Using First and Second Language Models to Correct Preposition Errors in Second Language Authoring

Matthieu Hermet
School of Information Technology and Engineering
University of Ottawa
800, King Edward, Ottawa, Canada
mhermet@site.uottawa.ca

Alain Désilets
Institute for Information Technology

National Research Council of Canada
Bldg M-50, Montreal Road, Ottawa, K1A 0R6, Canada
alain.desilets@nrc-cnrc.gc.ca

Abstract

In this paper, we investigate a novel approach to correcting grammatical and lexical errors in texts written by second language authors. Contrary to previous approaches which tend to use unilingual models of the user's second language (L2), this new approach uses a simple roundtrip Machine Translation method which leverages information about both the author's first (L1) and second languages. We compare the repair rate of this roundtrip translation approach to that of an existing approach based on a unilingual L2 model with shallow syntactic pruning, on a series of preposition choice errors. We find no statistically significant difference between the two approaches, but find that a hybrid combination of both does perform significantly better than either one in isolation. Finally, we illustrate how the translation approach has the potential of repairing very complex errors which would be hard to treat without leveraging knowledge of the author's L1.

1 Introduction

In this paper, we investigate a novel approach to correcting grammatical and lexical errors in texts written by second language learners or authors. Contrary to previous approaches which tend to use unilingual models of the user's second language (L2), this new approach uses a translation model based on both the user's first (L1) and second languages. It has the advantage of being able to model linguistic interference phenomena, that is, errors which are produced through literal translation from the author's first language. Although we apply this method in the context of French-as-a-Second-Language, its principles are largely independent of language, and could also be extended to other classes of errors. Note that this is preliminary work which, in a first step, focuses on error correction, and ignores for now the preliminary step of error detection which is left for future research.

This work is of interest to applications in Computer-Assisted-Language-Learning (CALL) and Intelligent Tutoring Systems (ITS), where tutoring material often consists of drills such as fill-in-the-blanks or multiple-choice-questions. These require very little use of a learner's language production capacities, and in order to support richer free-text assessment capabilities, ITS systems thus need to use error detection and correction functionalities (Heift and Schulze, 2007).

Editing Aids (EA) are tools which assist a user in producing written compositions. They typically use rules for grammar checking as well as lexical heuristics to suggest stylistic tips, synonyms or fallacious collocations. Advanced examples of such

Proceedings of the NAACL HLT Workshop on Innovative Use of NLP for Building Educational Applications, pages 64–72, Boulder, Colorado, June 2009.

tools include Antidote[1] for French and StyleWriter[2] for English. Text Editors like MS Word and Word Perfect also include grammar checkers, but their style checking capabilities tend to be limited. All these tools can provide useful assistance to editing style, but they were not designed to assist with many errors found typically in highly non-idiomatic sentences produced by L2 authors.

Recent work in the field of error correction, especially as applied to English in the context of English as a Second Language (ESL), show an increasing use of corpora and language models. These have the advantage of offering a model of correctness based on common usage, independently of any meta-information on correctness. Corpus-based approaches are also able to correct higher level lexical-syntactic errors, such as the choice of preposition which is often semantically governed by other parts of the sentence.

The reminder of this paper is organized as follows. In section 2, we give a detailed account of the problem of preposition errors in a Second Language Learning (SLL) context. Related work is reviewed in section 3 and the algorithmic framework is presented in section 4. An evaluation is discussed in section 5, and conclusions and directions for future research are presented in section 6.

2 The Preposition Problem

Prepositions constitute 14% of all tokens produced in most languages (Fort & Guillaume 2007). They are reported as yielding among the highest error class rates across various languages (Izumi, 2004, for Japanese, Granger et al., 2001, for French). In their analysis of a small corpus of advanced-intermediate French as a Second Language (FSL) learners, Hermet et al. (2008) found that preposition choice accounted for 17.2 % of all errors. Prepositions can be seen as a special class of cognates, in the sense that the same L1 preposition used in different L1 sentences, could translate to several different L2 prepositions.

