

Measuring the Impact of Spelling Errors on the Quality of Machine Translation

Irina Galinskaya, Valentin Gusev, Elena Meshcheryakova, Mariya Shmatova

Yandex

16, Leo Tolstoy St., Moscow 119021, Russia

E-mail: (galinskaya, vgoussev, mescheryakova, mashashma)@yandex-team.ru

Abstract

In this paper we show how different types of spelling errors influence the quality of machine translation. We also propose a method to evaluate the impact of spelling errors correction on translation quality without expensive manual work of providing reference translations.

Keywords: machine translation, spelling errors, quality estimation

1. Introduction

It has been shown in numerous works (e.g. Clark, 2003; Bertoldi et al. 2010; Banerjee et al. 2012) that noise, especially frequent in user-generated content, leads to severe degradation of translation quality. This problem is especially relevant for free online translation systems, used to translate any type of content with any level of noise.

One possible way to improve the quality of machine translation is to correct the input text. When translating from a language unknown to the user, the original text processing should be done automatically, as a part of the machine translation algorithm (Kirchhof et al., 2007; Bertoldi et al. 2010; Roturier et al., 2012).

However, different types of spelling errors, including wrong punctuation and use of capital letters, may affect the translation differently. If it is true, it would be useful to estimate the relative influence of various types of spelling errors on the quality of machine translation. Such a study can help to enhance the pre-processing routine and ultimately improve the translation quality.

Note also that each language has its special spelling rules, which, if neglected, can lead to certain errors in translation queries. For example, in German and French many words contain letters with diacritical symbols, while in English there are almost no such words; so we can suppose that omitting diacritics can damage French or German text to a greater extent, than the English one. On the other hand, according to German orthographic rules all nouns should begin with a capital, so omitting capitals results in a higher error rate; see examples in Section 3. We might expect therefore that the effect of different types of spelling errors on the machine translation quality would be different for each input language.

In this paper we consider the impact of most common spelling errors on the quality of statistical machine translation in three languages pairs: English-Russian, German-Russian and French-Russian. We also propose a new way to estimate the impact of error correction that is less time-consuming and requires less qualified human expertise.

In the following section we cite previous works that dealt with related issues. In Section 3 we discuss problems associated with translation of noisy user-generated texts.

To evaluate the impact of different types of spelling errors on the translation quality, we carried out a series of experiments described in Section 4. In Section 5 we propose a new way to evaluate the difference in translation quality without using time-consuming procedure of creating reference translations.

2. Related Work

A number of papers have addressed problems in translating user-generated content; see for example (Aikawa et al., 2007; Banerjee et al., 2012). The impact of spelling errors of any type on the quality of machine translation has been considered in several works; see (Kirchhof et al., 2007; Subramaniam et al., 2009; Carrera et al., 2009; Bertoldi et al., 2010; Jiang et al., 2012; Plesco and Rychtycky, 2012). All these authors present their own classification of errors, from very large text-level classes (Kirchhof et al., 2007) to those dealing only with word-level misspellings (Bertoldi et al., 2010). The classification of error types given in (Clark, 2003) is especially close to what we use here; Clark distinguishes four classes of errors: (1) wrong token boundaries (word breaking errors); (2) wrong sentence boundaries (sentence-final punctuation); (3) misspellings of words; (4) wrong capitalization.

Spelling errors correction can be a part of text pre-processing for any automatic analysis (see e.g. Shoukry, Rafea, 2012 on Arabic twitter texts normalizing for sentiment analysis) and, in particular, for machine translation (Kirchhof et al., 2007; Bertoldi et al., 2010; Roturier et al., 2012).

3. User-Generated Input for Machine Translation

Online machine translation systems often deal with user-generated noisy texts that may contain lots of misspellings, incorrect use of punctuation marks and capital letters. Translation of copy-pasted text is also sometimes problematic: for instance, elements of a web page being pasted into a translation box, turn into sets of words and phrases not separated by any punctuation marks.

