Measuring the Influence of L1 on Learner English Errors in Content Words within Word Embedding Models

Kanishka Misra (kmisra@purdue.edu)
Hemanth Devarapalli (hdevarap@purdue.edu)
Julia Taylor Rayz (jtaylor1@purdue.edu)
Purdue University
West Lafayette, IN 47906 USA

Abstract

Recent works in Second Language Acquisition Literature and Corpus Linguistics have shown the interference of a person's first language (L1) when they process words in a new language. In this work, we build on the findings in two recent studies that explore the various differences in the lexico-semantic models of a person's L1 and L2 (English in their case), and test their hypotheses within the framework of two popular word vector models. This test is carried out by extracting erroneous content word errors from an annotated corpus of essays written by learners of English who belong to 16 different first languages. Specifically, we compare the vectors representations of the incorrect and correct-replacement word pairs in English as well as in the person's first language and find a moderate correlation between L1 and English. Additionally, we find certain inconsistencies between the two word embedding models when observed under the radar of language typology, suggesting new avenues for future work.

Keywords: L1 influence on L2; Natural Language Processing; Semantic Overlaps between L1 and L2

Introduction

While writing in a non-native language, people often make wrong word choices. For example, French speakers often use *scene* in place of *stage* when writing in English. Observations such as these are often a result of a transfer of properties from the speakers Native Language (L1) during their Second Language (L2) acquisition. In this paper, we investigate whether models for distributed representations of words capture this transfer of L1 semantic knowledge based on the errors made by learners of English; and if they do, whether the observations are similar to results from previously conducted experiments.

Patterns of lexical choice in content produced by non-native speakers have been widely studied by Second Language Acquisition (SLA) and Natural Language Processing (NLP) researchers. It has been shown that a person's native language L1 influences their L2 acquisition in morphological, phonological, syntactical and semantic aspects (Groot, 1992; Koda, 1993; De Groot & Keijzer, 2000; Hopman, Thompson, Austerweil, & Lupyan, 2018). The semantic influence of L1 over L2 has been studied by SLA researchers in behavioral studies (Prior, MacWhinney, & Kroll, 2007; Degani & Tokowicz, 2010; Bracken, Degani, Eddington, & Tokowicz, 2017) as well as corpus analysis (Gilquin, Granger, et al., 2011). Within NLP, errors in lexical choice have been analyzed based on their detection or correction (Ng et al., 2014; Rozovskaya & Roth, 2010, 2011; Chang, Chang, Chen, & Liou, 2008; Futagi, Deane, Chodorow, & Tetreault, 2008; Dahlmeier & Ng, 2011).

Word Choice by Second Language Acquisition Research

Degani and Tokowicz (2010) found that translation ambiguity occurs when there is an indirect mapping between translations of a word. Earlier works in SLA have highlighted the role of cross-lingual translation and semantic ambiguity in L2 acquisition. In an experiment with word translations from 40 English and Spanish bilinguals, Prior et al. (2007) found that the overlap between the words across the two languages was highly correlated with the translation choices made by the bilinguals. This was further confirmed by Boada, Sánchez-Casas, Gavilán, García-Albea, and Tokowicz (2013), where the presence of translation ambiguity proved to be challenging to recognize words for Spanish and Catalan bilinguals, as compared to when words only had one translation in the L2. A more recent study by Bracken et al. (2017) introduced a new metric known as Translation Semantic Variability (TSV) that measures the meaning similarity between translations, as conducted by participants who were trained to translate German-English word pairs. The TSV was found to be a predictor in measuring the learning of translation-ambiguous German words, i.e., the accuracy of learning fell when the relatedness between the German and English word was low (Bracken et al., 2017), further highlighting the importance of ambiguity in early acquisition of an L2.