Automatic error detection/correction methods often process prepositions and determiners in the same way because they both fall in the class of function-words. However, one can make the argument that preposition errors deserve a different and deeper kind of treatment, because they tend to be more semantically motivated (event hough some prepositions governed by verbs draw a purely functional relation). In contrast, determiners are not semantically motivated and only vary on the register of quantity (or genre in some languages).

For example, there are 37 determiners in French, most of which can be used interchangeably without significantly affecting the syntax of a sentence, and often, not even its meaning ("I'll have one coffee"/"I'll have a coffee"/"I'll have some coffee"/"I'll have my coffee"/"I'll have coffee" are all rather alike). Comparatively, there are 85 simple prepositions and 222 compounds ones and they cannot be used interchangeably without significantly modifying the sense of an utterance, except for cases of synonymy.

In this paper, we focus our attention on preposition correction only, as it seems to be a more complex problem than determiners. While in principle the methods described here could handle determiner errors, we feel that our framework, which involves parsing in combination with a very large language model and Machine Translation, constitutes heavier machinery than is warranted for that simpler problem.

There are two major causes of preposition errors in a SLL context. The first kind is caused by lexical confusion within the second language itself. For example, a L2 author writing in English may erroneously use a location preposition like *"at"* where another location preposition like *"in"* would have been more appropriate. The second kind involves linguistic interference between prepositions in L1 and prepositions in L2 (Granger et al., 2001). For example, a Second Language Learner who wants to render the following two English sentences in French *"I go to Montreal"* and *"I go to Argentina",* might use the same French preposition *"à"* for *"to"*, when in fact, French usage dictates that you write *"à Montréal"*, and *"en Argentine"*. Note that the situation varies greatly from language to language. The same two English sentences rendered in Italian and German would in fact employ a same preposition, whereas in Spanish, different prepositions would also be required as in French.

[1] www.druide.com
[2] www.stylewriter-usa.com

Studies have found that the majority of errors made by L2 authors (especially intermediate to advanced ones) are caused by such linguistic interference (Wang and Garigliano, 1992, Cowan, 1983, p 109). Note that this kind of linguistic interference can often lead to much more severe and hard to repair errors, as illustrated by the following example, taken from an actual SLL corpus. Say a native English author wants to render *"Police arrived at the scene of the crime"* into French (her L2). Because she is not fluent in French, she translates the last part of the sentence to *"à la scène de la crime"*. This literal translation turns out to be highly unidiomatic in French, and should instead be written as *"sur les lieux du crime"* (which in English, would translate literally to *"on the location of the crime"*).

One might suspect that preposition errors of the first type would be solvable using unilingual L2 language models, but that the second type might benefit from a language model which also takes L1 into account. This is the main question investigated in this paper.

3 Related Work

Historically, grammatical error correction has been done through parsing-based techniques such as syntactic constraint-relaxation (L'haire & Vandeventer-Feltin, 2003), or mal-rules modeling (Schneider and McCoy, 1998). But generating the rule-bases needed by these types of approaches involves a lot of manual work, and may still in the end be too imprecise to convey information on the nature and solution of an error. Recently, more effort has been put in methods that rely on automatically built language models. Typically, this kind of work will focus either on a restricted class of errors or on specific domains. Seneff and Lee (2006) propose a two-phased generation-based framework where a n-gram model re-ranked by a stochastic context-free-grammar model is used to correct sentence-level errors in the language domain of flight reservation. Brockett et al. (2006) used a Brown noise channel translation model to record patterns of determiner error correction on a small set of mass-nouns, and reducing the error spectrum in both class and semantic domain, but adding detection capabilities. Note that although they use a translation model, it processes only text that is in one language. More specifically, the system learned to "translate" from poorly written English into correctly written English.