For our study we analyzed translation queries to the in-house online translation system and closely examined

	Source	MT result
(1) French; word breaking errors		
(a)original	Bonjour! Je m'appelle <u>X</u> .J'ai seize ans. <u>J</u> 'habite à Québec. <u>C</u> 'est la capitale du Québec. <u>la</u> province francophone du Canada...Québec est très joli.	Hello! I'm <u>X</u> . <u>J</u> 'have sixteen ans. <u>J</u> ' lives Québec. <u>C</u> is the capital of Quebec, the French-speaking province of Quebec Canada ...is very pretty.
(b)corrected	Bonjour! Je m'appelle <u>X</u> . J'ai seize ans. J'habite à Québec. C'est la capitale du Québec. <u>la</u> province francophone du Canada... Québec est très joli.	Hello! My name is <u>X</u> . I'm sixteen. I live in Quebec. It is the capital of Quebec, the French-speaking province of Canada ... Quebec is very pretty.
(2) French; misspellings		
(a)original	Je t'embrace!	I <u>fire</u> you!
(b)corrected	Je t'embrasse!	I <u>kiss</u> you!
(3) German; misspellings		
(a)original	Die Natur wird von Verkehrsmitteln, Flugzeugen, Autos <u>betroht</u> .	Nature is <u>betroht</u> of transport, aircraft, cars.
(b)corrected	Die Natur wird von Verkehrsmitteln, Flugzeugen, Autos <u>bedroht</u> .	The nature is <u>threatened</u> by transport, aircraft, cars.
(4) French; wrong punctuation		
(a)original	santé amis famille	family friendly health
(b)corrected	santé amis famille	health, friends, family

Table 1: Examples of different spelling errors and their influence on the translation quality.

all kinds of errors in them. We consider here four types of spelling errors:

- 1) word breaking errors;
- 2) misspellings;
- 3) wrong capitalization;
- 4) wrong punctuation.

Compared to (Clark 2003), we added to the fourth class the sentence-internal punctuation, because — as we have noticed during manual analysis of machine translated texts — it may also affect the translation (cf. examples (4a and b) in Table 1).

In this work we limit ourselves to relatively simple types of errors and leave more complex cases for future.

Here are some typical examples:

1. Word breaking errors. English: *happi ness* → *happiness*; German: *Wenn Sie mit den Änderungen einverstanden sind, müssen Sie nichts tun.* → *Wenn Sie mit den Änderungen einverstanden sind, müssen Sie nichts tun* 'If you agree with the changes, you should do nothing'; French: *profi l* → *profil* 'profile'. Such misprints are often systematic, and since many online translation systems cannot discern words that are not properly divided, such errors lead to a sharp decrease of translation quality.
2. Misspellings: substitution, omission or addition of one or several letters, or omission of diacritics. English: *Russin* → *Russian*; *countri* → *country*; French: *était* →

était 'was'; *Frabce* → *France*; German: *mit disem* → *mit diesem* 'with this'; *mude* → *müde* 'tired'.

3. Capitalization errors: English: *i* → *I*; German: *der zug* → *der Zug*; *sind diese probleme für dich so wichtig* → *Sind diese Probleme für dich so wichtig?* 'Are these problems so important to you?'; French: *sont arrivés ce matin À Blois* > *sont arrivés ce matin à Blois* 'have arrived this morning in Blois'.

4. Wrong punctuation: English: *Is it far* → *Is it far?*, French *santé amis famille* → *santé, amis, famille*.

It should be noted that one query may and very often does contain several types of spelling errors.

In Table 1 we give just a few examples to illustrate how spelling errors can affect the translation results. Examples (a) are original ones, in examples (b) the relevant errors have been corrected; note that these are not fully corrected queries, so they illustrate the impact of the given type of errors only.

From these examples it is clear that the noisier the input text is, the poorer quality of the translation we can expect.

4. Experiments

4.1 Data

For our study we chose 3 translation directions: English-Russian, German-Russian and French-Russian. For each pair we took a test set of 500 randomly selected

translation queries to the in-house online translation service. Each query was no longer than 1000 symbols.

