Assessing Machine Translation Quality with Error Analysis

Maarit Koponen
University of Helsinki
Department of Modern Languages

Abstract

Translation quality can be evaluated with regard to different aspects, such as accuracy (fidelity), fluency and fitness for purpose. In using a machine translation system for information purposes, accuracy of semantic content is the key aspect of quality. Automated quality metrics developed in the machine translation field have been criticized for conflating fluency of form with accuracy of content and for failing to provide any information on the types of errors in the translations. Our research aims to discover criteria for assessing translation quality specifically in terms of accuracy of semantic content in translation. This paper demonstrates how an error analysis with a view to identifying different error types in machine translations can serve as a starting point for such criteria. The error classification described focuses on mismatches of semantic components (individual concepts and relations between them) in the source and target texts. We present error analysis results, which show differing patterns both between human translators and machine translation systems on the one hand and two different kinds of translation systems on the other.

Keywords: machine translation, machine translation quality evaluation, error analysis

1 Introduction

Translation quality assessment is important in the context of both human translation, where it may be used for quality control in professional settings or in translator training, and machine translation, where it may be used by the developer or potential user to evaluate system performance. Nevertheless, translation quality remains an elusive concept with no one universal definition of quality or one generally accepted method for quality assessment. Quality assessment involves various aspects, such as accuracy (fidelity), fluency and fitness for purpose, and different aspects have been deemed important for different situations.

Human translation assessment (see Secară 2005; Williams 2001) has been moving from microtextual, word- or sentence-level error analysis methods toward more macrotextual methods focused on the function, purpose and effect of the text. At the same time, machine translation assessment has mainly been microtextual and focused on the aspects of accuracy and fluency (e.g. LDC 2005). In addition to methods involving human evaluators, the machine translation field has developed automated metrics, such as the widely used BLEU metric (Papineni et al. 2002). Automated quality metrics are generally based on a statistical comparison of the machine translation to one or more reference translations produced by human translators. Such metrics have been claimed to correlate well with human assessments of accuracy and fluency but they are not without problems. Studies have shown that a higher score by the metric does not guarantee better translation quality (see Callison-Burch et al. 2006, for example).

The statistical comparison metrics have been argued to only measure superficial similarity and to conflate fluency of form with accuracy of content. Since translating for information purposes is a common and possibly the most fruitful use of machine translation, semantic accuracy should perhaps be the first and foremost concern over fluency. While attempts have been made to include semantic features into quality metrics (Giménez and Màrquez 2007; Padó et al. 2009), the need for criteria geared toward semantic accuracy remains.

As a first step, more detailed information about different types of errors in machine translated texts is necessary for developing semantic quality criteria. Such information should separate form and function, and focus on errors affecting the accuracy of meaning (translation errors) over errors related to fluency only (language errors). Analysis of the effect of errors on meaning will be important as not all errors are equally critical. Even some changes to semantic content on the word or sentence level may not be destructive to transfer of meaning on the text level, because context and extra-linguistic knowledge may help the reader to identify errors and reconstruct meaning (see Bensoussan and Rosenhouse 1990; Jones et al. 2007). On the other hand, certain errors can lead to particularly significant misinterpretations, such as errors affecting the argumentation structure (Williams 2001) or semantic roles (Wu and Fung 2009).

This paper demonstrates how an error analysis of machine translation can serve as the starting point for discovering semantic criteria by helping to identify different error types. The error analysis presented here is a pilot study conducted on three text passages translated by two different types of machine translation systems and a comparison with human translations. Comparing the error results show that this kind of error analysis can bring out interesting differences between translations produced by humans and machines as well as between translations by different machine translation systems. In this way, the error definition discussed forms the first step in developing a classification of errors and analyzing the connection between errors and meaning. In further studies, the central aspect will be determining how different error types truly affect the preservation of the source text semantic content.

2 Materials

The material studied comprises three English source texts selected to represent different text types: a European Commission Green Paper (Commission of the European Communities 2009), an article from the National Geographic magazine (Hall 2008), and a software user guide (Symantec Corporation 2006b). Different genres were selected to observe whether machine translation accuracy varies with text type. The European Commission text and the magazine article contain fairly long and complex sentences, whereas the user guide has many short imperative sentences and sentence fragments. The magazine article has the most diverse vocabulary while the user guide has the least variation but contains many common nouns used as name-like terms (e.g. *Next* and *Finish* as the names of buttons).