Chodorow et al. (2007) employed a maximum entropy model to estimate the probability of 34 prepositions based on 25 local context features ranging from words to NP/VP chunks. They use lemmatization as a means of generalization and trained their model over 7 million prepositional contexts, achieving results of 84% precision and 19% recall in preposition error detection in the best of the system's configurations. Gamon et al. (2008) worked on a similar approach using only tagged trigram left and right contexts: a model of prepositions uses serves to identify preposition errors and the Web provides examples of correct form. They evaluate their framework on the task of preposition identification and report results ranging from 74 to 45% precision on a set of 13 prepositions.

Yi et al. (2008) use the Web as corpus and send segments of sentences of varying length as bag-of-constituents queries to retrieve occurrence contexts. The number of the queried segments is a PoS condition of "check-points" sensitive to typical errors made by L2 authors. The contexts retrieved are in turn analyzed for correspondence with the original input. The detection and correction methods differ according to the class of the error. Determiner errors call for distinct detection and correction procedures while collocation errors use the same procedure for both. Determiner errors are discovered by thresholds ratios on search hits statistics, taking into account probable ambiguities, since multiple forms of determiners can be valid in a single context. Collocation errors on the other hand, are assessed only by a threshold on absolute counts, that is, a form different from the input automatically signals an error and provides its correction. This suggests that detection and correction procedures coincide when the error ceases to bear on a function word.

Similarly, Hermet et al. (2008) use a Web as corpus based approach to address the correction of preposition errors in a French-as-a-Second-Language (FSL) context. Candidate prepositions are substituted for erroneous ones following a taxonomy of semantic classes, which produces a set of alternate sentences for each error. The main interest of their study is the use of a syntax-based sentence generalization method to maximize the likelihood that at least one of the alternatives will have at least one hits on the Web. They achieve accuracy of 69% in error repair (no error detection), on a small set of clauses written by FSL Learners.

Very little work has been done to actually exploit knowledge of a L2 author's first language, in correcting errors. Several authors (Wang and Garigliano, 1992, Anderson, 1995, La Torre, 1999, Somers, 2001) have suggested that students may learn by analyzing erroneous sentences produced by a MT system, and reflecting on the probable cause of errors, especially in terms of interference between the two languages. In this context however, the MT system is used only to generate exercises, as opposed to helping the student find and correct errors in texts that he produces.

Although it is not based on an MT model, Wang and Garigliano propose an algorithm which uses a hand-crafted, domain-specific, mixed L1 and L2 grammar, in order to identify L1 interference errors in L2 sentences. L2 sentences are parsed with this mixed grammar, giving priority to L2 rules, and only employing L1 rules as a last resort. Parts of the sentence which required the user of L1 rules are labeled as errors caused by L1 interference. The paper does not present an actual evaluation of the algorithm.

Finally, a patent by Dymetman and Isabelle (2005) describes several ways in which MT technology could be used to correct L2 errors, but to our knowledge, none of them has been implemented and evaluated yet.

4 Algorithmic Framework

As discussed in section 2, L2 authoring errors can be caused by confusions within the L2 itself, or by linguistic interference between L1 and L2. In order to account for this duality, we investigate the use of two correction strategies, one which is based on unilingual models of L2, and one which is based on translation models between L1 and L2.

Input Sentence

Il y a une grande fenêtre qui permet au soleil <à> entrer

(there is a large window which lets the sun come in)

Syntactic Pruning and Lemmatization

permettre <à> entrer

(let come in)

Generation of alternate prepositions

semantically related: *dans, en, chez, sur, sous, au, dans, après, avant, en, vers*

most common: *de, avec, par, pour*

Query and sort alternative phrases

permettre d'entrer: 119 000 hits
permettre avant entrer: 12 hits
permettre à entrer: 4 hits
permettre en entrer: 2 hits

...

→ preposition <*d'*> is returned as correction

Figure 1. Typical processing carried out by the *Unilingual* approach.