Then we asked human editors (one editor for each language) to correct separately different types of spelling errors. This resulted in the following 6 test sets for each language:

- original one (orig);
- set with corrected word breaking errors (wbr);
- set with corrected misspellings (msp);
- set with corrected capitalization (caps);
- set with corrected punctuation marks (punct);
- set where all types of errors have been corrected (all).

Table 2 presents comparative data on different error types in our 3 test sets (percentage of words with the given type of errors). We see that, for instance, in English most spelling errors result from incorrect punctuation, in German set word breaking errors are the most frequent, etc. The rate of capitalization errors is the lowest in all test sets.

Error type	EN	DE	FR
Word breaking errors	1.8%	4.3%	4.2%
Misspellings	1.9%	2.3%	4.3%
Capitalization	0.5%	0.6%	0.6%
Punctuation	3.8%	2.1%	1.9%
All	7,4%	8,5%	10%

Table 2: Share of words with different types of spelling errors in test sets.

The test sets with all types of errors corrected (English, French and German) were translated into Russian by professional translators in order to obtain reference translations. After that we translated all the sets with three publicly available online statistical machine translation services (S1, S2, S3). We consider these three systems as blackboxes, presuming that we know nothing about their internal structure, except the fact that they are statistical ones. Then we used BLEU metric (Papineni et al., 2002) for quality evaluation (see Roturier et al. 2012 on using automatic metrics for evaluating the impact of source correction).

4.2 Evaluation Results

Tables 3 and 4 demonstrate the BLEU scores for each translation system (S1, S2 and S3) and each language depending on the type of errors that were corrected. Figures in parentheses indicate the difference in BLEU compared to the original set.

We see that correction of all — both spelling and punctuation — errors can raise the BLEU scores of statistical machine translation by 2 to 5 BLEU points, or 10-15% relative. As for specific types of misspellings, the impact of each of them may differ depending on the source language and translation system.

If we measure case-insensitive BLEU (Table 3), the impact of incorrect capitalization on the translation quality is minimal and close to zero — though the amount

	orig	wbr	msp	caps	punct	all
English						
S1	30,3	30,9 (0,6)	31,1 (0,8)	30,3 (0,0)	31,2 (0,9)	32,8 (2,5)
S2	29,5	30,1 (0,6)	30,4 (0,9)	29,6 (0,1)	31,4 (1,9)	33,5 (4,0)
S3	27,5	27,8 (0,3)	28,3 (0,8)	27,6 (0,1)	29,6 (2,1)	30,7 (3,2)
German						
S1	25,7	27,7 (2,0)	26,9 (1,2)	25,8 (0,1)	26 (0,3)	29,8 (4,1)
S2	22,6	24,4 (1,8)	23,6 (1,0)	22,6 (0,0)	23 (0,4)	26 (3,4)
S3	20,3	21,3 (1,0)	21,1 (0,8)	20,3 (0,0)	20,5 (0,2)	22,6 (2,3)
French						
S1	28,5	29,9 (1,4)	30,3 (1,8)	28,9 (0,4)	29 (0,5)	33,4 (4,9)
S2	24,3	25,6 (1,3)	25 (0,7)	24,6 (0,3)	25,3 (1,0)	28,2 (3,9)
S3	23,3	23,7 (0,4)	24,3 (1,0)	23,4 (0,1)	24,2 (0,9)	25,9 (2,6)

Table 3: BLEU scores calculated using reference translations (case-insensitive)

of capitalization errors in German and French is high enough (see Table 2). This is due to the fact that we disregard here the correct use of capitals in translation. If we measure case-sensitive BLEU, the figures will change (see below).

Correction of wrong punctuation leads to a noticeable increase of BLEU scores in all three systems (evidently because of high percentage of these errors in the test sets), especially for S3. On the other hand, for German, the main factor that affects the translation quality in all systems is incorrect word breaking — though the rate of queries containing such errors is not so high.

If we look at BLEU scores that take into account capitalization (see Table 4), the results will be somewhat different.

