Correcting ESL Errors Using Phrasal SMT Techniques

Chris Brockett, William B. Dolan, and Michael Gamon

Natural Language Processing Group

Microsoft Research

One Microsoft Way, Redmond, WA 98005, USA

{chrisbkt,billdol,mgamon}@microsoft.com

Abstract

This paper presents a pilot study of the use of phrasal Statistical Machine Translation (SMT) techniques to identify and correct writing errors made by learners of English as a Second Language (ESL). Using examples of mass noun errors found in the *Chinese Learner Error Corpus (CLEC)* to guide creation of an engineered training set, we show that application of the SMT paradigm can capture errors not well addressed by widely-used proofing tools designed for native speakers. Our system was able to correct 61.81% of mistakes in a set of naturally-occurring examples of mass noun errors found on the World Wide Web, suggesting that efforts to collect alignable corpora of pre- and post-editing ESL writing samples offer can enable the development of SMT-based writing assistance tools capable of repairing many of the complex syntactic and lexical problems found in the writing of ESL learners.

1 Introduction

Every day, in schools, universities and businesses around the world, in email and on blogs and websites, people create texts in languages that are not their own, most notably English. Yet, for writers of English as a Second Language (ESL), useful editorial assistance geared to their needs is surprisingly hard to come by. Grammar checkers such as that provided in Microsoft Word have been designed primarily with native speakers in mind. Moreover, despite growing demand for ESL proofing tools, there has been remarkably little progress in this area over the last decade. Research into computer feedback for ESL writers remains largely focused on small-scale pedagogical systems implemented within the framework of CALL (Computer Aided Language Learning) (Reuer 2003; Vanderventer Faltin, 2003), while commercial ESL grammar checkers remain brittle and difficult to customize to meet the needs of ESL writers of different first-language (L1) backgrounds and skill levels.

Some researchers have begun to apply statistical techniques to identify learner errors in the context of essay evaluation (Chodorow & Leacock, 2000; Lonsdale & Strong-Krause, 2003), to detect non-native text (Tomokiyo & Jones, 2001), and to support lexical selection by ESL learners through first-language translation (Liu et al., 2000). However, none of this work appears to directly address the more general problem of how to robustly provide feedback to ESL writers—and for that matter non-native writers in any second language—in a way that is easily tailored to different L1 backgrounds and second-language (L2) skill levels.

In this paper, we show that a noisy channel model instantiated within the paradigm of Statistical Machine Translation (SMT) (Brown et al., 1993) can successfully provide editorial assistance for non-native writers. In particular, the SMT approach provides a natural mechanism for suggesting a correction, rather than simply stranding the user with a flag indicating that the text contains an error. Section 2 further motivates the approach and briefly describes our SMT system. Section 3 discusses the data used in our experiment, which is aimed at repairing a common type of ESL error that is not well-handled by current grammar checking technology: mass/count noun confusions. Section 4 presents experimental results, along with an analysis of errors produced by the system. Finally we present discussion and some future directions for investigation.

Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL, pages 249–256,
Sydney, July 2006.

2 Error Correction as SMT

2.1 Beyond Grammar Checking

A major difficulty for ESL proofing is that errors of grammar, lexical choice, idiomaticity, and style rarely occur in isolation. Instead, any given sentence produced by an ESL learner may involve a complex combination of all these error types. It is difficult enough to design a proofing tool that can reliably correct individual errors; the simultaneous combination of multiple errors is beyond the capabilities of current proofing tools designed for native speakers. Consider the following example, written by a Korean speaker and found on the World Wide Web, which involves the misapplication of countability to a mass noun:

> And I knew many informations about Christmas while I was preparing this article.

The grammar and spelling checkers in Microsoft Word 2003 correctly suggest *many* → *much* and *informations* → *information*. Accepting these proposed changes, however, does not render the sentence entirely native-like. Substituting the word *much* for *many* leaves the sentence stilted in a way that is probably undetectable to an inexperienced non-native speaker, while the use of the word *knew* represents a lexical selection error that falls well outside the scope of conventional proofing tools. A better rewrite might be:

> And I learned a lot of information about Christmas while I was preparing this article.

or, even more colloquially:

> And I learned a lot about Christmas while I was preparing this article

Repairing the error in the original sentence, then, is not a simple matter of fixing an agreement marker or substituting one determiner for another. Instead, wholesale replacement of the phrase *knew many informations* with the phrase *learned a lot* is needed to produce idiomatic-sounding output. Seen in these terms, the process of mapping from a raw, ESL-authored string to its colloquial equivalent looks remarkably like translation. Our goal is to show that providing editorial assistance for writers should be viewed as a special case of translation. Rather than learning how strings in one language map to strings in another, however, "translation" now involves learning how systematic patterns of errors in ESL learners' English map to corresponding patterns in native English

2.2 A Noisy Channel Model of ESL Errors

If ESL error correction is seen as a translation task, the task can be treated as an SMT problem using the noisy channel model of (Brown et al., 1993): here the L2 sentence produced by the learner can be regarded as having been corrupted by noise in the form of interference from his or her L1 model and incomplete language models internalized during language learning. The task, then, is to reconstruct a corresponding valid sentence of L2 (target). Accordingly, we can seek to probabilistically identify the optimal correct target sentence(s) T^* of an ESL input sentence S by applying the familiar SMT formula:

$$T^* = \arg\max_T \{\mathrm{P}(T \mid S)\}$$

$$= \arg\max_T \{\mathrm{P}(S \mid T)\,\mathrm{P}(T)\}$$

In the context of this model, editorial assistance becomes a matter of identifying those segments of the optimal target sentence or sentences that differ from the writer's original input and displaying them to the user. In practice, the patterns of errors produced by ESL writers of specific L1 backgrounds can be captured in the channel model as an emergent property of training data consisting ESL sentences aligned with their corrected edited counterparts. The highest frequency errors and infelicities should emerge as targets for replacement, while lesser frequency or idiosyncratic problems will in general not surface as false flags.

2.3 Implementation

In this paper, we explore the use of a large-scale production statistical machine translation system to correct a class of ESL errors. A detailed description of the system can be found in (Menezes & Quirk 2005) and (Quirk et al., 2005). In keeping with current best practices in SMT, our system is a phrasal machine translation system that attempts to learn mappings between "phrases" (which may not correspond to linguistic units) rather than individual words. What distinguishes

this system from other phrasal SMT systems is that rather than aligning simple sequences of words, it maps small phrasal "treelets" generated by a dependency parse to corresponding strings in the target. This "Tree-To-String" model holds promise in that it allows us to potentially benefit from being able to access a certain amount of structural information during translation, without necessarily being completely tied to the need for a fully-well-formed linguistic analysis of the input—an important consideration when it is sought to handle ungrammatical or otherwise ill-formed ESL input, but also simultaneously to capture relationships not involving contiguous strings, for example determiner-noun relations.

In our pilot study, this system was employed without modification to the system architecture. The sole adjustment made was to have both Source (erroneous) and Target (correct) sentences tokenized using an English language tokenizer. N-best results for phrasal alignment and ordering models in the decoder were optimized by lambda training via Maximum Bleu, along the lines described in (Och, 2003).

3 Data Development

3.1 Identifying Mass Nouns

In this paper, we focus on countability errors associated with mass nouns. This class of errors (involving nouns that cannot be counted, such as *information, pollution*, and *homework*) is characteristically encountered in ESL writing by native speakers of several East Asian languages (Dalgish, 1983; Hua & Lee, 2004).[1] We began by identifying a list of English nouns that are frequently involved in mass/count errors in by writing by Chinese ESL learners, by taking the intersection of words which:

- occurred in either the *Longman Dictionary of Contemporary English* or the *American Heritage Dictionary* with a mass sense
- were involved in $n \geq 2$ mass/count errors in the *Chinese Learner English Corpus CLEC* (Gui and Yang, 2003), either tagged as a mass noun error or else with an adjacent tag indicating an article error.[2]

This procedure yielded a list of 14 words: *knowledge, food, homework, fruit, news, color, nutrition, equipment, paper, advice, haste, information, lunch, and tea*.[3] Countability errors involving these words are scattered across 46 sentences in the CLEC corpus.

For a baseline representing the level of writing assistance currently available to the average ESL writer, we submitted these sentences to the proofing tools in Microsoft Word 2003. The spelling and grammar checkers correctly identified 21 of the 46 relevant errors, proposed one incorrect substitution (*a few advice* → *a few advices*), and failed to flag the remaining 25 errors. With one exception, the proofing tools successfully detected as spelling errors incorrect plurals on lexical items that permit only mass noun interpretations (e.g., *informations*), but ignored plural forms like *fruits* and *papers* even when contextually inappropriate. The proofing tools in Word 2003 also detected singular determiner mismatches with obligatory plural forms (e.g. *a news*).