Word Choice in Corpus Analysis

The influence of L1 on errors in lexical choice in learner corpora has been studied based on functional words as well as combinations of content words. Rozovskaya and Roth (2010, 2011) improved on correcting errors in preposition usage made by learners of English by inducing error-probabilities made by learners in their L1 from external corpora. Siyanova and Schmidt (Siyanova & Schmitt, 2008) showed that learning of content word combinations and collocations has also been shown to be a challenging task for non-native speakers of English. Chang et al. (2008) introduced a system to detect and correct mis-collocations of words in English content produced by Chinese speakers. Their system benefited from consulting parallel English-Chinese collocation dictionaries.

More recently, Kochmar and Shutova (2016, 2017) analyzed the L1 effects on L2 semantic knowledge using three types of content word combinations (Adjective-Noun, Verb-Direct Object, and Subject-Verb). They addressed L2 acquisition across a spectrum of proficiency, as well as within different language families of the learner L1s. We are interested in three hypotheses (out of five) that were tested in these papers: (1) L1 lexico-semantic models influence lexical choice in L2; (2) L1 lexico-semantic models are portable to other typologically similar languages; (3) typological similarity between L1 and L2 facilitates semantic acquisition of knowledge in L2. For hypothesis (1), it was found that semantic models of lexical choice derived from a learners L1 helped in improving error detection in the content word combinations. This improvement was also observed in the case of errors made by learners belonging to typologically similar L1s, as hypothesized by (2). Additionally, within language typology (hypothesis (3)), lexical distributions of content word combinations were found to be closer to native English for distant L1s, as compared to closer L1s. This contradicted the authors original assumptions that Germanic L1s would be closest to Native English. In particular, their experiment showed that the lexical distributions of Romance L1s and Asian L1s were closer to that of Native English, as compared to that of Germanic L1s. The authors speculated that this result was due to (1) the usage of prefabricated word combinations by speakers of typologically different L1s, which makes their distribution more native-like, and (2) the adventurous experimentation carried out by proficient speakers, especially observed among those that speak languages closer to English, where new (although incorrect) expressions are created.

Word Embeddings

Recent research within NLP has seen the emergence of neural network-based models of distributed word representations, also called word embeddings. Neural word embeddings were first introduced by Bengio, Ducharme, Vincent, and Jauvin (2003) and, after their reemergence due to the popularity of word2vec (Mikolov, Chen, Corrado, & Dean, 2013), have become an integral part of NLP research (Bojanowski, Grave, Joulin, & Mikolov, 2016). These word representations have found to capture semantic information of words by treating words as multi-dimensional vectors, such that words with similar contexts have similar vectors. Recent development in the intrinsic evaluation of these embeddings have highlighted their competent performance in comparison to human judgments. Specifically, word embeddings have achieved high correlation to humans in tasks involving the judgment of semantic similarity and relatedness between words such as WS-353 (Finkelstein et al., 2002), MEN (Bruni, Boleda, Baroni, & Tran, 2012), SimLex-999 (Hill, Reichart, & Korhonen, 2015). Word embeddings also exhibit the capability to solve verbal analogies, for example, king - man + woman = queen, which has attracted the attention of the Cognitive Science community. A recent study (Chen, Peterson, & Griffiths, 2017) analyzed two popular word embedding models, GloVe (Pennington, Socher, & Manning, 2014) and word2vec (Mikolov et al., 2013), as accounts of analogy to evaluate their performance in a relational similarity task. Chen et al. (2017) showed that the models capture certain forms of similarities more than others. Word embeddings have been used in SLA literature as well. Word embedding based similarity measures were successful in predicting L2 word learning accuracy (Hopman et al., 2018). Vector representations of words have been successful in improving error detection on learner corpus of essays (Kochmar & Shutova, 2016). Since word embedding models have been shown to capture certain semantic properties observed in language, we explore whether they capture patterns that were found by earlier work in the analysis of content word errors made by learners of English. Specifically, we explore the relationship of word errors in L2 and the learners L1 using distributed representations of words, following Kochmar and Shutova (2016, 2017). We are interested in the following questions:

1. Do distributed representations of words reflect L1 influence on learner English error words?
2. Does distributed representation of learner English error words exhibit similar relationships between typologically similar languages?