We see that overall figures are lower, which is natural, because now it is also the capitalization that must be identical with the reference. However, we would expect that the difference in “caps” and in “all” columns would be higher: correcting capitals in source text should entail correct capitalization in translation. Generally this is what we see, but in S2 correcting only capitalization in German does not result in any difference in BLEU score, though if all errors are corrected, the difference in BLEU is higher, as it should be.

	orig	wbr	msh	caps	punct	all
English						
S1	27,9	28,4 (0,5)	28,6 (0,7)	28,5 (0,6)	28,9 (1,0)	31 (3,1)
S2	27,6	28,3 (0,7)	28,5 (0,9)	28,1 (0,5)	29,4 (1,8)	32 (4,4)
S3	25,8	26 (0,2)	26,5 (0,7)	26,3 (0,5)	27,9 (2,1)	29,4 (3,6)
German						
S1	23,3	25,3 (2,0)	24,4 (1,1)	23,6 (0,3)	23,6 (0,3)	27,6 (4,3)
S2	21,4	23,1 (1,7)	22,3 (0,9)	21,4 (0,0)	21,8 (0,4)	24,8 (3,4)
S3	19	19,9 (0,9)	19,8 (0,8)	19,1 (0,1)	19,3 (0,3)	21,3 (2,3)
French						
S1	26,3	27,9 (1,6)	28 (1,7)	27,4 (1,1)	26,8 (0,5)	32 (5,7)
S2	23,4	24,8 (1,4)	24,1 (0,7)	23,8 (0,4)	24,3 (0,9)	27,5 (4,1)
S3	22,2	22,6 (0,4)	23 (0,8)	22,4 (0,2)	23 (0,8)	24,8 (2,6)

Table 4: BLEU scores calculated using reference translations (case-sensitive)

5. Quality Evaluation without Human-Translated Reference Set

In the previous section we estimated the impact of different types of noise on the quality of machine translation in a usual way, when a set of automatically translated texts is compared with a human-created reference translation. This procedure is expensive and time-consuming. Besides, for each language pair we need a translator who would be fluent in both languages.

While it is relatively easy to find such a person for translating between two major languages, it may become a hard task when both source and target languages are not widespread. Is it possible to bypass this problem?

5.1 Using Monolingual BLEU

We compared the time needed for a qualified translator to translate a set of 100 randomly selected queries from English into Russian and to correct all types of spelling errors in the same set of English queries. The first task took four times as much time as the second one. Note that for the task of correction the editor need not to be as qualified as for the translation task: s/he does not need to master both languages, but only the source language, and may not even be a native speaker of it. So correcting a text

is at least four times cheaper than translating it (more than four times if we take into account that we need less qualified manpower).

As we have seen in Section 4.2, the translations of the fully corrected sets invariably had the highest quality scores. So we can count BLEU scores of test sets in source languages, taking the fully corrected sets as references. This metric, the so-called monolingual BLEU (*mBLEU*), is often used to evaluate the result of text correction (see e.g. Park and Levy, 2011) or pre-ordering in machine translation (Navrátil et al., 2012). If the difference in *mBLEU* between sets with different types of corrected errors correlates with the difference in BLEU between translations of the same sets, we can use only monolingual BLEU and thus avoid the task of creating reference translations.

To account for the difference between translations of fully corrected sets and translations of sets containing spelling errors, we introduce a P_{BLEU} value, which is counted as:

$$P_{BLEU} = \frac{BLEU}{BLEU_{all}} * 100\%$$

where $BLEU_{all}$ is the BLEU score of the test set with all types of spelling errors corrected. In other words, P_{BLEU} is a percentage of BLEU of a given translation from the best possible machine translation.

