



## NRC Publications Archive Archives des publications du CNRC

### Using Automatic Roundtrip Translation to Repair General Errors in Second Language Writing

Désilets, Alain; Hermet, Matthieu

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. /  
La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version  
acceptée du manuscrit ou la version de l'éditeur.

#### **Publisher's version / Version de l'éditeur:**

*Proceedings of the twelfth Machine Translation Summit, (MT Summit XII), 2009,  
2009-08-29*

#### **NRC Publications Record / Notice d'Archives des publications de CNRC:**

<https://nrc-publications.canada.ca/eng/view/object/?id=ea005cc4-0eca-4ae5-80a9-a1ac32fb64fe>  
<https://publications-cnrc.canada.ca/fra/voir/objet/?id=ea005cc4-0eca-4ae5-80a9-a1ac32fb64fe>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

**Questions?** Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the  
first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la  
première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez  
pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



# Using Automatic Roundtrip Translation to Repair General Errors in Second Language Writing

**Alain Désilets**

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

**Matthieu Hermet**

University of Ottawa  
800, King Edward  
Ottawa, Ont, Canada  
mhermet@site.uottawa.ca

## Abstract

In this paper, we evaluate for the first time the use of Machine Translation technology to repair general errors in second language (L2) authoring. Contrary to previously evaluated approaches which rely exclusively on unilingual models of L2, this method takes into account both languages, and is thus able to model linguistic interference phenomena where the author produces an erroneous word for word translation of his L1 intent. We evaluate a simple roundtrip MT approach on a corpus of foreign-sounding errors produced in the context of French as a Second Language. We show that the roundtrip approach is better at repairing linguistic interference errors than non-interference ones, and that it is better at repairing errors which only involve function words. We also show that the first leg of the roundtrip (inferring the author's L1 intent) is more sensitive to error type and more error prone than the second leg (rendering a correct L1 intent back into L2).

## 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 most previous approaches which tend to use unilingual models of the learner's second language (L2), this new approach uses a bilingual translation model based on both the learner's first (L1) and second languages. It has the advantage of being able to model linguistic interference phe-

nomena, that is, errors which take their root in 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.

While there currently exist many Editing Aids which can assist a user in producing written compositions, very few of them target specifically the type of errors made by second language authors. These tools typically use rules for grammar checking as well as lexical heuristics to suggest stylistic tips, synonyms or fallacious collocations. Advanced examples of such tools include Antidote for French<sup>1</sup> and StyleWriter<sup>2</sup> for English. Text Editors like MS Word and Word Perfect also include grammar checkers, but their style checking capabilities tend to be limited. Most of the above tools were designed with native authors in mind, and do not deal well with errors found in foreign sounding sentences often produced by second language authors.

Recent work in the field of error correction, especially as applied to English in the context of English as a Second Language (ESL) and Computer Assisted Language Learning (CALL), show an increasing use of corpora and language models. The work presented in this paper is in a similar vein, as it is based on Statistical Machine Translation (MT) systems which are corpus-based.

The remainder of this paper is organized as follows. Section 2 discusses the problem of linguistic

---

<sup>1</sup> [www.druides.com](http://www.druides.com)

<sup>2</sup> [www.stylewriter-usa.com](http://www.stylewriter-usa.com)

interference and its influence on second language writing. Section 3 presents a review of related work. Section 4 discusses different ways in which MT could be used to help repair second language errors, and explains the simple roundtrip translation approach used in this paper. Sections 5 and 6 respectively describe the corpus and method used to evaluate performance of this roundtrip approach. Section 7 presents the results of this evaluation, while conclusions and directions for future work are discussed in Section 8.

## 2 The linguistic interference problem

It is widely accepted among linguists (Selinker, 1969; Sheen, 1980; Cowan, 1983) that a significant portion of the errors committed by Second Language students are caused by linguistic interference from the student's first language (L1). Some researchers have even observed empirically that the majority of errors made by second language students fall in that category (Wang and Garigliano, 1992, Cowan, 1983, p 109). Therefore, in the context of automatic correction of L2 errors, it is worth paying special attention to these so-called transfer errors. This type of error is also particularly interesting and challenging, because they often result in sentences which are perfectly grammatical, yet clearly sound foreign.