In order to approximate the extent of influence of L1, as represented by word embeddings, we take the incorrect-correct pairs in their present state (English), and compare them with their translated form in the learners' first language (L1). The influence is approximated by correlation between the closeness of the incorrect and correct words in each of the languages embedding spaces, i.e., a positive correlation might indicate some signal showing influence of L1 on the errors made in English. We compute the closeness of the incorrect and correct words based on their vector space neighborhood. Given the various word vectors, cosine similarities offer a good way to calculate a word's nearest neighbors, these represent words that are most related to the word (Hill et al., 2015). We assume that the closer two neighbors are in the L1 space, the easier they are to confuse in a typologically close L2 space. We introduce a metric that measures the closeness and using correlation between the closeness in L1 and L2, approximate a possible influence.

Methodology

In order to answer the questions presented above, we use an error annotated corpus where the errors are made by people whose native language is different from English. We use the Cambridge - First Certification in English (FCE) corpus (Yannakoudakis, Briscoe, & Medlock, 2011) which is a small subset of the Cambridge Learner Corpus (Nicholls, 2003). The FCE examination falls under the B2 proficiency category of the Common European Framework of Reference for Languages (CEFR). In the CEFR framework, language proficiency is organized in 6 categories, ranging from A1 (lowest) to C2 (highest). The FCE corpus contains error annotated short essay responses by learners of English taking the First Certification in English examination. There are 16 different L1 backgrounds represented in the 2488 different short essays. The errors in the corpus are annotated, including the linguistic information such as the type of error and the part of speech involved in the correction, as well as the correct replacement. The annotation follows the scheme provided by (Nicholls, 2003). We chose this corpus because it is the only freely available corpus for learner English with error annotations and suggested replacements.

We only consider the annotations involving a replacement of a content word. Based on the annotation scheme, the replacement category for content word errors have been labelled as RX where X indicates the part of speech of the word in that context. For the purposes of this research, only Nouns (N), Adjectives (J), Verbs (V), and Adverbs (Y) have been considered as content words. Furthermore, we ignore the semantic errors containing multi-word expressions or phrases, or errors counted as replacements but also containing misspellings. The incorrect-correct content word pairs were extracted based on the given criteria, resulting in a total of 5521 cases of incorrect usage of a content word, and its replacement.

Translation of Error Pairs into L1

Since each of the essays contained learners L1, the extracted incorrect words as well as the corrected words suggested by annotation we will refer to then as incorrect and correct word pairs were translated from English (L2) into the learners L1 using the Microsoft Azure Text Translator API. This was used in place of the widely used (for instance, in Hopman et al. (2018)) off-the-shelf Google Cloud Translator API, since the latter only provides one-to-one word translations, without providing much choice about the part of speech, or the confidence with which it predicts a certain translation, both of which were available in the Azure API. Translations that resulted in word utterances rather than a single word, as well as errors made by Dutch L1 speakers (only 5 cases) were discarded, resulting in a total of 4932 incorrect and correct word pairs (known as L1 and L2 pairs respectively, hereafter). Table 1 describes the number of semantic error cases for the various L1s used in the experiment.

Table 1: Number of Error Cases per language (L1).

L1	***n***	**L1**	***n***
Spanish	796	German	285
French	794	Portuguese	284
Greek	353	Turkish	272
Russian	340	Japanese	192
Italian	335	Korean	185
Catalan	325	Thai	122
Chinese (Simplified)	310	Swedish	44
Polish	295		

Distributed Representation of Words

Word embeddings provide mapping between words and their vectors in a multi-dimensional space, such that the semantic properties of the words are preserved. Since our final selection consists of content word-based errors and has a multilingual element to it, we use embeddings trained on corpora in multiple languages. Moreover, we compare different models that were produced using different parameters and different corpora. Specifically, we use:

1. **polyglot:** a word representation with embeddings for over 100 languages (Al-Rfou, Perozzi, & Skiena, 2013). This embedding learns a 64-dimension vector for each word by scoring the word's surrounding context, and a corrupted context (the selected word swapped out randomly).