Table 5 and Figure 1 show the correlation between P_{BLEU} and *mBLEU* scores calculated for each language.

	orig	wbr	msh	caps	punct	all
English						
	87,3	87,8 (0,5)	90 (2,7)	90,8 (3,5)	91,7 (4,4)	100 (12,7)
German						
	86,7	87 (0,3)	98,3 (11,6)	91,2 (4,5)	88,6 (1,9)	100 (13,3)
French						
	82,3	82,7 (0,4)	91,8 (9,5)	85,6 (3,3)	84,5 (2,2)	100 (17,7)

Table 5. *mBLEU* scores (case-sensitive)

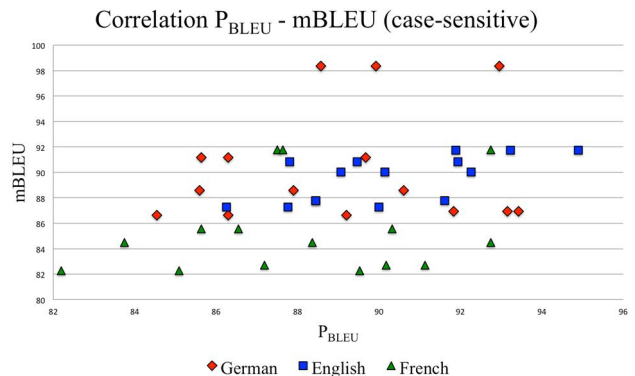


Figure 1. P_{BLEU} vs. *mBLEU* scores

The usual way to analyze the performance of an evaluation metric in machine translation is to compute the correlation between the automatic metric and human judgments (Papineni et al., 2002; Koehn, 2004; Lin and Och, 2004; Stent et al., 2005). Here we computed the Pearson correlation coefficients between $mBLEU$ and P_{BLEU} . The result we have got is 0,75, that is, the correlation is high enough (see Table 7). We also computed the Spearman correlation, which is lower (0,58, see Table 8).

5.2 Using Machine-Created Translations as a Reference Set

We have seen that it is possible to estimate approximately the expected improvement of translation considering only the difference between original sets of texts. However, correction of an error does not necessarily lead to the improvement of the translation result. This may occur in cases of minor errors, such as, for instance, missing diacritics. So French *vetement* ('clothes', instead of *vêtements*) or *voila* ('here is', instead of *voilà*) are translated correctly by all three translation systems. Since plain letters and letters with diacritics are considered as different symbols, this difference has been reflected in monolingual BLEU scores, and we would expect an improvement in translation, but this is not the case. Some more serious errors may also be processed correctly, cf. a German example *Sonnenenergie* (instead of *Sonnenergie*), which, despite two extra letters, has been correctly translated by one of translation systems as 'Solar energy'. Note also that different systems may deal with one and the same error differently. In the following example substitution of the first lowercase letter with the capital changed the translation by S2, but not the translation by S3:

French

- (a) *ils dépriment parce qu'ils ont manqué de soleil.*
they depressed because they missed the sun. (S2)
 They depressed because they lack of Sun. (S3)
- (b) *Ils dépriment parce qu'ils ont manqué de soleil.*
They get depressed because they missed the sun. (S2)
 They depressed because they lack of Sun. (S3)

Therefore, for practical reasons it does not make so much sense to estimate an improvement in translation irrelative of a particular translation system. This improvement may depend on the translation system and, maybe, on the target language. Thus it should be done not before the translation, but after it.

To solve this problem, we propose to use a pseudo-reference translation which is the result of machine translation of the fully corrected set instead of the human-created set of reference translations. In this case the impact of each type of spelling errors on the quality of translation will be measured as the difference in quality between the translation of sets containing this type of errors and the best possible result.

We call the value which shows the relation between the given translation and the best possible one " $aBLEU$ " (from a[utomated reference]BLEU). The correlation between case-sensitive $aBLEU$ and P_{BLEU} scores for all systems and for all translation directions are shown in Table 6 and on Figure 2.