The first approach, called the *Unilingual* strategy, is illustrated by the example in Figure 1. It uses a web search engine (Yahoo) as a simple, unilingual language model, where the probability of a L2 phrase is estimated simply by counting its number of occurrences in Web pages of that language. A severe limitation of this kind of model is that it can only estimate the probability of phrases that appear at least once on the Web. In contrast, an N-gram model (for example) is able to estimate the probability of phrases that it has never seen in the training corpus. In order to deal with this limitation, syntactic pruning is therefore applied to the phrase before it is sent to the search engine, in order to eliminate parts which are not core to the context of use of the preposition, thus increasing the odds that the pruned sentence will have at least one occurrence on the Web.

This pruning and generalization is done by carrying out syntactic analysis with the Xerox Incremental Parser for the syntactic analysis (Ref XIP). XIP is an error robust, symbolic, dependency parser, which outputs syntactic information at the constituency and dependency levels. Its ability to produce syntactic analyses in the presence of errors is

Category	Prepositions
Localization	*in front, behind, after, before, above, in, at, on, below, above...*
Temporal	*at, in, after, before, for, during, since...*
Cause	*for, because of*
Goal	*for, at*
Manner	*in, by, with, according to...*
Material	*in, of*
Possession/Relation	*to, at, with respect to...*
Most common	*to, at, on, with, by, for*

Table 1. Categories of prepositions – the list is given in English, and non exhaustive for space reasons.

particularly interesting in the context of second language authoring where the sentences produced by the authors can be quite far from grammatical correctness. The input sentence is fed to the parser as two segments split at error point (in this case, at the location of the erroneous preposition). This ensures that the parses are correct and not affected at dependency level by the presence of error. The syntactic analyses are needed to perform syntactic pruning, which is a crucial step in our framework, following Hermet et. al (2008). Pruning is performed by way of chunking heuristics, which are controlled by grammatical features, provided by XIP's morphological analysis (PoS tagger). The heuristics are designed to suppress syntactically extraneous material in the sentence, such as adverbs, some adjectives and some NPs. Adverbs are removed in all cases, while adjectives are only removed when they are not in a position to govern a Prepositional Phrase. NPs are suppressed in controlled cases, based on the verb sub-categorization frame, when a PP can be attached directly to the preceding verb. In case of ambiguity in the attachment of the PP, two versions of the pruned sentence can be produced reflecting two different PP attachments. Lemmatization of verbs is also carried out in the pruning step.

After pruning, the right and left sides of the sentences are re-assembled with alternate prepositions. The replacement of prepositions is controlled by way of semantics. Since prepositions are richer in sense than strict function words, they can therefore be categorized according to semantics. Saint-Dizier (2007) proposes such a taxonomy, and in our framework, prepositions have been grouped in 7 non-exclusive categories. Table 1 provides details of this categorization. The input preposition is mapped to all the sets it belongs to, and corresponding alternates are retrieved as correction candidates. The 6 most frequent French preposition are also added automatically to the candidates list.

The resulting sentences are then sent to the Yahoo Search Engine and hits are counted. The number of hits returned by each of the queries is used as decision criteria, and the preposition contained in the query with the most hits is selected as the correction candidate.

While the above *Unilingual* strategy might work for simple cases of L1 interference, one would not expect it to work as well in more complex cases where both the preposition and its governing parts have been translated too literally. For example, in the case of the example from section 2, while the *Unilingual* strategy might be able to effect correction *"**sur** la scène du crime"* which is marginally better than the original *"**à** la scène du crime"* (12K hits versus 1K), it cannot address the root of the problem, that is, the unidiomatic expression *"scène du crime"* which should instead be rendered as *"lieux du crime"* (38K hits). In this particular case, it is not really an issue because it so happens that *"sur"* is the correct preposition to use for both *"lieux du crime"* and *"scène du crime"*, but in our experience, that is not always the case. Note also that the *Unilingual* approach can only deal with preposition errors (although it would be easy enough to extend it to other kinds of function words), and cannot deal with more semantically deep L1 interference.