As pointed out by Wang and Garigliano (1992), there are many different types of transfer errors. A *lexical transfer* error is when the author improperly renders a content L2 word using one of its many possible translations in L1, but this turns out to be an inappropriate choice in the particular context of the L1 sentence he is writing. For example, a French as a Second Language (FSL) author might write “je vais *commencer* une famille” (“I will *start* a family”), where “*fonder*” (“to *found*”) should have been used instead of “*commencer*”.

A *syntactic transfer* error is when the author writes an L1 sentence which borrows improperly from a L2 grammatical or syntactic structure. For example, a FSL author might write “si *j'avais eu* donné du temps”, which has a one-to-one mapping with “if I had been given time”, when in fact, the proper French formulation would be “si on *m'avait* donné du temps”.

An *idiomatic transfer* error is when the author literally translates an idiom or collocation, resulting in a sentence which is either nonsensical or for-

eign sounding in L2. For example, a FSL author might write “*regarder pour* une maison” (“look for a house”), when in French, the collocation “look for” should be rendered as “*chercher*” (“search”, without a preposition).

To these three categories, we add a fourth which was not mentioned by Wang and Garigliano: *orthographic transfer* errors. This is when the author “invents” a word which does not actually exist in L2, by using an L2-ish orthography for a L1 word. For example, an FSL author might write “*construc*ter” in an attempt to render “to construct”, when in fact the proper French word is “*construire*”.

Note that the boundaries between these different categories are not crisply defined, and it is not always easy to decide which category a particular error fits in. For example, “*commencer* une famille” could be construed either as a lexical transfer error (bad choice for “start”), or an idiomatic transfer error (bad rendering of the collocation “start a family”). However, it is generally easy to determine whether or not an error is a transfer error (whatever its sub-category) or not. In this paper, we considered an error to be a transfer error if we could come up with a native sounding English sentence which had the correct meaning, and translated word for word to the L1 error.

Given that transfer errors are caused by literal translation of a L1 thought, it seems reasonable to try and leverage knowledge of the authors' L1 to repair them. In this paper, we evaluate for the first time how Statistical Machine Translation might be used to do this.

## 3 Related Work

Historically, grammatical error correction has been done through parsing-based techniques such as syntactic constraint-relaxation (L'haire & Vandeventer-Felton, 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. This is particularly true of work done in the context of Second Language learning or authoring, and Computer Assisted Language Learning (CALL).

Typically, this kind of work will either focus on controlled inputs (ex: a small number of predetermined sentences used as exercises in a CALL context), on a specific domain (ex: flight reservation), or on a specific class of errors which usually involves function words only (ex: determiners or prepositions).

Shei and Pain (2001) propose a unilingual method for correcting L2 collocation errors, where words in the collocation are substituted by synonyms taken from a dictionary, and the likelihood of these combinations is evaluated by looking up in library of collocations pre-compiled from a corpus. The paper does not actually evaluate the approach.

Lee and Seneff (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 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 yield at least one hit on the Web. They achieve accuracy of 69% in error repair (no error detection), on a small set of clauses written by FSL Learners.

There has also been some work on bilingual approaches, or use of MT in error correction settings. 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 students find and correct errors in texts that they produce.

Schuster (1986) describes a system which uses L1 information to correct L2 errors in verb-particle constructs (ex: "go over", "put on"). The system is designed to work only for controlled example sentences in a CALL context. It uses hand-crafted L2 grammar to identify where student's translation of those examples differ from a correct translation, with respect to verb-particle constructs. Direct-translation tables of the verbs and prepositions are then used to find the likely L1 construct, and information about the L1 verb is used to explain to the

student, the difference between that construct in L1 and L2. Performance of the system was not evaluated.

Although it is not based on an MT model, Wang and Garigliano (1992) 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 use of L1 rules are labeled as errors caused by L1 interference. The paper does not present an actual evaluation of the algorithm.

In a broad patent, Dymetman and Isabelle (2007) propose a range of methods for correcting single-word errors. The method computes the probability of different corruption paths, starting backward from a potentially incorrect L2 word written by the author, going to various L1 words which might have erroneously been rendered as that L2 word. It then goes back to L2, generating alternative renderings of those L1 words, and those with highest probability are suggested as potential repairs. To the best of our knowledge, no evaluation has been published for any embodiment of this general method.