In Tables 7 and 8 we present the correlations between P_{BLEU} , $mBLEU$ and $aBLEU$ for each translation direction.

	orig	wbr	msp	caps	punct	all
English						
S1	83,1	85,5 (2,4)	86,1 (3,0)	85,9 (2,8)	89,9 (6,8)	100 (16,9)
S2	82,5	85 (2,5)	86,2 (3,7)	84,7 (2,2)	89,5 (7,0)	100 (17,5)
S3	82,4	84 (1,6)	85,9 (3,5)	84 (1,6)	91,1 (8,7)	100 (17,6)
German						
S1	77,8	84,9 (7,1)	83,6 (5,8)	79,7 (1,9)	80,4 (2,6)	100 (22,2)
S2	71,2	81,3 (10,1)	77 (5,9)	73,2 (2,0)	74,8 (3,6)	100 (28,8)
S3	76,4	82,7 (6,3)	82,1 (5,7)	77,8 (1,4)	80,3 (3,9)	100 (23,6)
French						
S1	75,2	79,7 (4,5)	84,5 (9,3)	79 (3,8)	77,6 (2,4)	100 (24,8)
S2	71,7	78,6 (6,9)	79,8 (8,1)	75,2 (3,5)	75,3 (3,6)	100 (28,3)
S3	77,9	79,8 (1,9)	87,2 (9,3)	79,7 (1,8)	82,2 (4,3)	100 (22,1)

Table 6. $aBLEU$ scores (case-sensitive)

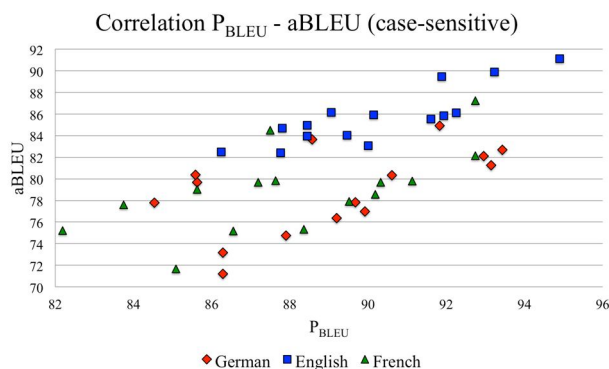


Figure 2. P_{BLEU} vs. $aBLEU$ scores

	P_{BLEU}			
	English	German	French	All
$mBLEU$	0,92	0,62	0,80	0,75
$aBLEU$	0,96	0,89	0,91	0,90

Table 7. Pearson correlation between P_{BLEU} , $mBLEU$ and $aBLEU$

	P_{BLEU}			
	English	German	French	All
$mBLEU$	0,80	0,48	0,56	0,58
$aBLEU$	0,89	0,72	0,81	0,76

Table 8. Spearman correlation between P_{BLEU} , $mBLEU$ and $aBLEU$

As we see, the correlation between P_{BLEU} and $aBLEU$ is significantly higher than the correlation between P_{BLEU} and $mBLEU$: average 0,90 against 0,75 (Pearson correlation) and 0,76 against 0,58 (Spearman correlation). One of the reasons for this is that machine translation systems handle some errors automatically, so that these errors are reflected in $mBLEU$, but have no effect on translation result, which is shown by $aBLEU$.

Note also that for German and French difference between $mBLEU$ and $aBLEU$ is much higher than for English. Evidently these two languages have more minor spelling errors which do not affect translation; among them may be missing diacritical marks (see examples in Section 5.2). So we may conclude that $aBLEU$ is a reliable technique to estimate the impact of pre-editing on the quality of machine translation. At the same time, $aBLEU$ requires less human work than the usual BLEU metrics.

6. Conclusion

In this paper we analyzed how the correction of different types of spelling errors affects the quality of machine translation. While correction of all errors can raise the BLEU score by 10 to 15% relative, the impact of particular types of errors is different. We found out that misspellings and wrong punctuation affect translation results more than other types of errors. Correction of these two error types then would be most profitable for improving the translation quality.

We also proposed a method to evaluate the change in machine translation results without using human-created reference translations. This method consists of creating a fully corrected text in the source language, automatically translating it and then using it as a reference for measuring BLEU score. This procedure gives us a reliable correlation with the usual measuring methods and at the same time saves a lot of human translators' time. The proposed method could be used to evaluate the performance of spelling correction in automatic pre-editing.

