Searching for poor quality machine translated text : learning the difference between human writing and machine translations

Carter, Dave; Inkpen, Diana

Publisher's version / Version de l'éditeur:
https://doi.org/10.1007/978-3-642-30353-1
Advances in Artificial Intelligence, pp. 49-60, 2012-05

Searching for Poor Quality Machine Translated Text: Learning the Difference between Human Writing and Machine Translations

Dave Carter[1,2] and Diana Inkpen[3]

[1] University of Ottawa, School of Electrical Engineering and Computer Science
[2] National Research Council Canada, Institute for Information Technology
david.carter@cnrc-nrc.gc.ca
[3] University of Ottawa, School of Electrical Engineering and Computer Science
diana@site.uottawa.ca

Abstract. As machine translation (MT) tools have become mainstream, machine translated text has increasingly appeared on multilingual websites. Trustworthy multilingual websites are used as training corpora for statistical machine translation tools; large amounts of MT text in training data may make such products less effective. We performed three experiments to determine whether a support vector machine (SVM) could distinguish machine translated text from human written text (both original text and human translations). Machine translated versions of the Canadian Hansard were detected with an F-measure of 0.999. Machine translated versions of six Government of Canada web sites were detected with an F-measure of 0.98. We validated these results with a decision tree classifier. An experiment to find MT text on Government of Ontario web sites using Government of Canada training data was unfruitful, with a high rate of false positives. Machine translated text appears to be learnable and detectable when using a similar training corpus.

1 Introduction

Machine translated text often seems to be intuitively identifiable by proficient speakers of a language. This paper aims to determine whether the differences relative to human-written text (whether written in a given language or translated from a second language) are detectable by a machine. English and French human-written texts, and French and English machine translations thereof, respectively, are considered.

State-of-the-art general-purpose machine translation (MT) tools available online are a boon for many users, opening up the essence of web resources written in foreign languages. Machine translation tools like Google Translate, Bing Translator, and Yahoo Babel Fish are, however, misused by some web authors as a stand-in for a professional translator. When such machine-translations are posted to the web statically, they offer the worst of both worlds: they miss out on the constant updates and improvements to the machine translation tools (benefits realized when translation is performed on-the-fly by a reader via browser plugins or translation toolbar code offered by MT products for inclusion by webmasters) and are largely inferior to professional human translations[1]. Such webmasters falsely presume that copy-and-pasting MT text into their sites is better than nothing; when, in fact, they are doing users and natural language processing researchers and developers a disservice.

Text extracted from multilingual websites is used in many natural language processing (NLP) experiments and tools. The presence of machine-translated text in web-based corpora (i.e., the Google ngram dataset) presents a problem. When building models for statistical machine translation (SMT) systems, for example, it is important to not use machine translations in the source materials; SMT systems rely on large corpora to determine possible and disallowed ngrams, and the presence of poor quality machine translated text in the training data may incorrectly suggest that some impossible ngrams are in fact grammatical. The ability to detect portions of machine translated text in a web corpus at training time would be useful for improving the results of such a system.

Organizations that offer large public-facing websites might benefit from the ability to detect portions of machine translated text as a part of quality control procedures. This may be of particular concern to government organizations that have a legal mandate to provide service in multiple languages. Organizations that outsource translations of important documents (manufacturers' product manuals, for example) might benefit from the ability to quickly scan for poor quality MT text in the returned results.

Differences between original text and human-translated text have been investigated[2] and can be detected automatically by a machine[3]. The objective of this study is to detect a third and distinct class of text: that which has been translated by a mainstream machine translation tool. This was achieved to varying degrees of success by running experiments with three bilingual (English and French) parallel corpora: a portion of the Canadian Hansard; a basket of six Canadian federal government web sites; and a large set of web pages gathered from the Canadian province of Ontario government web sites. Experiments with the Hansard and the Government of Canada data used both the human-written data and machine translations thereof. To test whether the technique can be generalized, the models developed on the Government of Canada data were applied to the Ontario web sites, with the goal of finding machine translated text on the latter. The first two experiments were successful; the third, less so.

A key concern in the federal and provincial government data is that there may be machine translated text buried in the training data, which may reduce recall substantially. This was not a concern with the Hansard texts.

All three experiments were run bidirectionally, considering English-to-French and French-to-English translations, in an effort to investigate machine translation detection independent of any particular language-specific features or MT performance differences attributable to individual languages.