2. **fasttext:** a word representation with embeddings for over 100 languages (Bojanowski et. al, 2016). In fasttext, each word vector is composed by summing up vectors of the subwords of the word (specifically, 3-6 character ngrams) and is trained using skipgrams along with negative sampling.

Error Pair Neighbor Overlap

To measure the differences between the incorrect and correct word in a given language, the semantic properties of their vectors in the distributed vector space are taken into account.

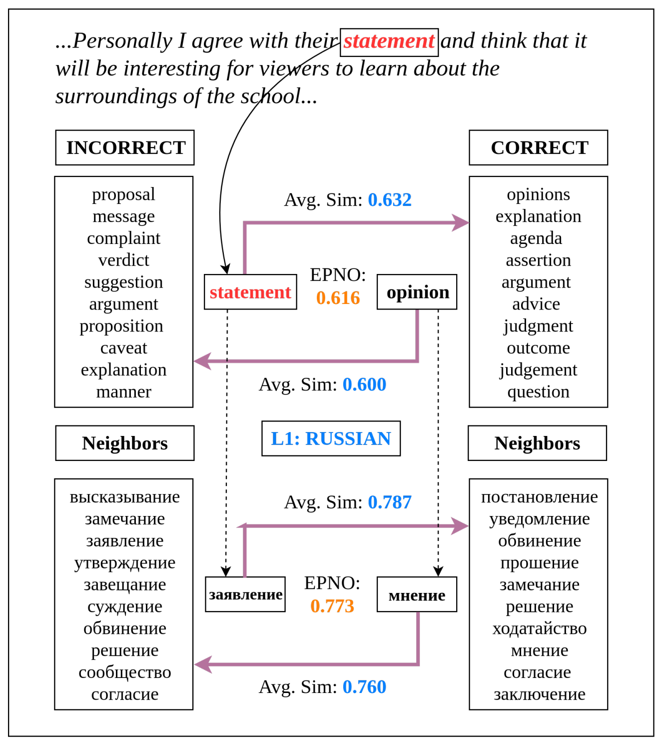

Figure 1: Visual Depiction of Computing EPNOs for (i, c) pairs in English and the person's L1 (Russian in this case). The context line is provided along with all the neighbors of the words.

More formally, given the incorrect-correct word pair, (i, c), the semantic overlap between i and c is computed. We introduced the Error Pair Neighbor Overlap (EPNO) to quantify the semantic relatedness between the incorrect word and correct word in terms of their nearest neighbors in the vector space, by relying on the idea that if the two words have a high semantic overlap, they will have related neighboring vectors. Mathematically, the *EPNO* for words i and c in language L is computed as:

$$EPNO_L(i,c) = \frac{1}{2k}[\sum_{c' \in NN_k^L(c)} cos(i,c') + \sum_{i' \in NN_k^L(i)} cos(c,i')] \quad (1)$$

where $NN_k^L(x)$ is a set of k nearest neighbors for word x in vector space for language L, and $cos(x,y)$ is the cosine similarity between vectors x and y. For our experiments, k is kept as 10. While cosine similarity shows a direct similarity between two vectors, EPNO computes the degree to which a given word (x) is related to words that are most similar to the second word (y),and vice-versa. Figure 1 shows a visual example of an error made by a native speaker of Russian, where EPNO values are calculated for *statement* (incorrect) and *opinion* (correct replacement) as well as for their respective Russian translations. The nearest neighbors, along with a context where the incorrect word occurs, are also provided for both cases.

Research Question 1

The first question that we would like to explore is: *Whether distributed representations of words reflect L1 influence on learner English error words.*

Experiment In order to approximate the influence of L1 on learner errors, EPNO values are computed for the L1 and the respective translated L2 word pairs over the fasttext as well as polyglot vector spaces. Japanese L1s were left out of the polyglot embeddings due to difficulty in feeding the text into the polyglot package. In order to check whether embeddings capture the L1 influence on learner English, the Spearman's Rank Correlation Statistic (ρ) between the overlaps in English as well as the L1 pairs was computed. Spearman's ρ calculates the monotonic relationship between the two variables. A significant correlation between the overlaps sustained across languages would indicate a potential role of L1 in influencing errors made by the learner. To test the significance for ρ for different languages, the p-values are computed along with the 95% bootstrap confidence intervals over 1000 resamples for each language. The resulting correlation estimates between the overlaps along with their p-values are shown in Table 2, while the bootstrap confidence intervals are shown in Figure 2.