From each source text, a passage of approximately 400 words (see Table 1) was taken for a detailed analysis. Each passage was then translated from English into Finnish using two different machine translation systems representing two different types of machine translation strategies: rule-based and statistical. The rule-based machine translation system, a demo by Sunda Systems Oy,[1] is based on thousands of hand-coded lexical and grammatical rules. The statistical system, Google Translate[2] by Google Inc., is based on the use of statistical learning methods on large monolingual and parallel corpora to create a language model and a bilingual phrase table for translation. In addition to the machine translations, the published Finnish translations of the source texts (Euroopan yhteisöjen komissio 2009; Hall 2009; Symantec Corporation 2006a) were also analyzed for comparison.

Table 1 Length of source text passages for analysis

	Number of words (tokens)	**Number of sentences**	**Words per sentence**	**Type-token ratio**
Green Paper	464	14	33	0.38
User guide	411	38	11	0.33
Magazine	455	17	26	0.54

3 Method

With the aim of drafting an error classification emphasizing semantic accuracy, a translation error was roughly defined as “semantic component not shared by source text and target text”. Here, semantic components refer to individual concepts and the semantic relations between two concepts (head and dependent). Concepts are represented by content words and they can be units larger than individual words, for example in the case of compound nouns, names and idioms. Relations are expressed through function words, inflection and word order, for instance.

The unit of analysis was set to the sentence level because that is the largest unit processed by the machine translation systems, and source and target text sentences can therefore be expected have a one-to-one correspondence. In cases where the human translator had split a source text sentence into two or more target text sentences, these sentences were treated as one. Similarly, if the translator had combined two or more source text sentences into one target text sentence, these were split to correspond to the source text. Kernel analysis and back-transformation of the source and target sentences was used as a way to break complex sentences into smaller units and identify the components present in each sentence.

The semantic components were then compared to identify differences between the source text and the three target text versions (human translation and two machine translations). In comparing concepts, lexical choices were assessed as acceptable if they conveyed the correct meaning regardless of whether this choice was the most frequent or idiomatic one. In comparing relations, a relation was considered to be present in the target sentence only if it could be parsed without hesitation.

At first, mismatches between source and target concepts were divided into four categories: **omitted**, **added**, **mistranslated** and **untranslated**. Example 1 demonstrates each of these cases. The Finnish machine translation contains no equivalents for *places*, which is therefore classified as omitted. On the other hand, the source sentence contains no match for the Finnish *henkilöllä* ‘person’, which is therefore classified as added. A mistranslated concept is seen in *inhimillisen* ‘human (adj.)’ or ‘humane’, which would be a possible translation for *human* in the adjectival sense of ‘characteristic of humans’ but not for the noun *human* in the sense of the human species. Untranslated concepts are source language words (other than names) appearing the target, such as *locale* in Example 1.

> Example 1.
> The locale places them at one of the most important geographical intersections of prehistory, and the date puts them squarely at the center of one of the most enduring mysteries in all of human evolution. (Hall 2008)
> Locale henkilöllä on yksi tärkeimmistä maantieteellinen leikkauspisteistä esihistoriaan, ja päivä tuo ne reilusti keskelle yksi pysyviä salaisuuksia kaikilla inhimillisen kehityksen. (Statistical MT)
> [Locale person has one of the most important geographic of intersections into prehistory, and the day brings them squarely into the center one enduring secrets on all humane development’s.]

This classification was refined during the analysis, as it did not adequately account for certain acceptable lexical choices. Firstly, in the software user guide (Symantec Corp. 2006a) the verb *click* is often replaced with a more generic concept *valita* ‘select’. The words do not appear to be equivalent as such, and according to the strict classification described above, this would count as two errors (omission of *click* and addition of *valita*). However, translators and writers of comparable texts often use *valita* to convey this meaning in this context, and it is a valid substitution rather than an error. Secondly, some additions do not add new information, such as superordinate concepts added as grammatical support words. An example is adding the *ohjelma* ‘program’ to the name *Norton AntiVirus,* which would be awkward to use in Finnish without this addition. These observations led to the forming of two new categories and reclassification of the relevant cases as **substitution** or **explicitation**.