The next section describes other work in the field, particularly the related investigation of differences between original text and human-translated text. Section 3 describes the three experiments performed. Section 4 examines the successful results achieved with the Hansard and Government of Canada corpora, and the poor results searching for machine translated text in the Ontario web sites using models trained on the federal government data.

2 Related Work

The field of translationese has been well studied since early investigations by Gellerstam[4] and others in the mid-1980s. Various grammatical and linguistic features of human translationese have been proposed by Santos[5][6]. Baroni and Bernardini successfully trained a machine to distinguish original Italian text from translations and, critically, concluded that a relatively naive machine learning algorithm could classify text with better accuracy than humans[2]. Kurokawa, Goutte, and Isabelle were able to discern original text from human translations with 90% accuracy[3]. Koppel and Ordan determined that translations to a given target language from varying source languages (for example, English translated from German versus English translated from Spanish) are detectably different from each other and can be classified by a machine[7].

Carpuat examines some characteristics of the output of statistical machine translation, noting that SMT has more regularity in lexical choice than human translation[8]; and that some SMT systems arbitrarily choose synonyms for a given word that are inappropriate in a given subject domain, which thereby cause translation errors[8]. Both of these properties (unusual lexical consistency and incorrect selection of out-of-domain synonyms) should be machine learnable.

Current machine translation systems are based on rules written by translators, by statistical techniques that use parallel translated texts, or a combination of the two. Since human translators' text is detectable, and statistical MT has unique characteristics, machine translation should be detectable.

Baroni and Bernardini[2] suggest in passing that parallel corpus extractors might improve if able to distinguish original and human-translated text; an argument that we would like to extend to include machine-translated text. Lembersky, Ordan, and Wintner briefly discuss the problem of removing MT text from training sets while more broadly positing that language models built with translated text outperform those built on original text[9].

The nature of translationese and its learnable features are examined and compared in depth by Ilisei and Inkpen[10] as well as Ilisei et al.[11]; they conclude that the strongest learnable features of translationese are at the morphological level, a useful result of which we will take advantage. Popescu was able to identify translationese using machine learning methods with character-level features[12].

Language detection is a well-studied field. The experiments herein could be thought of as a language detection system where the four supported languages are human-written English, human-written French, machine translated English, and machine translated French. This could be conceptually similar to discerning similar yet distinct cognates like European Portuguese from Brazilian Portuguese, for example.

Machine translation output evaluation seeks to identify the worst parts of a machine translation so that they can be corrected by a human, as investigated by Uchimoto, Hayashida, Ishida and Isahara[13], Russell[14], and Melamed[15]. These papers presuppose that all source text is machine translated and has translation flaws; by contrast, our goal is to find the MT text in a larger collection of human-written/human-translated text. The focus on finding low-quality machine translated text is similar in both cases.

Related concepts are applied in cross-language plagiarism detection, particularly the work by Somers et al.[16], who use computational linguistics techniques (rather than machine learning) to discover texts or portions thereof that are machine translated (and thereby plagiarized) in translation students' assignments.

The useful results we build upon, then, are that translated text written by humans differs from original text; that the magnitude of these translationese characteristics vary depending on the source language (and, by extension, that English-to-French and French-to-English human translations have particular signatures); that machine translations tend to have unusual lexical consistency or odd synonym choices; that machine learning can be used to detect all of the aforementioned differences; that such differences tend to manifest at the unigram level. Our contribution, then, is to demonstrably detect these differences using machine learning methods in English and French texts.

3 Data and Methods

Three experiments were performed, each considering a different data set: the Canadian Hansard, a basket of six public Government of Canada web sites, and a basket of Government of Ontario web sites.

In each experiment, a support vector machine (SVM) was used to classify text as human-written English, human-written French, machine-translated English, or machine-translated French. Since SVM can successfully classify translationese[2][10][11], it offered promise in finding MT text. LibSVM was used.

In all three experiments, the training data consisted of text of the four categories: original and machine translated English and French, labelled *hu-e*, *hu-f*, *mt-e*, and *mt-f*. The human-written text, which, having been created by human writers and translators, represents a combination of original and human-translated text, and thus contains elements of both original text and translationese. These texts were translated en masse by Microsoft's Bing Translator, a free online translation tool. Bing was chosen for two reasons: at the time of the experiments, the tool did not have usage quotas; and its English and French translation performance appears to trail that of the dominant Google Translate service, which was a desirable property for these experiments. Its performance was presumed to be a satisfactory approximation of machine translation tools that may have been used to create the public government web sites as they have been updated over the last several years; whereas advances in some cutting-edge SMT tools may not reasonably approximate the MT text of older systems. Finally, the Bing Translation service appears to have not been trained on Hansard, Government of Canada, and Government of Ontario web sites, unlike some free online translation tools like Apertium, which returns the word-for-word translations contained in the Hansard and several Government of Canada web sites.