Results As can be seen from Table 2 and Figure 2, the fasttext and polyglot EPNOs between L1 and English incorrect-correct word pairs have a moderately positive Spearman's ρ. In the case of Polyglot, errors committed by learners who speak Thai had a non-significant negative correlation, the rest (apart from Japanese L1) showed a significant correlation estimate between L1 and English. All languages within fasttext had significant positive correlations overall ($p < 10^{-3}$).

Discussion The results demonstrate a significant positive relationship between the EPNOs of error word pairs in English and the learners L1 for almost all languages, with the exceptions of Thai (non-significant) and Japanese (not included) in the case of Polyglot. A significant positive correlation shows that the incorrect-correct word pairs that are highly overlapping with each other in a person's L1 also highly overlap in English, indicating equal strength between the similarities in L1 and L2. These observations are consistent with findings reported by Kochmar and Shutova (2016), where L2 error detection accuracy improved when L1 lexico-semantic models were used as predictors, where their model showed improvement in differentiating error words from correctly used ones.

Table 2: Spearman's ρ between L1 and L2 overlaps in the error word pairs for fasttext and polyglot embeddings.

L1	fasttext	polyglot
Catalan	0.403 (<.001)	0.312 (<.001)
Chinese (Simplified)	0.588 (<.001)	0.322 (<.001)
French	0.477 (<.001)	0.373 (<.001)
German	0.505 (<.001)	0.384 (<.001)
Greek	0.489 (<.001)	0.351 (<.001)
Italian	0.565 (<.001)	0.355 (<.001)
Japanese	0.457 (<.001)	*NA*
Korean	0.366 (<.001)	0.281 (<.001)
Polish	0.546 (<.001)	0.356 (<.001)
Portuguese	0.543 (<.001)	0.369 (<.001)
Russian	0.552 (<.001)	0.129 (.025)
Spanish	0.539 (<.001)	0.351 (<.001)
Swedish	0.573 (<.001)	0.516 (<.001)
Thai	0.373 (<.001)	0.006 (.953)
Turkish	0.492 (<.001)	0.369 (<.001)

Note: Correlation Estimates and p values are listed as estimate (p-value)

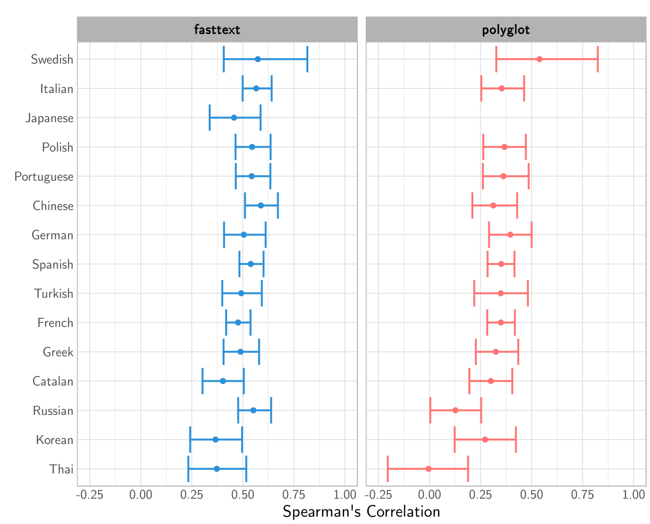

Figure 2: Spearman's ρ estimates of EPNOs computed for L1 and English incorrect-correct word pairs.