Hermet and Désilets (2009) compare the performance of a web-as-corpus with shallow syntactical pruning, against that of a roundtrip MT approach, for repairing preposition errors. They also evaluate a hybrid approach where the roundtrip approach is used as a fallback for cases where the web-as-corpus approach produces no suggestions at all. While they found no significant difference between the repair rates of the first two approaches, they found the hybrid method to perform significantly better than either approach in isolation.

None of the work cited above on bilingual models tried to *evaluate* this type of approach in situations with *open-ended* input, domain, and error type all at once. To the best of our knowledge, this is a unique feature of the present paper.

#### 4 Roundtrip Translation as a Means for Second Language Error Repair

There are many ways in which MT could be used to correct L2 errors, many of which were first proposed in Dymetman and Isabelle (2007). For this

very first evaluation of this kind of approach, we choose the simplest possible implementation of the concept (roundtrip translation), and only apply it to error repair (but see section 7 for a discussion of future research on error detection).

Given an L2 sentence S2, for which we already know that specific words are erroneous, we carry out an automatic roundtrip translation to generate an alternative L2 sentence S2\*. The roundtrip is carried out in two separate steps, first producing the most probable L1 translation S1 of S2, then producing the most probable L2 translation S2\* of S1. We then follow the word alignments in the two legs of this path, in order to identify the words in S2\* which align with the erroneous words in S2. Those aligned words in the roundtrip sentence are presented as the correction. One can think of this approach as trying to reverse engineer the correct L1 intent behind the L2 error, and then trying to produce a better rendering of that intent into L2.

Note that one limitation of this simple approach is that it carries out the roundtrip in two steps, and may not return a path which maximizes the overall probability of S2\* given S2. This choice was made because of its ease of implementation. In particular, it allowed us to carry out the two legs of the roundtrip using the Google Translate service. One 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 have no way of assessing which parts of S2\* have a high probability of being better than their corresponding parts in the original S2. This is the main reason why we focus first on error repair and leave error detection as future work.

We have found this simple approach to be unexpectedly effective, often resulting in a S2\* which addresses the erroneous parts of S2. This is somewhat surprising, since one would expect the roundtrip sentence to be worse than the original, on account of the "Chinese Whisper" effect<sup>3</sup>.

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

<sup>3</sup> Chinese Whisper:  
[http://en.wikipedia.org/wiki/Chinese\\_whispers](http://en.wikipedia.org/wiki/Chinese_whispers)

Original French	Closest correct and native sounding English	Error classification
<i>jouer les sports</i>	<i>to play sports</i>	<b>Interference.</b> Word for word translation resulting in bad lexical choice (should be “ <i>pratiquer des sports</i> ”). <b>Not just function words.</b> Requires changing content word “ <i>jouer</i> ”
je suis celle <i>que</i> vous avez besoin	I am the one you need	<b>Not interference.</b> Erroneous word “ <i>que</i> ” has no equivalent in the correct native-sounding English. <b>Function words only.</b>
ça fait <i>un</i> longtemps	it's been <i>a</i> long time	<b>Interference.</b> Superfluous determiner “ <i>un</i> ” corresponds to “ <i>a</i> ” in correct English. <b>Function words only.</b>
si j'avais eu donner	if I had been given	<b>Interference.</b> Word for word translation resulting in bad choice of words, and bad choice of verb tense. <b>Not just function words.</b> Although error mostly involves function words, it requires a change of tense in main verb “ <i>donner</i> ”.
elle veut <i>amaigrir</i>	she wants to <i>lose weight</i>	<b>Not interference.</b> The bad lexical choice <i>amaigrir</i> is not a word for word translation of the proper English collocation <i>lose weight</i> . <b>Not just function words.</b>
je dois trouver <i>le</i>	I must <i>find it</i>	<b>Interference.</b> Word for word translation resulting in inversion of two words. <b>Not just function words.</b> Although the error concerns placement of determiner “ <i>le</i> ”, fixing the error requires that it be swapped with content word “ <i>trouver</i> ”.