3.2 Training Data

The errors identified in these sentences provided an informal template for engineering the data in our training set, which was created by manipulating well-formed, edited English sentences. Raw data came from a corpus of ~484.6 million words of Reuters Limited newswire articles, released between 1995 and 1998, combined with a ~7,175,000-word collection of articles from multiple news sources from 2004-2005. The resulting dataset was large enough to ensure that all targeted forms occurred with some frequency.

From this dataset we culled about 346,000 sentences containing examples of the 14 targeted words. We then used hand-constructed regular expressions to convert these sentences into mostly-ungrammatical strings that exhibited characteristics of the *CLEC* data, for example:

- much → many: much advice → many advice
- some → a/an: some advice → an advice
- conversions to plurals: much good advice → many good advices

[1] These constructions are also problematic for hand-crafted MT systems (Bond et al., 1994).

[2] *CLEC* tagging is not comprehensive; some common mass noun errors (e.g., *make a good progress*) are not tagged in this corpus.

[3] Terms that also had a function word sense, such as *will*, were eliminated for this experiment.

Data Size	Whole	Partial	Correctly Left	New Error	Missed	Word Order Error
45K	55.28	0.81	8.13	12.20	21.14	1.63
30K	36.59	4.07	7.32	16.26	32.52	3.25
15K	47.15	2.44	5.69	11.38	29.27	4.07
cf. Word	29.27	0.81	10.57	1.63	57.72	N/A

Table 1. Replacement percentages (per sentence basis) using different training data sets

- deletion of counters: `piece(s)/item(s)/sheet(s) of)`
- insertion of determiners

These were produced in multiple combinations for broad coverage, for example:

`I'm not trying to give you` **`legal advice`**`. →`

- `I'm not trying to give you` **`a legal advice`**`.`
- `I'm not trying to give you` **`the`** `legal advice.`
- `I'm not trying to give you` **`the legal advices`**`.`

A total of 24128 sentences from the news data were "lesioned" in this manner to create a set of 65826 sentence pairs. To create a balanced training set that would not introduce too many artifacts of the substitution (e.g., *many* should not always be recast as *much* just because that is the only mapping observed in the training data), we randomly created an equivalent number of identity-mapped pairs from the 346,000 examples, with each sentence mapping to itself.

Training sets of various sizes up to 45,000 pairs were then randomly extracted from the lesioned and non-lesioned pairs so that data from both sets occurred in roughly equal proportions. Thus the 45K data set contains approximately 22,500 lesioned examples. An additional 1,000 randomly selected lesioned sentences were set aside for lambda training the SMT system's ordering and replacement models.

4 Evaluation

4.1 Test Data

The amount of tagged data in *CLEC* is too small to yield both development and test sets from the same data. In order to create a test set, we had a third party collect 150 examples of the 14 words from English websites in China. After minor cleanup to eliminate sentences irrelevant to the task,[4] we ended up with 123 example sentences to use as test set. The test examples vary widely in style, from the highly casual to more formal public announcements. Thirteen examples were determined to contain no errors relevant to our experiment, but were retained in the data.[5]

4.2 Results

Table 1 shows per-sentence results of translating the test set on systems built with training data sets of various sizes (given in thousands of sentence pairs). Numbers for the proofing tools in Word 2003 are presented by way of comparison, with the caveat that these tools have been intentionally implemented conservatively so as not to potentially irritate native users with false flags. For our purposes, a replacement string is viewed as correct if, in the view of a native speaker who might be helping an ESL writer, the replacement would appear more natural and hence potentially useful as a suggestion in the context of that sentence taken in isolation. Number disagreement on subject and verb were ignored for the purposes of this evaluation, since these errors were not modeled when we introduced lesions into the data. A correction counted as Whole if the system produced a contextually plausible substitution meeting two criteria: 1) number and 2) determiner/quantifier selection (e.g., *many informations → much information*). Transformations involving bare singular targets (e.g., *the fruits → fruit*) also counted as Whole. Partial corrections are those where only one of the two criteria was met and part of the desired correction was missing (e.g., *an*

[4] In addition to eliminating cases that only involved subject-verb number agreement, we excluded a small amount of spam-like word salad, several instances of the word *homework* being misused to mean "work done out of the home", and one misidentified quotation from Scott's *Ivanhoe*.