Research Question 2

The second question that we explore is: *Whether the similarity between semantic information of English and typologically closer L1s can be captured by fasttext and Polyglot.*

Experiment In this analysis, the same initial assumption made by Kochmar and Shutova (2017) was followed, i.e., L1s belonging to the same typological family will have similar EPNOs. For example, Germanic L1s should be closest to English based on their EPNO. The closeness with English is measured by the difference between the L1 and the English EPNO values computed for the fasttext and polyglot spaces. Based on our corpus, five groups of languages are considered: Germanic, Romance, Asian, Slavic, and an Other category to store the rest of the L1s. While we report the results, we will discard the Other category in the analysis since this combination is linguistically meaningless. The L1-English EPNO differences are computed as average differences over 1000 random samples (with replacement) within each group for 10,000 iterations. The notations $d_{fasttext}$ and $d_{Polyglot}$ denote these differences. Then, a one-way ANOVA is carried out to test for significance between the group L1-English differences. Table 3 lists the various languages covered in each group and their EPNO differences with English.

Results Table 3 reveals that for fasttext, the Asian family of languages in the corpus had the least difference between the EPNO values, followed by Slavic, Romance, and finally the Germanic. In contrast, for polyglot the differences observed for Germanic were the lowest, followed by Romance, Asian, and finally the Slavic. From the ANOVA results, the group L1-English differences were found to be significantly different from each other for both fasttext ($F(4, 49995) = 16539$, $p < 2 \times 10^{-16}$), and polyglot ($F(4, 49995) = 128751$, $p < 2 \times 10^{-16}$). A post-hoc Tukey HSD test revealed statistically significant pairwise difference between each of the groups except those with Slavic ($p = 0.131$).

Discussion The results observed in Table 3 reveal contrasting (although statistically significant) observations between differences in overlaps computed in fasttext and Polyglot. Based on the typology of languages, English falls under the Germanic family. However, the difference in the overlaps between the error pairs of Germanic L1s and English is the highest when computed for fasttext, with the least being the Asian L1s. In case of differences observed in the polyglot space, the opposite observation is made. The observations made in fasttext align with the findings of Kochmar and Shutova (2017), where Asian L1s were found to be closest to English in case of certain word pairs in the B2 proficiency category (same as our corpus), while Germanic L1s were found to be the farthest. On the other hand, the polyglot differences between L1 and English aligned with the initial assumptions made by Kochmar and Shutova (2017). The inconsistencies between fasttext and polyglot can be attributed to several factors. First, their dimension size and vocabulary: fasttext contains 300 dimensional vectors and an average vocabulary size is in the order of 10 million, while polyglot has 64 dimensional vectors with an average vocabulary size between 10,000 to 100,000. The difference in vocabulary size may dictate the choice in the neighbors for each overlap computation. Second, the nearest neighbors: fasttext incorporates the usage of subwords in its training along with the context of the words themselves, while polyglot follows only the contextual route. For example, the word *almost* has the following neighbors in fasttext: *nearly, practically, virtually, almsot, Almost, amost, alsmost, alomst, damn-near, pretty-much*; while in Polyglot: *nearly, once, roughly, just, equally, virtually, somewhat, less, absolutely, slightly*. The neighbors in fasttext could contribute to the noise while measuring the overlaps, thus distorting the results.

Table 3: Differences between L1 and English EPNOs for each Language Family in the Corpus.

Group	Languages	$d_{fasttext}$	$d_{Polyglot}$
Germanic	German Swedish	0.135	**0.184**
Romance	Spanish Catalan Italian French Portuguese	0.129	0.188
Slavic	Russian Polish	0.127	0.226
Asian	Chinese Japanese* Korean Thai	**0.123**	0.217
Other	Turkish Greek	0.128	0.195

* Japanese was ignored in the analysis of Polyglot. The bold formatted values highlight the minimum value in the respective column.

Conclusion

By analyzing content word errors in a corpus of learner English using two different word embedding models, we found (1) a significantly positive relationship between the error words in a learner's L1 and English, and that (2) while fasttext vector spaces emulate the results reported by Kochmar and Shutova (2016), the polyglot vector spaces are consistent with their initial assumptions. We speculate that the inconsistencies between fasttext and polyglot could be attributed to their inherent differences, namely: the dimensionality and vocabulary size, resulting in nearest neighbor choices. Due to the small size of the corpus, we unable to analyze the specific relationships within the different parts of speech used in the content word set, which could shed more light on the differences between the two embedding models.