Table 1: Examples of errors with their classifications in the two axes.

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, and this turns out to mirror exactly what L2 authors do when they write poorly formed L2 sentences by translating too literally from their L1 intent. As a result, that part of 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 part back to L2, it is therefore able to use longer entries from its phrase table, and hence produce a better L2 translation than what the author originally produced. While this theory sounds plausible, we have not been able to verify it in this work, as we did not have access to the phrases used by Google Translate in the course of roundtrip translation.

## 5 Evaluation Corpus

In order to evaluate the roundtrip approach, we collected a total of 829 erroneous sentences from a corpus of texts written by 30 students, during one semester in an advanced-intermediate French as a Second Language (FSL) course given to university students. The class included native English students, as well as allophones for whom English was

a second language and French was a third. In pulling out errors from this corpus, we focused only on errors which were clearly unlikely for a native speaker. Most sentences presented several errors, including some that could have been made by native speakers (ex: spelling mistakes, agreement). All sentences were fed as is to the roundtrip translation procedure, without any pre-massaging. In particular, we did not run them through standard spelling and grammar checkers.

Each error was classified along two axes. Table 1 shows different examples of how this was done. The first axis looked at whether the error was due to linguistic interference from English or not. For an error to be categorized as such, we had to be able to think of a proper native-sounding English sentence which had the same meaning as the French one, and for which the faulty part of the French was a word for word translation from the English. The reason for this classification was to assess the degree to which the roundtrip approach might work better on interference errors, since it is essentially trying to model the process that produces them.

The second axis of classification looked at whether the error could be fixed simply by substituting a contiguous sequence of function words (prepositions, determiners, conjunctions, auxiliaries, etc.), for another contiguous sequence of function words. The reason for this second classifi-

S2 (original L2 sentence)	S1 (intermediate L1 translation)	S2* (roundtrip L2 translation)
C'est une jeune fille d'une vingtaine d'années qui <i>&lt;fait promener&gt;</i> les chiens de ses voisins.	This is a girl of twenty years which <i>&lt;makes walking&gt;</i> dogs from its neighbors. <b>Not corrected.</b> Failed to infer the proper English collocation "walk the dogs".	Il s'agit d'une jeune fille de vingt ans qui <i>&lt;fait marcher&gt;</i> les chiens de ses voisins. <b>Corrected.</b> Eventhough "Il s'agit" may have slightly changed the meaning of that part of the sentence, the roundtrip did fix the part which was flagged for repair, by using a proper French collocation "faire marcher les chiens".
J'ai deux frères et beaucoup <i>&lt;des&gt;</i> oncles.	I have two brothers and many <i>&lt;&gt;</i> uncles. <b>Corrected.</b>	J'ai deux frères et <i>&lt;de&gt;</i> nombreux oncles. <b>Corrected.</b> Although the roundtrip unnecessarily changed content word "beaucoup" to "nombreux" (which is still appropriate), the sentence is grammatical, native sounding, and preserves the author's original meaning.
La vieille dame a <i>&lt;formé instantanément une amitié&gt;</i> avec les huit garçons	The old lady has <i>&lt;formed an instant friendship&gt;</i> with the eight boys. <b>Corrected.</b> Although the choice of tense is dubious, the intermediate translation does address the error which was flagged for repair, namely, failure to use a proper collocation for "form an instant friendship".	La vieille dame a <i>&lt;formé un instant d'amitié&gt;</i> avec les huit garçons. <b>Not corrected.</b> The meaning has been changed to "had a moment of friendship".
Je donnerais de l'argent pour construire des maisons pour ces qui n'en ont pas <i>&lt;une&gt;</i> .	I would give the money to build houses for those who do not <i>&lt;&gt;</i> . <b>Not corrected.</b> Meaning has changed from "those who do not have a house" to "those who do not build a house".	Je souhaite donner de l'argent pour construire des maisons pour <i>&lt;ceux qui ne le font pas&gt;</i> . <b>Not corrected.</b> Meaning has changed from "those who do not have a house" to "those who do not build a house".

Table 2: Examples of roundtrip translations from original L2 sentence (S2), through L1 sentence (S1) and back to L1 (S2\*).

cation was to assess the degree to which the roundtrip approach could deal with "difficult" errors which involve more than changing a few contiguous function words.