[5] This test set may be downloaded at http://research.microsoft.com/research/downloads

Input	Shanghai residents can buy **the fruits** for a cheaper price than before.
Replacement	Shanghai residents can buy **fruit** for a cheaper price than before .
Input	Thank u for giving me **so many advice**.
Replacement	thank u for giving me **so much advice** .
Input	Acquiring **the knowledge** of information warfare is key to winning wars
Replacement	acquiring **knowledge** of information warfare is key to winning wars
Input	**Many knowledge** about Li Bai can be gain through it.
Replacement	**much knowledge** about Li Bai can be gain through it .
Input	I especially like drinking **the tea**.
Replacement	i especially like drinking **tea** .
Input	Icons printed on **a paper** have been brought from Europe, and were pasted on boards on Taiwan.
Replacement	icons printed on **paper** have been brought from Europe , and were pasted on boards on Taiwan .

Table 2. Sample corrections, using 45K engineered training data

equipments → *an equipment* versus the targeted bare noun *equipment*). Incorrect substitutions and newly injected erroneous material anywhere in the sentence counted as New Errors, even if the proposed replacement were otherwise correct. However, changes in upper and lower case and punctuation were ignored.

The 55.28% per-sentence score for Whole matches in the system trained on the 45K data set means that it correctly proposed full corrections in 61.8% of locations where corrections needed to be made. The percentage of Missed errors, i.e., targeted errors that were ignored by the system, is correspondingly low. On the 45K training data set, the system performs nearly on a par with Word in terms of not inducing corrections on forms that did not require replacement, as shown in the Correctly Left column. The dip in accuracy in the 30K sentence pair training set is an artifact of our extraction methodology: the relatively small lexical set that we are addressing here appears to be oversensitive to random variation in the engineered training data. This makes it difficult to set a meaningful lower bound on the amount of training data that might be needed for adequate coverage. Nonetheless, it is evident from the table, that given sufficient data, SMT techniques can successfully offer corrections for a significant percentage of cases of the phenomena in question.

Table 2 shows some sample inputs together with successful corrections made by the system. Table 3 illustrates a case where two valid corrections are found in the 5-best ranked translations; intervening candidates were identical with the top-ranked candidate.

4.3 Error Analysis

Table 1 also indicates that errors associated with the SMT system itself are encouragingly few. A small number of errors in word order were found, one of which resulted in a severely garbled sentence in the 45K data set. In general, the percentage of this type of error declines consistently with growth of the training data size. Linearity of the training data may play a role, since the sentence pairs differ by only a few words. On the whole, however, we expect the system's order model to benefit from more training data.

The most frequent single class of newly introduced error relates to sporadic substitution of the word *their* for determiners *a/the*. This is associated with three words, *lunch*, *tea*, and *haste*, and is the principal contributor to the lower percentages in the Correctly Left bin, as compared with Word. This overgeneralization error reflects our attempt to engineer the discontinuous mapping *the X of them* → *their X*, motivated by examples like the following, encountered in the *CLEC* dataset:

Input:	And we can learn **many knowledge** or new information from TV
Candidate 1:	And we can learn **much knowledge** or new information from TV
Candidate 5:	And we can learn **a lot of knowledge** or new information from TV

Table 3. Multiple replacement candidates generated by 45K training set

In this equal world, lots of people are still concerned on **the colors of them …**

The inability of our translation system to handle such discontinuities in a unitary manner reflects the limited ability of current SMT modeling techniques to capture long-distance effects. Similar alternations are rife in bilingual data, e.g., *ne…pas* in French (Fox, 2002) and separable prefixes in German (Collins et al. 2005). As SMT models become more adept at modeling long-distance effects in a principled manner, monolingual proofing will benefit as well.

The Missed category is heterogeneous. The SMT system has an inherent bias against deletion, with the result that unwanted determiners tended not to be deleted, especially in the smaller training sets.

Other errors related to coverage in the development data set. Several occurrences of greengrocer's apostrophes (*tea's*, *equipment's*) caused correction failures: these were not anticipated when engineering the training data. Likewise, the test data presented several malformed quantifiers and quantifier-like phrases (*plenty tea → plenty of tea*, *a lot information → a lot of information*, *few information → too little information*) that had been unattested in the development set. Examples such as these highlight the difficulty in obtaining complete coverage when using handcrafted techniques, whether to engineer errors, as in our case, or to handcraft targeted correction solutions.