The Bing Translation service was not able to translate all sentences. Accordingly, a small portion of sentences in the human-written training data (less than one percent) were not machine translated and were removed from the output.

The features extracted from these training texts were simple unigram frequencies, scaled for the length of each document. Documents were considered wholly; paragraphs were not considered separately, for example. For each document, the type-token ratio of unstemmed unigrams was calculated, as was the average unigram length; these additional features were added to try to model lexical simplification characteristics of MT text. Unigrams representing numbers, cardinals, and symbols were removed, leaving only words and word-like tokens. Files that contained fewer than twenty tokens were removed, as were any pages that appeared to be error or redirection pages.

3.1 Hansard data

The first experiment was performed on the Hansards of the 36th Parliament of Canada[17], a set of transcribed and expertly human-translated debates from the two parliamentary bodies of the government of Canada. These data have been used in successful English/French experiments[3] and are of high quality.

Hansards are free of MT text, and thus constitute a clean training set. Both English and French texts contain original and human-translated text[3].

The Hansard texts were machine translated; analyzed for unigrams, type-token ratio, and average unigram length; and finally classified together with the original texts using LibSVM by ten-fold cross validation and by holding out a small (roughly 6%) test set.

The cleaned corpus consisted of 949 human-written documents (226 MB) and their machine translations for training, and 58 human-written documents (28 MB) and their machine translations held aside for testing.

3.2 Government of Canada data

A second experiment was performed on the collected text of six selected Canadian federal government web sites covering a range of scientific, law enforcement, and financial domains. These six web sites were collected using JSpider. Non-textual files (graphics, stylesheets, PDFs) were discarded, and text was extracted from the HTML pages.

The text of all six federal government web sites was machine translated; analyzed for unigrams, type-token ratio, and average unigram length; and classified using LibSVM by ten-fold cross validation. Each site was classified with a model trained on the other five sites (cross-validation).

This corpus was gathered so that the technique could be applied to real world imperfect data. The sites selected were an attempt at compromise between text that was similar in style (government business writing) while being broad enough to have a notable amount of natural variation (selecting work from different writers and translators at different government departments dealing with different lines of business). It is conceivable that a large portion of the publicly-visible text on the web sites of smaller government departments might be written (or edited) by a single person, and thereby has consistent and detectable style traits; whereas, in our case, enough text was collected so that the models developed would be insensitive to the writing style of any single individual.

Of note, while Government of Canada web sites are carefully maintained and largely human-translated, it is presumed that some MT text may exist, and may thus pollute the training data to a small extent. Our goal was to find any such text. No attempt was made to excise such text from the training sets.

The cleaned corpus contained 21 436 original documents (230 MB of text) and 21 436 MT documents (187 MB).

3.3 Government of Ontario data

A third and final experiment examined Government of Ontario data. JSpider was again used to gather as many pages as possible from 139 Ontario domains representing various ministries, boards, and programs. Text was extracted from almost 19 000 web pages, removing HTML tags in the process.

The Government of Canada models trained previously were used to classify the Ontario data in order to test whether the techniques applied to the Hansard data and Government of Canada data could be applied more generally to a similar (yet distinct) domain; and whether the time- and processing-intensive step of machine translating a corpus of interest could be avoided.

The corpus contained 17 583 nominally human-written documents (204 MB of text) and no machine-translated text.

4 Evaluation

4.1 Hansard data

The training data drawn from the Canadian Hansards were classified using 10-fold cross-validation with LibSVM. An accuracy of 99.89% was achieved overall, relative to a baseline of roughly 25% (as the training classes were all roughly equal in size) with an F-measure of 0.999 in each class (human-written and machine-translated English and French).

Of 474 human-written French documents, one was mis-classified as being machine translated; of 474 machine-translated English documents, one was mis-classified as being human-written. Sets of 475 human-written English documents and their corresponding machine translations were classified with 100% accuracy.

In addition, a matched set of 58 documents in each class was held out as a test set when the training models were built. The training models built had 100% accuracy classifying the testing models (58 documents of each class).

Such strong results might suggest overtrained models. However, our features are simple (unigrams, average token length, and type-token ratio) and not hand-tuned.