Selection of the erroneous phrases and their classification into the two axes was done in two phases. For each of those tasks, one of the authors of the present paper did a first pass, and his choices were subsequently validated by the other author. All differences of opinion were resolved through discussion.

## 6 Evaluation procedure

For each error in our corpus, we carried out roundtrip translation from French to English, and then back to French, using the Google Translate web service. We then looked at both legs of this trip, and assessed whether or not the error had been repaired. The reason for assessing both legs was to evaluate the degree to which the system was actually able to infer the the author's L1 intent (S1) based on the erroneous L1 sentence (S2). Regarding the first leg, we should point out that we were

always able to manually infer the author's intended meaning beyond doubt. This is because we had access to the full length of the text in which the sentence appeared, and the intended meaning was always clear from this context, even if the sentence was severely garbled.

As we pointed out earlier, this simple roundtrip approach cannot distinguish between false positive and true positives. Therefore, we focused only on error correction, that is, given that we know a particular part of the original S2 sentence to be erroneous, we looked at whether the roundtrip S2\* reformulated that part in a way that fixed the error or not. In particular, this means that if S2\* introduced new errors in parts of S2 that were not considered for repair, the roundtrip could still be considered a success if it did repair the part of S2 which *had* been flagged for repair.

An erroneous part of the original S2 was deemed to have been repaired if its corresponding part in S2\*:

- was grammatically correct

	N	P(S2*)	P(S1)	P(S2* S1)
All	829	0.556	0.669	0.763
Interf	497	0.597	0.746	0.759
Non-Interf.	332	0.493	0.557	0.772
Fct Wrds	520	0.625	0.733	0.789
Non Fct	309	0.440	0.566	0.708

Table 3: Number of errors of each type (N), probability that the error was repaired in the overall roundtrip (P(S2\*)), and in each leg of the trip (P(S1) and P(S2\*|S1)).

- was native sounding (Google counts used in case of doubt)
- preserved the meaning intended by the author in the corresponding part of S2

The last point means that, as long as the first two conditions were met, S2\* was allowed to make larger reformulations than were strictly necessary. For example, if it changed an erroneous preposition, but also changed its support verb for a synonym, the change could still be considered a success event though changing the support verb was not absolutely necessary.

Table 2 provides examples of various sentences and how the success of the roundtrip was evaluated.

## 7 Results and Discussion

Table 3 summarizes the results of the evaluation. For the various types of errors, it provides the probability that the error was repaired in the overall roundtrip (P(S2\*) column) as well as in each leg of the roundtrip (P(S1) and P(S2\*|S1) columns). The probabilities for P(S2\*|S1) were estimated by taking the sample mean for the cases where S1 repaired the error. Note that P(S2) is not necessarily equal to the product P(S1) and P(S2\*|S1) because there were cases where S2\* was correct, even though S1 was incorrect.

A first observation is that in our data, only 60% of the foreign-sounding errors were caused by linguistic interference. Although we did not collect native-sounding errors in our corpus of texts, we noticed that there was a large number of them, none of which can (by definition) be transfer errors. Therefore, it is unlikely that the majority of errors present in our texts were caused by linguistic

interference. This contradicts what was previously reported (Selinker, 1969; Sheen, 1980; Cowan, 1983; Wang and Garigliano, 1992).

We can offer two possible explanations for this. Firstly, it could be that previous authors used “transfer errors” in the sense of “an error which a native speaker would not make”. In contrast, we adopted a stricter definition where there had to be a one-to-one mapping between a correct L1 sentence with the appropriate sense, and the incorrect L2 sentence. Another hypothesis is that, because our corpus was partly produced by learners whose first language was neither English or French, it may contain a large number of errors which involve linguistic interference through a language other than English. More research needs to be done to clarify this.

Looking at the P(S2\*) column, we see that roundtrip translation repaired 55.6% of all errors. If we compare the Interf and Non-Interf rows, we see that the approach is better at repairing interference errors, and that this holds whether we look at the overall roundtrip or each leg separately. This was to be expected, since the error repair strategy specifically models the process through which L2 authors produce transfer errors. All differences were found to be statistically significant ( $p < 0.05$ ), except for the difference for the second leg of the trip P(S2\*|S1).