References

- Aikawa, T.; Schwartz, L.; King, R.; Corston-Oliver, M. and Lozano, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. In *Proceedings of MT Summit XI*, pp. 1--7. Copenhagen, Denmark, 11 September 2007.
- Banerjee, P.; Naskar, S. K.; Roturier, J.; Way, A. and van Genabith, J. (2012). Domain Adaptation in SMT of User-Generated Forum Content Guided by OOV Word Reduction: Normalization and/or Supplementary Data. In *Proceedings of the 16th Annual Meeting of the European Association for Machine Translation*, pp. 169--176. Trento, Italy, 26--30 May 2012.
- Bertoldi, N.; Cettolo, M. and Federico, M. (2010). Statistical machine translation of texts with misspelled words. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 412--419. Association for Computational Linguistics, Los Angeles, California, USA, 2--4 June 2010.
- Carrera, J.; Beregovaya, O. and Yanishevsky, A. (2009). *Machine translation for cross-language social media*, http://www.promt.com/company/technology/pdf/machine_translation_for_cross_language_social_media.pdf.
- Clark, A. (2003). Pre-processing very noisy text. In *Proceedings of Workshop on Shallow Processing of Large Corpora*, pp. 12--22. Lancaster, UK, March 2003.
- Jiang, J.; Way, A. and Haque, R. (2012). Translating user-generated content in the social networking space. In *Proceedings of AMTA-2012, the Tenth Biennial Conference of the Association for Machine Translation in the Americas*. San Diego, USA, October 2012.
- Kirchhoff, K.; Rambow, O.; Habash, N. and Diab, M. (2007). Semi-automatic error analysis for large-scale statistical machine translation systems. In *Proceedings of the MT Summit XI*. Copenhagen, Denmark, 11 September 2007.
- Koehn, P. (2004). Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP, Conference on Empirical Methods in Natural Language Processing*, pp. 388--395. Barcelona, Spain, 25--26 July 2004.
- Lin, C.-Y. and Och, F. J. (2004). Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 605--612. Barcelona, Spain, 21--26 July 2004.
- Navrátil, J.; Visweswariah, K. and Ramanathan, A. (2012). A Comparison of Syntactic Reordering Methods for English-German Machine Translation. In *Proceedings of COLING 2012: Technical Papers, The 24th International Conference on Computational Linguistics*, pp. 2043--2058. Mumbai, India, 8--15 December 2012.
- Papineni, K.; Roukos, S.; Ward, T. and Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*,

- pp. 311--318. University of Pennsylvania, Philadelphia, Pennsylvania, USA, 7--12 July 2002.
- Park, Y. A. and Levy, R. (2011). Automated whole sentence grammar correction using a noisy channel model. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1:934--944. Portland, Oregon, USA, 19--24 June 2011.
- Plesco, C. and Rychtycky, N. (2012). Machine Translation as a Global Enterprise Service at Ford. In *Proceedings of AMTA-2012, the Tenth Biennial Conference of the Association for Machine Translation in the Americas*. San Diego, USA, October 2012.
- Roturier, J.; Mitchell, L.; Grabowski, R. and Siegel, M. (2012). Using Automatic Machine Translation Metrics to Analyze the Impact of Source Reformulations. In *Proceedings of AMTA-2012, the Tenth Biennial Conference of the Association for Machine Translation in the Americas*. San Diego, USA, October 2012.
- Shoukry, A. and Rafea, A. 2012. Preprocessing Egyptian Dialect Tweets for Sentiment Mining. In *Proceedings of the Fourth Workshop on Computational Approaches to Arabic Script-based Languages, AMTA-2012*, pp. 47--56. San Diego, USA, October 2012.
- Stent, A.; Marge, M. and Singhai, M. (2005). Evaluating evaluation methods for generation in the presence of variation. In *Proceedings of CICLING, Conference on Computational Linguistics and Natural Language*, pp. 341--351. Mexico City, Mexico, 13--19 February 2005.
- Subramaniam, L. V.; Roy, Sh.; Faruque, T. A. and Negi S. 2009. A survey of types of text noise and techniques to handle noisy text. In *Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data*, pp. 115--122. Barcelona, Spain, 23--24 July 2009.