A decision tree classifier was trained and examined to verify that there were no extraneous unigrams introduced during data processing that could give the SVM classifier strong hints (untranslated words in the MT output, for example). The decision tree model appeared to choose non-rare words in both English and French, which suggests that the SVM, despite its strong performance, did not have a trivial task.

As the Hansard texts are of high translation quality and do not contain any machine translations, further experiments were conducted on real-world data from several Government of Canada web sites.

4.2 Government of Canada data

The training data drawn from six Government of Canada web sites were classified using 10-fold cross-validation with LibSVM. The classifier performs well, exceeding a chance baseline of 25% accuracy, achieving an average F-measure of 0.98 (Table 1).

Table 1. Analysis of 10-fold cross validation of federal government data

Class	TP	FP	Precision	Recall	F-Measure
hu-e	0.995	0.020	0.961	0.995	0.977
hu-f	0.961	0.001	0.997	0.961	0.978
mt-e	0.944	0.002	0.989	0.944	0.966
mt-f	0.996	0.006	0.988	0.996	0.992
Weighted Avg.	0.980	0.009	0.981	0.980	0.980

Examining the precision, recall, and F-measures of this model, the results seem quite promising if considered as a normal NLP application. Unusually for work in this field, the training data may be polluted with the out-of-class data being sought: the human-written English training data may contain some English text that was machine translated from French, and similarly, the French training data may also contain text machine translated from English. A concern, then, is that very high precision and recall would suggest that the model is not going to be very effective in finding machine translated text hidden in the nominally human-written text; or that such text does not exist. However, delving deeper into the results mollifies this concern to some extent. Fifty-six (nominally) human-written English pages have been classified as machine translated English, and 163 pages in the human-written French class have been predicted to be machine translated. These represent predictions of the machine translations that we sought. The relative scale of these mis-classifications suggests that the model is probably working well. One could imagine a terrible model that randomly classifies cases would, for the *hu-e* class for example, generate a roughly equal number of predictions for the *hu-f*, *mt-e*, and *mt-f* classes. The number of cases predicted to be *mt-e* (56) is much higher than either *hu-f* (13) or *mt-f* (8), which suggests that the model is, at a minimum, able to detect language with some accuracy, and further gives some confidence that the model is working.

The prediction results for the machine translated classes are not directly useful (as they do not provide any direct evidence whether machine translated text is detectable); nonetheless, the high accuracy provides evidence that the models are working well. Further, in the English case, all incorrect predictions classified the data as human-written English (and not as French). The French data are a little more concerning, as there were an uncomfortably high number of machine translated French documents that were predicted to be English relative to the number that were predicted to be human-written French. The relatively small number of such misclassifications (less than 0.3% of cases were predicted to be English) is reassuring.

A more detailed experiment was performed, holding out each of the six federal government web sites and training detection models on the remaining five (Table 2). This is a further extension of the MT detection technique, which seeks to omit domain-specific lexical features unique to a given testing set from the corresponding training models; this is a step closer to having models trained on one data set that could be used to classify a different arbitrary data set.

Table 2. Site-by-site analysis of federal government websites

Web site data	F-measure				Weighted average
	hu-e	hu-f	mt-e	mt-f	
Site 1	0.994	1.000	0.994	1.000	0.997
Site 2	1.000	0.982	0.988	1.000	0.999
Site 3	0.958	0.983	0.954	0.984	0.970
Site 4	0.983	0.994	0.983	0.994	0.988
Site 5	0.769	0.910	0.298	0.941	0.756
Site 6	0.986	0.984	0.967	0.993	0.985

Results for five out of the six sites have excellent F-measures. The results of the tests on Site 5—that with the worst results—bear further examination. Statistical analysis for this particular model appears in Table 3.

The poor results for the machine translated English class stem from very poor recall. As the experiment is not one of finding human-written text in the machine translated text, this is perhaps not a problem that needs to be addressed (whereas such numbers could be considered a fatal flaw if they occurred in the *hu-e* or *hu-f* classes); it is slightly concerning nonetheless. The results for this

Table 3. Analysis of worst-performing federal government model

Class	TP	FP	Precision	Recall	F-Measure
hu-e	0.973	0.232	0.636	0.973	0.769
hu-f	0.918	0.026	0.903	0.918	0.910
mt-e	0.179	0.006	0.875	0.179	0.298
mt-f	0.939	0.024	0.942	0.939	0.941

site model are acceptable, and the results for the other five site models seem to be rather good.

Some MT predictions in the nominally human-written text were evaluated by hand (Table 4), seeking to find instances of MT text on the government sites.