Comparing the Fct Wrds and Non Fct rows, we see that the approach is better at repairing function word errors, and that this holds whether we look at the overall roundtrip or each leg separately. In this case, all three differences were found to be statistically significant. This indicates that, although roundtrip translation is able to deal with errors that involve more than just function words, it still performs better on errors of that type.

Comparing the P(S1) and P(S2\*|S1) columns, we see that the first leg of the trip exhibits more variance than the second one, meaning that reverse engineering the author’s L1 intent is more sensitive to the type of error than rendering a correct L1 intent back into L2. We also see that with the exception of interference errors, the first leg tends to be more successful than the second leg of the roundtrip. All differences (with the exception of the interference row) were found to be statistically significant. This is was to be expected since MT systems are trained mostly on well formed sentences. Therefore, they are less likely to produce a



correct target sentence when the source sentence is ill-formed. The one exception to this seems to be cases where the ill-formed source sentence is a word for word translation of a well formed target sentence, in which case, performance seems to be as good as with well-formed source sentences.

## 8 Conclusions and Future Work

We have presented a very first attempt at evaluating the use of MT technology for repairing errors in second language authoring.

While performances may seem moderate (55.6% repair rate), one must remember that this simple generic approach was used to tackle a very wide range of errors, including choice of preposition and determiners, choice of verb tense, lexical choice for content words, syntactic word order and collocations. Some of those (ex: lexical choice of content words) are particularly challenging. More work needs to be done to compare the performance of MT based correction on such wide range of errors, to that of unilingual approaches such as the ones cited in section 3.

It would be interesting to test the approach using other MT systems. In particular, given that many of the errors in our corpus involved improper use of L1 grammar and syntax, it would be interesting to see if a hybrid SMT system which takes syntax and grammar into account (see Eisele 2008 for a survey) would perform better than a purely phrase-based system like the one we used in this study (Google Translate).

Certainly, the fact that even this simplest possible use of MT in a L2 correction context was able to correctly infer the L1 intent behind 74.6% of interference errors, is quite encouraging. Future work should also be done in order to evaluate more complex algorithms based on MT. For example, instead of choosing S1 and S2\* which respectively maximize likelihood for each leg of the roundtrip, one might instead choose the S2\* which maximizes likelihood of the complete path. This would involve the generation of a combined phrase lattice for the overall roundtrip, instead of separate lattices for each leg.

More sophisticated approaches are also needed to deal with error detection, as opposed to repair. For example, one might analyze the combined phrase lattice to identify S2\* segments whose

probability given S1 are sufficiently high to be considered better than their counterpart in S1. These S1 segments would then be flagged as errors.

## 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. Also to Howard Johnson, for help with statistical analysis of the results.

## 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. Robinson, 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.
- Eisele, A. 2008. *Hybrid Architectures for Multi-Engine Machine Translation*, ASLIB Translating and the Computer, 30, London, 27-28 November 2008.
- 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.
- 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, Morocco.
- Hermet, M., Désilets, A. 2009. *Using First and Second Language Models to Correct Preposition Errors in*

- Second Language Authoring*. In proc. of the 4th Workshop on Innovative Use of NLP for Building Educational Applications, Boulder, Co, USA, June 5, 2009.
- La Torre, M. D. 1999. *A web-based resource to improve 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.
- Schuster, E. 1986. *The role of native grammars in correcting errors in second language learning*. Computational Intelligence, Vol 2(1).
- Selinker, L. 1969. *Language Transfer*. General Linguistics, 9, pp 69-92.
- Sheen, R. 1980. *The Importance of Negative Transfer in Speech of Near-Bilinguals*, International Review of Applied Linguistics, 18(2), pp 105-119.
- Shei, C. C., Pain, H. 2001. *Learning a foreign Language Through Machine Translation: Focusing on Sentence Stems and Collocations*. In Proceedings of the CALLWorkshop held in conjunction with AIED'01.
- Somers, Harold. 2001. *Three Perspectives on MT in the Classroom*, MT SUMMIT VIII Workshop on Teaching Machine Translation, Santiago de Compostela, pages 25-29.
- 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.