To address these issues, we experimented with a second strategy which we will refer to as the *Roundtrip Machine Translation* approach (or *Roundtrip MT* for short). Note that our approach is different from that of Brockett et al. (2006), as we do make use of a truly multi-lingual translation model. In contrast, Brockett's translation model was trained on texts that were written in the same language, with the sources being ill-written text in the same language as the properly-formed target texts. One drawback of our approach however is that it may require different translation models for speakers with different first languages.

There are many ways in which error-correction could be carried out using MT techniques. Several of these have been described in a patent by Dymetman and Isabelle (2005), but to our knowledge, none of them have yet been implemented and evaluated. In this paper, we use the simplest possible implementation of this concept, namely, we carry out a single round-trip translation. Given a potentially erroneous L2 sentence written by a second language author, we translate it to the author's L1 language, and then back to L2. Even with this simple approach, we often find that errors which were present in the original L2 sentence have been repaired in the roundtrip version. This may sound surprising, since one would expect the roundtrip sentence to be worse than the original, on account of the "Chinese Whisper" effect. Our current theory for why this is not the case in practice goes as follows. In the course of translating the original L2 sentence to L1, when the MT system encounters a part that is ill-formed, it will tend to use single word entries from its phrase table, because longer phrases will not have been represented in the well-formed L2 training data. In other words, the system tends to generate a word for word translation of ill-formed parts, which mirrors exactly what L2 authors do when they write poorly formed L2 sentences by translating too literally from their L1 thought. As a result, the L1 sentence produced by the MT system is often well formed for that language. Subsequently, when the MT system tries to translate that well-formed L1 sentence back to L2, it is therefore able to use longer entries from its phrase table, and hence produce a better L2 translation of that part than what the author originally produced.

We use Google Translate as a translation engine for matter of simplicity. A drawback of using such an online service is that it is essentially a closed box, and we therefore have little control over the translation process, and no access to lower level data generated by the system in the course of translation (e.g. phrase alignments between source and target sentences). In particular, this means that we can only generate one alternative L2 sentence, and have no way of assessing which parts of this single alternative have a high probability of being better than their corresponding parts in the original L2 sentence written by the author. In other words, we have no way of telling which changes are likely to be false positives, and which changes are likely to be true positives. This is the main reason why we focus only on error repair in this preliminary work.

The roundtrip sentences generated with Google Translate often differ significantly from the original L2 sentence, and in more ways than just the erroneous preposition used by the author. For example, the (pruned) clause *"avoir du succès **en** le recrutement" ("to be successful **in** recruiting"*) might come back as as *"réussir **à** recruter" ("to succeed **in** recruiting"*). Here, the translation is acceptable, but the preposition used by the MT system is not appropriate for use in the original sentence as written by the L2 author. Conversely, a roundtrip translation can be ill-formed, yet use a preposition which would be correct in the original L2 sentence. For example, *"regarder **à** des films" ("look at some movies"*) might come back as *"inspecter des films" ("inspect some films"*). Here, the original meaning is somewhat lost, but the system correctly suggested that there should be no preposition before *"des films"*.

Hence, in the context of the *Roundtrip MT* approach, we need two ways of measuring appropriateness of the suggested corrections for given clauses. The first approach, which we call the *Clause* criteria, looks at whether or not the whole clause has been restored to a correct idiomatic form (including correct use of preposition) which also preserves the meaning intended by the author of the original sentence. Hence, according to this approach, an MT alternative may be deemed correct, even if it chooses a preposition which would have been incorrect if substituted in the original L2 sentence as is. In the second approach, called the *Prep* criteria, we only look at whether the preposition used by the MT system in the roundtrip translation, corresponds to the correct preposition to be used in the original L2 clause. Hence, with this approach, an MT alternative may be deemed correct, even if the preposition chosen by the MT system is actually inappropriate in the context of the generated roundtrip translation, or, even worse, if the roundtrip modified the clause to a point where it actually means something different than what the author actually intended.