Table 4. Human evaluation of MT predictions in nominally human-written text

Data set	Source pages analysed		Out-of-class predictions		
	hu-e	hu-f	mt-e (# correct)	mt-f (# correct)	opposite lang (# correct)
Site 1	272	272	1 (0)	0 (0)	0 (0)
Site 2	3251	84	1 (0)	0 (0)	1 (0)
Site 3	455	456	0 (0)	15 (14)	0 (0)
Site 4	2375	2494	0 (0)	30 (10)	0 (0)
Site 5	1904	1354	33 (2)	110 (9 of 11 sampled)	20 (19)
Site 6	5665	2559	0 (0)	77 (8 of 11 sampled)	3 (3)
Total	13922	7219	35 (2)	232 (170 est.)	24 (22)

Overall, the detection models work well for detecting French machine-translated text, and are poor at detecting English machine-translated text.

Each document was evaluated in a binary fashion: either as containing some amount of text that is judged by a human to be machine translated text; or as a false positive. In cases where only part of a document is machine translated, the MT portion may have been overlooked, which suggests that there may be more correct MT predictions than listed above. These evaluations can be considered a floor; Baroni and Bernardini determined that machines were better at classifying translationese versus original text[2]—a property which may well apply in this human evaluation as well, as the task is quite similar.

As human evaluation was rather time-consuming and not perfectly reliable, an evenly-distributed sample was taken when many documents were predicted to be machine translated (for sites 5 and 6). In most of the false positive cases, where machine translations were predicted but not present, the language and sentence structure of the page is noticeably different from that of other pages. They tended to be frequently asked question pages, biographies, site maps, glossaries, dialogue/speeches, and lists of factoids.

Most of the cases of machine translated French text were identifiable by connective phrases common in English that were translated literally (and that do not exist as such in French), by incorrect verb forms (missing participles or incorrect verb tense), or by verb or adjective translations where the wrong sense of a word with multiple senses had been chosen; this last attribute is supported by and agrees with the results seen by Carpuat[8].

Disappointingly, all but two of the cases of machine translation that were detected were in French. Both English cases identified included machine translated dialogue. While this is a successful result, the experiment would have been more compelling had more machine translated English text and/or different kinds of machine translated English text been found.

4.3 Government of Ontario data

A final experiment was performed in order to determine whether the somewhat successful Government of Canada prediction models could be applied to find machine translated text on Government of Ontario web sites. The training data drawn from six federal web sites were used to create a LibSVM model that was tested on Ontario data that were not machine translated for classification; in the test set, the English and French machine translation classes were intentionally left empty. A successful experiment would lessen the burden of finding MT text in previously-unseen data in a given topic domain.

Table 5. Experimental analysis of Ontario data using federal government models

Class	TP	FP	Precision	Recall	F-Measure
hu-e	0.855	0.03	0.966	0.855	0.907
hu-f	0.617	0.015	0.973	0.617	0.755
mt-e	-	0.075	-	-	-
mt-f	-	0.183	-	-	-

The experiment was largely unsuccessful (Table 5). The sheer number of predictions of machine translations (1223 English pages and 3147 French pages) suggests a problem; it seems unlikely that a government web site serving a bilingual population could be so rife with detectable machine translated text. A small random set (234 documents—a 5% sample) of the documents classified as machine translations was manually examined and appeared to be, in fact, largely human written (if poorly, at times). Fewer than 20% were judged to contain any MT text, and almost entirely in small quantities; they were likely classified as MT by chance.

As the models were trained largely on unigram features, these results demonstrate that the models trained on one set of web sites and its translations are not applicable to web sites in a similar domain with substantially different vocabulary. This is perhaps not a surprising result, as such machine learning experiments are most successful when the training and test sets are quite similar.

Language detection was still successful despite the simple features selected.

5 Conclusions and Further Research

Just as the traits of human translations can be found with machine learning techniques, the traits associated with machine translation are machine learnable and detectable. Three data sets were examined with varying results: tests to classify machine translations of the Canadian Hansard and several Canadian federal government web sites were successful (with 99.8% and 98% accuracy, respectively), while models trained on the latter performed poorly on Ontario government web sites.

Different machine translation tools may generate different kinds of characteristic suboptimal translations, so building a single model to detect all machine translations may not be sufficient. A natural follow-up would be to identify several commonly-used MT tools and build separate models for each. Since MT systems are updated and improved over time, it would be particularly interesting to do multiple translations of a particular document over time to better detect the output of early MT systems and recent MT systems; these outputs likely differ greatly in quality and in detectable traits.

