the frequency matches the Laplace smoothing in the Naive Bayes Models to avoid zero numerator or denominator.

Language	Sentences	N-grams	Prepositions
English	254,366	708,126	42 / 42
Mandarin	108,372	218,411	33 / 42
Hebrew	25,094	174,395	32 / 42
Japanese	30,797	198,247	35 / 42

Table 2: Number of sentences used for training in each language. The columns from left to right indicate the number of sentences from CHILDES that include at least one of the prepositions, the resulting number of n-grams (2- to 5-grams) used for training, and the fraction of preposition types in the training data out of all the preposition types present in the questions.

Results

We compare participants' choices to three types of models. The naming conventions for the models is as follows: First, the baseline models are denoted by the first 3 letters of the language name and *-Base*: e.g., *Man-Base* corresponds to a baseline model based on only the frequency (or base rate) of the prepositions in the Mandarin data. Second, single language models are denoted by *L1:* and the first 3 letters of the language name: e.g., *L1:Heb* corresponds to the model trained on the frequency of n-grams in the Hebrew data. Lastly, L2 learner models, which is a mixed model trained on L1 and L2 data, are denoted by *L1:X + L2:Eng*, where *X* can be *Man*, *Heb*, or *Jap*. For this type of model, we always use all of the L1 data and a random selection of 25% of the English data.

Analysis of all choices We first analyze the correlation of each model with the participants' responses, with the goal of identifying models that correlate well with experimental results of (Xiao et al., 2018). The analysis includes all the multiple-choice candidates, 4 for each of the 100 questions. Figure 2 shows a scatter plot of the participants' choice percentage on the y-axis and each model's choice probability on the x-axis along with the linear regression line for each model. We analyze the correct choices (based on the gold-standard for each question) and the incorrect choices (the three distractors for each question) separately.

For the correct choices, the *Man-Base* model has the highest correlation ($r = 0.246$) and a significant positive slope (Effect size=0.164, t(98)=2.515, p=0.0135). This result suggests that the frequency of the prepositions in the native language predicts the degree to which participants are likely to select the correct choice as an answer. When the frequency of the correct choice is high in Mandarin (RHS of the x-axis), participants easily make the right selection. When the correct candidate has low frequency in Mandarin (LHS of the x-axis), participants fail to make the correct choice.

The *Eng-Base* model has a positive but non-significant correlation with the percentages of correct choices (r=0.151, t(98)=1.516, p=0.133). A linear model with interaction effect shows that the *Man-Base* and *Eng-Base* do not generate significantly different fits (t(196)=0.491, p=0.624). The other models also do not correlate positively and significantly with participants' choice behavior. Although the correct candidates are drawn from the grammatical use in English, the participants' choice probabilities are not as influenced by L2 knowledge as it is by their own native language.

Looking at the data for the incorrect choices, we observe that the strongest correlation comes from the *L1:Man + L2:Eng* model (left panel of Figure 2); however, none of the models yields significant correlation. An inspection of the training data suggests that the child-directed speech training data does not contain enough occurrences to analyze choices for low-frequency prepositions (single occurrence). To address this limitation, we look at the top choices—the preposition selected by most of the participants for each question. This method naturally eliminates low-frequency prepositions from our evaluation and puts a magnifying glass on the choices that are most agreed upon by the participants.

Analysis of top choices In this evaluation, we focus on only the top choice, or the most popular selection, made by the participants for each question. Thus, this analysis covers 1 preposition for each of the 100 questions corresponding to 66 correct choices and 34 incorrect choices (see Figure 3). For the correct choices, the *Man-Base* model based on the frequency of the prepositions in Mandarin (L1) is still the strongest, although there is no significant correlation (r=0.128, t(64)=1.034, p=0.305). However, since this evaluation does not include the choices with lower percentage (note the y-axis begins from 0.3), the correlation loses its strength. Our analysis shows that top choices that are also correct choices include mostly prepositions with high frequency in Mandarin (L1), making it harder to correlate across the full range.[4]

[4]In accordance with common methodology in corpus linguistic research, a modified baseline model based on log frequency of the