Of course, in the case of the *Prep* evaluation criteria, having the MT system return a sentence which employs the proper preposition to use in the context of the original L2 sentence is not the end of the process. In an error correction context, one must also isolate the correct preposition and insert it in the appropriate place in the original L2 sentence. This part of the processing chain is not currently implemented, but would be easy to do, if we used an MT system that provided us with the alignment information between the source sentence and the target sentence generated. The accuracy figures which we present in this paper assume that this mapping has been implemented and that this particular part of the process can be done with 100% accuracy (a claim which, while plausible, still needs to be demonstrated in future work).

We also investigate a third strategy called *Hybrid*, which uses the *Roundtrip MT* approach as a backup for cases where the *Unilingual* approach is unable to distinguish between different choices of preposition. The latter typically occurs when the system is not able to sufficiently prune and generalize the phrase, resulting in a situation where all pruned variants yield zero hits on the Web, no matter what preposition is used. One could of course also use the *Unilingual* approach as a backup for the *Roundtrip MT* approach, but this would be harder to implement since the MT system always returns an answer, and our use of the online Google Translate system precludes any attempt to estimate the confidence level of that answer.

In conclusion to this section, we use three preposition correction strategies: *Unilingual*, *Roundtrip MT* and *Hybrid*, and in the case of the *Roundtrip MT* approach, appropriateness of the corrections can be evaluated using two criteria: *Prep* and *Clause*.

5 Evaluation and Results

5.1 Corpus and Evaluation Metric

For evaluation, we extracted clauses containing preposition errors from a small corpus of texts written by advanced-intermediate French as a Second Language (FSL) student in the course of one semester. The corpus contained about 50, 000 words and 133 unique preposition errors. While relatively small, we believe this set to be sufficiently rich to test the approach. Most clauses also presented other errors, including orthographic, tense, agreement, morphologic and auxiliary errors, of which only the last two affect parsing. The clauses were fed as is to the correction algorithms, without first fixing the other types of errors. But to our surprise, XIP's robust parsing has proven resistant in that it produced enough information to enable correct pruning based on chunking information, and we report no pruning errors. Chodorow et al. (2008) stress the importance of agreement between annotators when retrieving or correcting preposition errors. In our case, our policy has been to only retain errors reported by both authors of this paper, and correction of these errors has raised little matter of dispute.

Algorithm	Repair rate (%)
Unilingual	68.7
Roundtrip MT (Clause)	44.8
Roundtrip MT (Prep)	66.4
Hybrid (Prep)	82.1

Table 2. Results for 3 algorithms on 133 sentences.

We evaluated the various algorithms in terms of repair rate, that is, the percentage of times that the algorithm proposed an appropriate fix (the absence of a suggestion was taken to be an inappropriate fix). These figures are reported in Table 2.

5.2 Discussion

ANOVA of the data summarized in Table 2 reveals a statistically significant ($p < 0.001$) effect of the algorithm on repair rate. Although *Roundtrip MT* performed slightly worse than *Unilingual* (66.4% versus 68.7%), this difference was not found to be statistically significant. On one hand, we found that round-trip translation sometimes result in spectacular restorations of long and clumsy phrases caused by complex linguistic interference. However, too often the Chinese whispers effect destroyed the sense of the original phrase, resulting in inappropriate suggestions. This is evidenced by the fact that repair rate of the *Roundtrip MT* approach was significantly lower ($p < 0.001$) when using the

Clause criteria (44.8%) than when using the *Prep* criteria (66.4%). It seems that, in the case of preposition correction, roundtrip MT is best used as a way to to generate an L2 alternative from which to mine the correct preposition. Indeed, flawed as they are, these distorted roundtrip segments corrected prepositions errors in 66.4% of the cases. However, for a full picture, the approach should be tried on more data, and on other classes of errors. Particularly, we currently lack sufficient data to test the hypothesis that the approach could address the correction of more complex literal translations by SL Learners.