The system performed poorly on words that commonly present both mass and count noun senses in ways that are apt to confuse L2 writers. One problematic case was *paper*. The following sentences, for example, remained uncorrected:

He published many paper in provincial and national publication.

He has published thirty-two pieces of papers.

Large amounts of additional training data would doubtless be helpful in providing contextual resolutions to the problems. Improved alignment models may also play a role here in capturing complex structures of the kind represented by constructions involving counters.

5 Discussion

The artificially-engineered training data that we relied on for our experiments proved surprisingly useful in modeling real errors made by non-native speakers. However, this is obviously a less than ideal data source, since the errors introduced by regular expressions are homogenously distributed in a way that naturally-occurring errors are not, creating artifacts that undoubtedly impair our SMT models.

Artificial data of this sort may be useful as proof of concept, but hand engineering such data plainly does not present a viable path to developing real world applications. In order to be able to handle the rich panoply of errors and error interactions encountered in the text of second language learners large quantities of naturally-occurring "before" and "after" texts will need to be collected. By way of illustration, Table 4 shows the output of results of "translating" our test data into more natural English by hand and dumping the pre- and post-editing pairs to the 45K training set.[6] Although we were unable to exactly recover the target sentences, inspection showed that 25 sentences had improved, some significantly, as Table 4 shows. Under the right conditions, the SMT system can capture contextual morphological alternations (*nutrition/nutritious*), together with complex mappings represented by the dependencies *learn ← knowledge ← many* (ESL) and

[6] Since a single example of each pair was insufficient to override the system's inherent bias towards unigram mappings, 5 copies of each pair were appended to the training data.

Input sentence	And we can **learn many knowledge** or new information from TV.
45K system output	and we can **learn much knowledge** or new information from TV .
45K + translation system output	we can **gain a lot of knowledge** or new information from TV .
Input sentence	The following is **one of the homework** for last week.
45K system output	the following is **one of their homework** for last week .
45K + translation system output	the following is **one of the homework assignments** for last week .
Input sentence	i like mushroom,its **very nutrition**
45K system output	i like mushroom , its **very nutrition**
45K + translation system output	i like mushroom , its **very nutritious**

Table 4. Contextual corrections before and after adding "translations" to 45K training data

gain ← *knowledge* ← *a lot of* (English). In a rule-based correction system, an immense amount of hand-coding would be required to handle even a small subset of the potential range of such mismatches between learner and native-like English. This knowledge, we believe, is best acquired from data.

5.1 The Need for Data Collection

Given a sufficiently large corpus of aligned sentences containing error patterns produced by ESL writers of the same L1 background and their corrected counterparts we expect eventually to be able to capture the rich complexity of non-native error within a noisy-channel based SMT model.

As a practical matter, however, parallel data of the kind needed is far from easy to come by. This does not mean, however, that such data does not exist. The void left by commercial grammar checkers is filled, largely unobserved, by a number of services that provide editorial assistance, ranging from foreign language teachers, to language helpdesks in multinational corporations, to mentoring services for conferences. Translation bureaus frequently offer editing services for non-native speakers. Yet, unlike translation, the "before" and "after" texts are rarely recycled in a form that can be used to build translation models. Although collecting this data will involve a large investment in time, effort, and infrastructure, a serious effort along these lines is likely to prove fruitful in terms of making it possible to apply the SMT paradigm to ESL error correction.

5.2 Feedback to SMT

One challenge faced by the SMT model is the extremely high quality that will need to be attained before a system might be usable. Since it is highly undesirable that learners should be presented with inaccurate feedback that they may not have the experience or knowledge to assess, the quality bar imposed on error correction is far higher than is that tolerated in machine translation. Exploration of error correction and writing assistance using SMT models may thus prove an important venue for testing new SMT models.

5.3 Advantages of the SMT Approach

Statistical Machine Translation has provided a hugely successful research paradigm within the field of natural language processing over the last decade. One of the major advantages of using SMT in ESL writing assistance is that it can be expected to benefit automatically from any progress made in SMT itself. In fact, the approach presented here benefits from all the advantages of statistical machine translation. Since the architecture is not dependent on hard-to-maintain rules or regular expressions, little or no linguistic expertise will be required in developing and maintain applications. As with SMT, this expertise is pushed into the data component, to be handled by instructors and editors, who do not need programming or scripting skills.