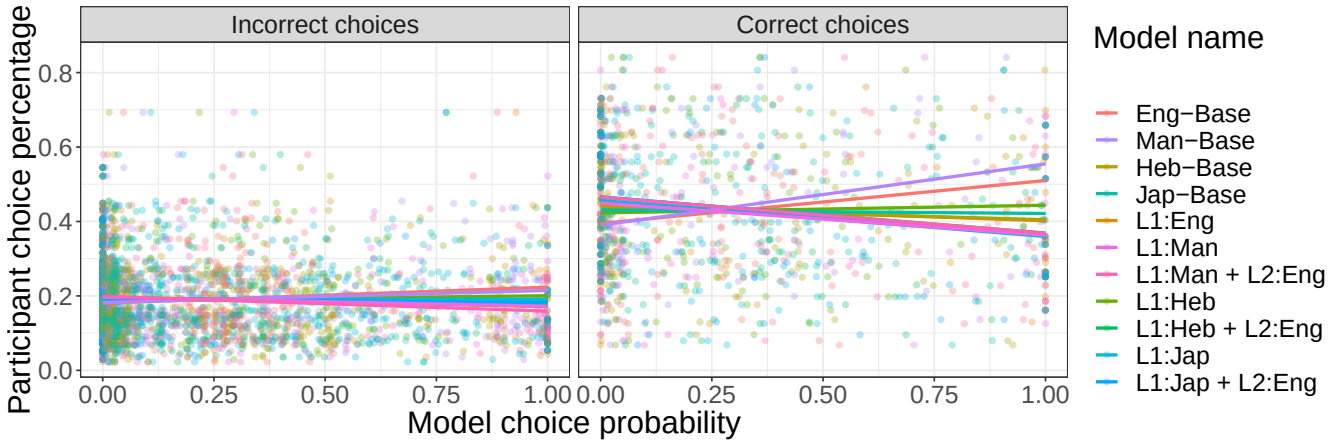

Figure 2: The correlation of our baseline models and Naive Bayes models with the participants' choice percentage over (1) the 100 correct choices (right panel), and (2) the 300 distractors (left panel). Only the frequency of the prepositions in Mandarin (*Man-Base*) predicts the percentages of participants' correct choices; none of the models predict those for the distractors.

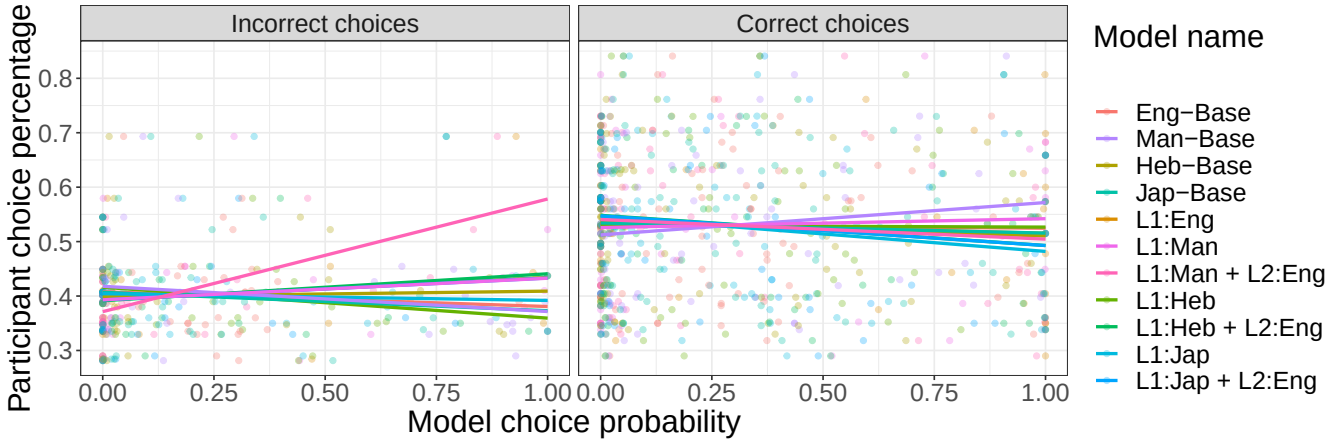

Figure 3: The correlation of our baseline models and Naive Bayes models with the participants' choice percentage over (1) the 66 correct choices chosen by participants as the top choice (right panel), and (2) the 34 incorrect choices selected by participants as the top choice (left panel). The percentage of participants' choice for the incorrect choices is predicted only by the probabilities provided by our model simulating a Mandarin speaker learning English (*L1:Man+L2:Eng*).