In the *Unilingual* approach, the Yahoo Web search engine proved to be an insufficient language model for 31 cases out of 133, meaning that even the pruned and generalized phrases got zero hits, no matter what alternative preposition was used. In those cases, the *Hybrid* approach would then attempt correction using *MT Roundtrip* approach. This turned out to work quite well, since it resulted in an overall accuracy of 82.1%. ANOVA on the data for *Hybrid* and the two pure approaches reveals a significant effect ($p < 0.001$) of the algorithm factor. Individual t-tests between the *Hybrid* approach and each of the two pure approaches also reveal statistically significant differences ($p < 0.001$). The improvements provided by the hybrid approach are fairly substantial, and represent relative gains of 19.5% over the pure *Unilingual* approach, and 23.6% over the pure *Roundtrip MT* approach. The success of this combined approach might be attributable to the fact that the two approaches follow different paradigms. *Roundtrip MT* uses a model of controlled incorrectness (errors of anglicism) and *Unilingual* a model of correctness (occurrences of correct forms). In this respect, the relatively low agreement between the two approaches (65.4%) is not surprising.

6 Conclusion and Future Work

In this paper, we have demonstrated for the first time that a bilingual Machine Translation approach can be used to good effect to correct errors in texts written by Second Language Authors or Learners. In the case of preposition error correction we found that, while the MT approach on its own did not perform significantly better than a unilingual approach, a hybrid combination of both performed much better than the unilingual approach alone. More work needs to be carried out in order to fully evaluate the potential of the MT approach. In particular, we plan to experiment with this kind of approach to deal with more complex cases of L1 interference which result in severely damaged L2 sentences.

In this paper, we compared the bilingual MT approach to a unilingual baseline which used a relatively simple Web as a corpus algorithm, whose accuracy is comparable to that reported in the literature for a similar preposition correction algorithm (Yi et al, 2008). Notwithstanding the fact that such simple Web as a corpus approaches have often been shown to be competitive with (if not better than) more complex algorithms which cannot leverage the full extent of the web (Halevy et al., 2009), it would be interesting to compare the bilingual MT approach to more sophisticated unilingual algorithms for preposition correction, many of which are referenced in section 3.

Error *detection* is another area for future research. In this paper, we limited ourselves to error correction, since it could be solved through a very simple round-trip translation, without requiring a detailed control of the MT system, or access to lower level information generated by the system in the course of translation (for example, intermediate hypotheses with probabilities and alignment information between source and target sentences). In contrast, we believe that error detection with an MT approach will require this kind of finer control and access to the guts of the MT system. We plan to investigate this using the PORTAGE MT system (Ueffing et al., 2007). Essentially, we plan to use the MT system's internal information to assign confidence scores to various segments of the roundtrip translation, and label them as corrections if this confidence is above a certain threshold. In doing this, we will be following in the footsteps of Yi et al. (2008) who use the same algorithm for error detection and error correction. The process of detecting an error is simply one of determining whether the system's topmost alternative is different from what appeared in the original sentence, and whether the system's confidence in that alternative is sufficiently high to take the risk of presenting it to the user as a suggested correction.

Acknowledgments

The authors are indebted to the following people (all from NRC) for helpful advice on how to best exploit MT for second language correction: Pierre Isabelle, George Foster and Eric Joanis.

References

Anderson, D. D. 1995. *Machine Translation As a Tool in Second Language Learning*. CALICO Journal, v13 n1 p68-97.

Brockett C., Dolan W. B., and Gamon M.. 2006. *Correcting ESL errors using phrasal SMT techniques*. In Proc. 21st International Conf. On Computational Linguistics and the 44th annual meeting of the ACL, p. 249–256, Sydney, Australia.