More general-purpose detection tools might be possible by providing more relevant features to the machine learning algorithms. As Ilisei and Inkpen experimented with the effects of dozens of word- and sentence-level features in trying to detect human-written translationese[10], so too might the detection of MT text be improved with further experimentation.

Results relevant to the linguistic community might be achievable by separating the human-written text into original and translated classes; perhaps MT text is more similar to translationese than original language text or vice versa.

Applying these techniques to other well-regarded multilingual corpora (i.e., EuroParl) would be a natural extension of this work.

While research into the understanding of the features of translationese has been making progress, investigation into the detection of machine-translated text has been largely overlooked to date; further research is warranted.

References

1. Helft, M.: Googles Computing Power Refines Translation Tool. In: New York Times, March 8 2010, A1. Retrieved from http://www.nytimes.com/2010/03/09/technology/09translate.html?nl=technology&emc=techupdateema1 (2010)
2. Baroni, M., Bernardini, S.: A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text. Literary and Linguistic Computing 21, 3, 259-274 (2006)

3. Kurokawa, D., Goutte C., Isabelle, P.: Automatic detection of translated text and its impact on machine translation. MT Summit XII: proceedings of the twelfth Machine Translation Summit, August 26-30, 2009, Ottawa, Ontario, Canada, 81-88 (2009)
4. Gellerstam, M.: Translationese in Swedish Novels Translated from English. In Wollin, L. and Lindquist, H. (eds.) Translation studies in Scandinavia: proceedings from the Scandinavian Symposium on Translation Theory (SSOTT) II, Lund 14-15 June, 1985, 88-95 (1985)
5. Santos, D.: On the use of parallel texts in the comparison on languages. Actas do XI Encontro da Associação Portuguesa de Linguística, Lisboa, 2-4 de Outubro de 1995, 217-239 (1995)
6. Santos, D.: On grammatical translationese. In Koskenniemi, Kimmo (ed.), Short papers presented at the Tenth Scandinavian Conference on Computational Linguistics, Helsinki, 29-30 (1995)
7. Koppel, M., Ordan, N.: Translationese and Its Dialects. Proceedings of ACL, Portland OR, June 2011, 1318-1326 (2011)
8. Carpuat, M.: One Translation per Discourse. In: Agirre, E., Márquez, L., Wicentowski, R. (eds.) SEW-2009 Semantic Evaluations: Recent Achievements and Future Directions, 19-27 (2009)
9. Lembersky, G., Ordan, N., Wintner, S.: Language models for machine translation: original vs. translated texts. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, July 2731, 2011, 363-374 (2011)
10. Ilisei, I., Inkpen, D.: Translationese Traits in Romanian Newspapers: A Machine Learning Approach. In: Gelbukh, A (ed.) International Journal of Computational Linguistics and Applications, 2011 (in press)
11. Ilisei, I., Inkpen, D., Pastor, G., Mitkov, R.: Identification of Translationese: A Machine Learning Approach. Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science 6008, 503-511 (2010)
12. Popescu, M.: Studying Translationese at the Character Level. Proceedings of Recent Advances in Natural Language Processing, 634639 (2011)
13. Uchimoto, K., Hayashida, N., Ishida, T., Isahara, H.: Automatic detection and semi-automatic revision of non-machine-translatable parts of a sentence. LREC-2006: Fifth International Conference on Language Resources and Evaluation. Proceedings, Genoa, Italy, 22-28 May 2006, 703-708 (2006)
14. Russell, G.: Automatic detection of translation errors: the TransCheck system. Translating and the Computer 27: proceedings of the Twenty-seventh International Conference on Translating and the Computer, 24-25 November 2005, London, 17. Aslib, London (2005)
15. Melamed, D.: Automatic detection of omissions in translations. Coling 1996: the 16th International Conference on Computational Linguistics: Proceedings, August 5-9, 1996, Center for Sprogteknologi, Copenhagen, 764-769 (1996)
16. Somers, H., Gaspari, F., Niño, A.: Detecting inappropriate use of free online machine translation by language students. A special case of plagiarism detection. EAMT-2006: 11th Annual Conference of the European Association for Machine Translation, June 19-20, 2006, Oslo, Norway, 41-48 (2006)
17. Germann, U. (Editor).: Aligned Hansards of the 36th Parliament of Canada Release 2001-1a. Retrieved from http://www.isi.edu/natural-language/download/hansard/ (2001)