On the other hand, a linear regression analysis shows that the non-native learner model *L1:Man+L2:Eng* predicts the top incorrect choices (r=0.604, effect size=0.192, t(32)=4.284, p<0.001) above and beyond any of the other baseline models. Notably, the mistakes arise from the integration of linguistic properties that are known to play a role in language prediction—frequency and co-occurrence data, rather than training on data of learner's mistakes. Finally, models of L2 learners of other languages, here, Hebrew and Japanese, do not explain any of the participants behavior. The next best model, *L1:Heb+L2:Eng*, does not correlate significantly with participants' choice percentage (t(32)=1.028, p=0.312). Linguistic choices of L2 learners are explained by the target second language and the native language of the group of learners.

prepositions instead of frequency also do not yield significant correlation (r=-0.0204, t(64)=-0.163, p=0.871).

Discussion

Non-native learners are becoming a core population of the most frequently used languages. Teaching materials often go through topic-specific words or lists of all possible syntactic extensions, which are based entirely on the structure of the L2 to be learned. To the extent that computational tools make use of information about learners' native languages, they aim to improve detection and correction of errors, after they happen.

Toward the goal of developing models for teaching L2 languages to L1 speakers, we investigate development of models of second language learners. In contrast with prior work, our approach to modeling learner responses is based on properties of L1, augmented (in some cases) by those of L2. Unlike error detection model, we train our models only on language data without using data of common errors made by non-native speakers. Focusing on the problem of learning prepositions, we create a wide variety of models based on preposition baserates and n-gram frequencies in the learner's L1 and L2. We find that a model based on L1 preposition baserate is

a robust predictor of *correct* answers in L2, suggesting that learners' are strongly leveraging their prior knowledge when learning. Moreover, learners' frequent errors are strongly predicted by a model based on n-grams in their L1, combined by n-grams in L2. Results show that a host of alternative models based on baserates and n-grams from other languages do not explain the pattern of results. Overall, the results indicate robust influence of L1 on learning L2.

Our current results are limited by the size and scope of the training data. On the other hand, the word-to-word translation, uniquely provided by this dataset, enables the model to represent the context of preposition choice in sentence-specific data. While one-to-one mapping such as *subject-verb* may be captured by single word mapping across languages (Kochmar & Shutova, 2016), there is no one-to-one mapping of prepositions as a single word across languages (e.g. Beekhuizen & Stevenson, 2015). The translation of full sentences word-by-word provided a context-specific translation that enabled prediction of L2 mistakes.

Developing learner models is a first step toward models that improve teaching second languages by understanding the influence of L1. There are a number of important directions before this goal can be realized. First, our results show that we can predict errors, the next step would be to validate teaching of prepositions using this learner model. Second, our results are based on speakers of Mandarin who are learning English as a second language, and it will be important to consider speakers of other L1s learning other L2s. Third, prepositions are one small aspect of the overall language learning problem, and generalization to other linguistic features remains an open question. Because of the simplicity of the approach and the availability of the relevant training data, we are optimistic that this approach will allow development of improved models of second language teaching.

Acknowledgment

We thank Wenyan Xiao for sharing with us the experimental results of (Xiao et al., 2018). This research was supported in part by Department of Defense grant 72531RTREP, NSF SMA-1640816, NSF MRI 1828528, and NIH DC017217-01 to P.S.

References

Armon-Lotem, S. (1996). *The minimalist child: Parameters and functional heads in the acquisition of hebrew*. Tel Aviv University.

Beekhuizen, B., & Stevenson, S. (2015). Crowdsourcing elicitation data for semantic typologies. In *Proceedings of the 37th annual meeting of the cognitive science society, CogSci.*

Berzak, Y., Nakamura, C., Flynn, S., & Katz, B. (2017). Predicting native language from gaze. *arXiv preprint arXiv:1704.07398*.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on esl student writing. *Journal of second language writing*, *14*(3), 191–205.

Bley-Vroman, R., & Joo, H.-R. (2001). The acquisition and interpretation of english locative constructions by native speakers of korean. *Studies in Second Language Acquisition*, *23*(2), 207–219.

Brown, R. (1973). *1973: A first language: the early stages. cambridge, ma: Harvard university press*.

Chang, C.-J. (2004). Telling stories of experiences: Narrative development of young chinese children. *Applied Psycholinguistics*, *25*(1), 83–104.