Chodorow M., Tetreault J. R. and Han N.-R.. 2007. *Detection of Grammatical Errors Involving Prepositions*. In Proc. ACL-SIGSEM Workshop on Prepositions. Prague, Czech Republic.

Cowan, J. R. 1983. *Towards a Psychological Theory of Interference in Second Language Learning*. In Second Language Learning: Contrastive Analysis, Error Analysis, and Related Aspects, edited by B. W. Robinett, J. Schachter, pp 109-119, The Univ. of Michigan Press.

Dymetman M., Isabelle, P. 2007. *Second language writing advisor*. US Patent #20070033002 , Feb 8, 2007.

Fort K., Guillaume B. 2007. *PrepLex: un lexique des prépositions du français pour l'analyse syntaxique*. TALN 2007, Toulouse, June 5-8.

Gamon M., Gao J. F., Brockett C., Klementiev A., Dolan W. B., and Vanderwende L. 2008. *Using contextual speller techniques and language modeling for ESL error correction*. In Proceedings of IJCNLP 2008, Hyderabad, India, January.

Granger S., Vandeventer A. & Hamel M. J. 2001. *Analyse de corpus d'apprenants pour l'ELAO basé sur le TAL*. TAL 42(2), 609-621.

Halevy, A., Norvig, P., Pereira, F. 2009. *"The Unreasonable Effectiveness of Data."*, IEEE Intelligent Systems, March/April 2009, pp 8-12.

Heift, T. & Schulze, M. 2007. *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*. Routledge.

Hermet, M., Désilets, A., Szpakowicz, S. 2008. *Using the Web as a Linguistic Resource to Automatically Correct Lexico-Syntactic Errors*. In Proceedings of the LREC'08. Marrakech, Morroco.

Izumi, E., K. Uchimoto, and H. Isahara. 2004. *The overview of the sst speech corpus of Japanese learner English and evaluation through the experiment on automatic detection of learners' errors*. In LREC.

La Torre, M. D. 1999. *A web-based resource to imprive translation skills*. ReCALL, Vol 11, No3, pp. 41-49.

Lee J. and Seneff S. 2006. *Automatic grammar correction for second-language learners*. In Interspeech. ICSLP. p. 1978-1981. Pittsburgh.

L'haire S. & Vandeventer Faltin A. 2003. *Error diagnosis in the FreeText project*. CALICO 20(3), 481-495, special Issue Error Analysis and Error Correction in Computer-Assisted Language Learning, T. Heift & M. Schulze (eds.).

Schneider, D. and McCoy, K. F. 1998. *Recognizing syntactic errors in the writing of second language learners*. In Proceedings of COLING/ACL 98.

Saint-Dizier, P. 2007. *Regroupement des Prépositions par sens*. Undated Report. IRIT. Toulouse. http://www.irit.fr/recherches/ILPL/Site-Equipe/publi_fichier/prepLCS.doc

Somers, Harold. 2001. *Three Perspectives on MT in the Classroom*, MT SUMMIT VIII Workshop on Teaching Machine Translation, Santiago de Compostela, pages 25-29.

Tetreault J. and Chodorow M. 2008. *The Ups and Downs of Preposition Error Detection*. *COLING*, Manchester.

Ueffing, N., Simard, M., Larkin, S., Johnson, J. H. (2007), NRC's PORTAGE system for WMT 2007, ACL-2007 Workshop on SMT, Prague, Czech Republic 2007.

Wang, Y. and Garigliano, R. 1992. *An Intelligent Language Tutoring System for Handling Errors caused by Transfer*. In Proceedings of ITS-92, pp. 395-404.

Yi X., Gao J. F., Dolan W. B., 2008. *A Web-based English Proofing System for English as a Second Language Users*. In Proceedings of IJCNLP 2008, Hyderabad, India, January.