We expect it to be possible, moreover, once parallel data becomes available, to quickly ramp up new systems to accommodate the needs of

learners with different first-language backgrounds and different skill levels and to writing assistance for learners of L2s other than English. It is also likely that this architecture may have applications in pedagogical environments and as a tool to assist editors and instructors who deal regularly with ESL texts, much in the manner of either Human Assisted Machine Translation or Machine Assisted Human Translation. We also believe that this same architecture could be extended naturally to provide grammar and style tools for native writers.

6 Conclusion and Future Directions

In this pilot study we have shown that SMT techniques have potential to provide error correction and stylistic writing assistance to L2 learners. The next step will be to obtain a large dataset of pre- and post-editing ESL text with which to train a model that does not rely on engineered data. A major purpose of the present study has been to determine whether our hypothesis is robust enough to warrant the cost and effort of a collection or data creation effort.

Although we anticipate that it will take a significant lead time to assemble the necessary aligned data, once a sufficiently large corpus is in hand, we expect to begin exploring ways to improve our SMT system by tailoring it more specifically to the demands of editorial assistance. In particular, we expect to be looking into alternative word alignment models and possibly enhancing our system's decoder using some of the richer, more structured language models that are beginning to emerge.

Acknowledgements

The authors have benefited extensively from discussions with Casey Whitelaw when he interned at Microsoft Research during the summer of 2005. We also thank the Butler Hill Group for collecting the examples in our test set.

References

Bond, Francis, Kentaro Ogura and Satoru Ikehara. 1994. Countability and Number in Japanese-to-English Machine Translation. *COLING-94*.

Peter E Brown, Stephen A. Della Pietra, Robert L. Mercer, and Vincent J. Della Pietra. 1993. The Mathematics of Statistical Machine Translation. *Computational Linguistics*, Vol. 19(2): 263-311.

Martin Chodorow and Claudia Leacock. 2000. An Unsupervised Method for Detecting Grammatical Errors. *NAACL 2000.*

Michael Collins, Philipp Koehn and Ivona Kučerová. 2005. Clause Restructuring for Statistical machine Translation. *ACL 2005*, 531-540.

Gerard M. Dalgish. 1984. Computer-Assisted ESL Research. *CALICO Journal.* 2(2): 32-33

Heidi J. Fox. 2002. Phrasal Cohesion and Statistical Machine Translation. *EMNLP 2002.*

Shicun Gui and Huizhong Yang (eds). 2003 *Zhongguo Xuexizhe Yingyu Yuliaohu. (Chinese Learner English Corpus)*. Shanghai: Shanghai Waiyu Jiaoyu Chubanshe. (In Chinese).

Hua Dongfan and Thomas Hun-Tak Lee. 2004. Chinese ESL Learners' Understanding of the English Count-Mass Distinction. In *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004).*

Ting Liu, Ming Zhou, Jianfeng Gao, Endong Xun, and Changning Huang. 2000. PENS: A Machine-aided English Writing System for Chinese Users. *ACL 2000*.

Deryle Lonsdale and Diane Strong-Krause. 2003. Automated Rating of ESL Essays. In *Proceedings of the HLT/NAACL Workshop: Building Educational Applications Using Natural Language Processing*.

Arul Menezes, and Chris Quirk. 2005. *Microsoft Research Treelet Translation System: IWSLT Evaluation*. Proceedings of the International Workshop on Spoken Language Translation.

Franz Josef Och, 2003. Minimum error rate training in statistical machine translation. *ACL 2003*.

Franz Josef Och and Hermann Ney. 2000. Improved Statistical Alignment Models. *ACL 2000*.

Chris Quirk, Arul Menezes, and Colin Cherry. 2005. *Dependency Tree Translation: Syntactically Informed Phrasal SMT*. ACL 2005.

Veit Reuer. 2003. Error Recognition and Feedback with Lexical Functional Grammar. *CALICO Journal*, 20(3): 497-512.

Laura Mayfield Tomokiyo and Rosie Jones. 2001. You're not from round here, are you? Naive Bayes Detection of Non-Native Utterance Text. *NAACL 2001*.

Anne Vandeventer Faltin. 2003. Natural language processing tools for computer assisted language learning. *Linguistik online* 17, 5/03 (http://www.linguistik-online.de/17_03/vandeventer.html)