Chodorow, M., Tetreault, J. R., & Han, N.-R. (2007). Detection of grammatical errors involving prepositions. In *Proceedings of the 4th ACL-SIGSEM workshop on prepositions.*

Dale, R., Anisimoff, I., & Narroway, G. (2012). HOO 2012: A report on the preposition and determiner error correction shared task. In *Proceedings of the seventh workshop on building educational applications using NLP* (pp. 54–62).

Eaves Jr, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological review*, *123*(6), 758.

Eaves JR, B. S., & Shafto, P. (2016). Toward a general, scaleable framework for bayesian teaching with applications to topic models. *IJCAI 2016 workshop on Interactive Machine Learning: Connecting Humans and Machines*.

Goldin, G., Rabinovich, E., & Wintner, S. (2018). Native language identification with user generated content. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3591–3601).

Hemphill, L., Feldman, H. M., Camp, L., Griffin, T. M., Miranda, A.-E. B., & Wolf, D. P. (1994). Developmental changes in narrative and non-narrative discourse in children with and without brain injury. *Journal of Communication Disorders*, *27*(2), 107–133.

Kochmar, E., & Shutova, E. (2016). Cross-lingual lexico-semantic transfer in language learning..

Kuczaj II, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 589–600.

Li, H., & Zhou, J. (2015). *The effects of pragmatic skills of mothers with different education on children's pragmatic development.* (Master's thesis, East China Normal University. Shanghai, China)

Li, L., & Zhou, J. (2008). *Metacommunication development of children aged from three to six during their collaborative pretend play.* (Unpublished Master dissertation, The University of Hong Kong)

Li, L., & Zhou, J. (2011). *Preschool children's development of reading comprehension of picture storybook: from a perspective of multimodel meaning making.* Unpublished doctoral dissertation, East China Normal University. Shanghai, China.

Li, X., & Zhou, J. (2004). *The effects of pragmatic skills of mothers with different education on children's pragmatic development.* (Master's thesis, Nanjing Normal University, Shanghai, China)

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of child language*, *17*(2), 457–472.

Masur, E. F., & Gleason, J. B. (1980). Parent–child interaction and the acquisition of lexical information during play. *Developmental Psychology*, *16*(5), 404.

Miyata, S. (1992). Wh-questions of the third kind; the strange use of Wa-question in japanese children. *Bulletin of Aichi Shukutoku Junior College*, *31*, 151–155.

Miyata, S. (2012). Nihongo mlu (heikin hatsuwachoo) no gaidorain: Jiritsugo mlu oyobi keitaiso mlu no keisanhoo [guideline for japanese mlu: How to compute mluw and mlum]. *Journal of Health and Medical Science*, *2*, 1–15.

Ng, H. T., Wu, S. M., Briscoe, T., Hadiwinoto, C., Susanto, R. H., & Bryant, C. (2014). The conll-2014 shared task on grammatical error correction. In *Proceedings of the eighteenth conference on computational natural language learning: Shared task* (pp. 1–14).

Robenalt, C., & Goldberg, A. E. (2016). Nonnative speakers do not take competing alternative expressions into account the way native speakers do. *Language Learning*, *66*(1).

Rozovskaya, A., Roth, D., & Sammons, M. (2017). Adapting to Learner Errors with Minimal Supervision. *Computational Linguistics*.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. *Children's language*, *4*(4), 1–28.

Schneider, N., Hwang, J. D., Srikumar, V., Green, M., Suresh, A., Conger, K., ... Palmer, M. (2016). A corpus of preposition supersenses. In *Proceedings of the 10th linguistic annotation workshop held in conjunction with ACL 2016.*

Settles, B., Brust, C., Gustafson, E., Hagiwara, M., & Madnani, N. (2018). Second language acquisition modeling. *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications*.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive psychology*, *71*.

Suppes, P. (1974). The semantics of children's language. *American psychologist*, *29*(2), 103.

Tardif, T. Z. (1995). Adult-to-child speech and language acquisition in mandarin chinese.

Xiangjun, D., & Yip, V. (2018). A multimedia corpus of child mandarin: The tong corpus. *Journal of Chinese Linguistics*, *46*(1), 69–92.

Xiao, W., Wang, M., Zhang, C., Tan, Y., & Chen, Z. (2018). Automatic generation of multiple-choice items for prepositions based on word2vec. In *International conference of pioneering computer scientists, engineers and educators.*

Zhou, J. (2001). *Pragmatic development of mandarin speaking young children: From 14months to 32 months*. Unpublished doctoral dissertation, The University of Hong Kong.