References

Al-Rfou, R., Perozzi, B., & Skiena, S. (2013, July). Polyglot: Distributed Word Representations for Multilingual NLP. *arXiv:1307.1662 [cs]*.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, *3*(Feb), 1137–1155.

Boada, R., Sánchez-Casas, R., Gavilán, J. M., García-Albea, J. E., & Tokowicz, N. (2013). Effect of multiple translations and cognate status on translation recognition performance of balanced bilinguals. *Bilingualism: Language and Cognition*, *16*(1), 183–197.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016, July). Enriching Word Vectors with Subword Information. *arXiv:1607.04606 [cs]*.

Bracken, J., Degani, T., Eddington, C., & Tokowicz, N. (2017). Translation semantic variability: How semantic relatedness affects learning of translation-ambiguous words. *Bilingualism: Language and Cognition*, *20*(4), 783–794.

Bruni, E., Boleda, G., Baroni, M., & Tran, N.-K. (2012). Distributional semantics in technicolor. In *Proceedings of the 50th annual meeting of the association for computational linguistics: Long papers-volume 1* (pp. 136–145).

Chang, Y.-C., Chang, J. S., Chen, H.-J., & Liou, H.-C. (2008). An automatic collocation writing assistant for taiwanese efl learners: A case of corpus-based nlp technology. *Computer Assisted Language Learning*, *21*(3), 283–299.

Chen, D., Peterson, J. C., & Griffiths, T. L. (2017). Evaluating vector-space models of analogy. *arXiv preprint arXiv:1705.04416*.

Dahlmeier, D., & Ng, H. T. (2011). Correcting semantic collocation errors with l1-induced paraphrases. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 107–117).

Degani, T., & Tokowicz, N. (2010). Ambiguous words are harder to learn. *Bilingualism: Language and Cognition*, *13*(3), 299–314.

De Groot, A. M., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*(1), 1–56.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., et al. (2002). Placing search in context: The concept revisited. *ACM Transactions on information systems*, *20*(1), 116–131.

Futagi, Y., Deane, P., Chodorow, M., & Tetreault, J. (2008). A computational approach to detecting collocation errors in the writing of non-native speakers of english. *Computer Assisted Language Learning*, *21*(4), 353–367.

Gilquin, G., Granger, S., et al. (2011). From efl to esl: evidence from the international corpus of learner english.

Groot, A. M. de. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001.

Hill, F., Reichart, R., & Korhonen, A. (2015). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, *41*(4), 665–695.

Hopman, E. W. M., Thompson, B., Austerweil, J. L., & Lupyan, G. (2018). Predictors of L2 word learning accuracy: A big data investigation. In (p. 6).

Kochmar, E., & Shutova, E. (2016, August). Cross-Lingual Lexico-Semantic Transfer in Language Learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 974–983). Berlin, Germany: Association for Computational Linguistics.

Kochmar, E., & Shutova, E. (2017, September). Modelling semantic acquisition in second language learning. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 293–302). Copenhagen, Denmark: Association for Computational Linguistics.

Koda, K. (1993). Transferred l1 strategies and l2 syntactic structure in l2 sentence comprehension. *The Modern Language Journal*, *77*(4), 490–500.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, January). Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [cs]*.

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).

Nicholls, D. (2003). The Cambridge Learner Corpus - error coding and analysis for lexicography and ELT. , 10.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for english and spanish: The role of lexical variables, word class, and l2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, *39*(4), 1029–1038.

Rozovskaya, A., & Roth, D. (2010). Generating confusion sets for context-sensitive error correction. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 961–970).

Rozovskaya, A., & Roth, D. (2011). Algorithm selection and model adaptation for esl correction tasks. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies - volume 1* (pp. 924–933).

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429–458.

Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011). A New Dataset and Method for Automatically Grading ESOL Texts. , 10.