The final concept error categories are as follows:

- **Omitted concept**: ST concept that is not conveyed by the TT (Example 1: *places*).
- **Added concept**: TT concept that is not present in the ST (Example 1: *henkilöllä* ‘person’).
- **Untranslated concept**: SL words that appear in TT (Example 1: *locale*).
- **Mistranslated concept**: A TT concept has the wrong meaning for the context (Example 1: *inhimillinen kehitys* ‘humane development’ for *human evolution*).
- **Substituted concept**: TT concept is not a direct lexical equivalent for ST concept but can be considered a valid replacement for the context (*valitse* ‘select’ for *click*).

- **Explicitated concept**: TT concept explicitly states information left implicit in ST without adding information (addition of *ohjelma* ‘program’ to *Norton AntiVirus*).

Mismatches in relations between concepts were originally divided into five categories: **omitted**, **added**, **mistaken relation**, **mistaken dependent** and **mistaken head**. In Example 1, all of the relations involving *places* are omitted due to the concept being omitted, and the relation between *into the center* and *(of) one* becomes omitted because the erroneous word form (nominative instead of partitive) of *yksi* ‘one’ renders it unparseable although both concepts are present. Again in Example 1, an added possessive relation forms between the added concept *henkilöllä* ‘person’ and *yksi* ‘one’ in the beginning of the sentence, while an added relation without added concepts appears in Example 2 where the hyphen between *AntiVirus* and *CD-levyltä* ‘from a CD disk’ makes *AntiVirus* a modifier of *CD*. Example 3 shows a mistaken relation where the predicate/object relation between the transitive verb *click* and noun *Next* becomes a temporal relation between the adverbial *seuraavaksi* ‘next (adv.)’ or ‘subsequently’ and intransitive verb *napsahtavan* ‘to click’, and a mistaken dependent when the infinitival form of *napsahtavan* makes *tilanneilmaisimen* ‘progress bar’ its subject instead of the second person represented by the imperative *click*.

Example 2.
You can install Norton AntiVirus from a CD or from a file that you download. (Symantec Corp. 2006b)
Voit asentaa Norton AntiVirus-CD-levyltä tai tiedoston lataamista. (Statistical MT)
[You can install from Norton AntiVirus-CD or downloading of file.]

Example 3.
In the Activation panel, wait for the progress bar to stop, and then click Next. (Symantec Corp. 2006b)
Aktivointipaneelissa odottakaa tilanneilmaisimen pysähtyvän ja sitten napsahtavan seuraavaksi. (Rule-based MT)
[In the activation panel wait for the progress bar to stop and then to click subsequently.]

As with concepts, not all changes of relations necessarily lead to changed semantic content. Firstly, one of the participants in a relation (the head or dependent) could be substituted either by a substituted concept (see the concept classification) or another concept referring to the same entity. For example, the translator EC text (Commission of the European Communities 2009) had chosen to translate *citizens who are nationals of a significant number of Member States* with *kansalaista merkittävästä määrästä jäsenvaltioita* ‘citizens of/from a significant number of Member States’ replacing both *citizens* and *nationals* with the same word and making *of a significant number of Member States* modifier of *citizens* instead of *nationals*. Secondly, a changed relation may sometimes be acceptable. For example, in the same EC translation *The Treaty of Lisbon --- amends the Treaty on European Union and the Treaty establishing the European Community* had been translated as *Lissabonin sopimuksella muutetaan sopimusta Euroopan unionista ja Euroopan yhteisön perustamissopimusta* ‘with the Treaty of Lisbon, the Treaty on European Union and the Treaty establishing the European Community are amended’. Although *Lissabonin sopimuksella* is an adjunct expressing the instrument of the passive verb, the meaning remains the same because in terms of semantic roles, *the Treaty of Lisbon* is in fact an instrument in the source text, as well. This particular substitution is a well-known translational contrast between subject-prominent English and less subject-prominent languages. According to these observations, two new categories were established for **substituted participants** and **substituted relations**.

Further changes of the relation error categories were also deemed necessary. Firstly, it was considered more informative to divide omission and addition according to whether the omission or addition was due to an omitted or added concept or not. Secondly, the categories of mistaken head and mistaken dependent were combined into one category for mistaken participants.

The final relation error categories are as follows:

- **Omitted participant**: ST relation not conveyed by the TT due to an omitted head or dependent (Example 1: all relations involving the omitted concept *places*).
- **Omitted relation**: ST relation not conveyed by the TT due to morpho-syntactic errors that prevent parsing the relation although both concepts are present in TT (Example 1: relation between *the center* and *one* cannot be parsed in Finnish).
- **Added participant**: TT relation not present in ST introducing an added concept (Example 1: addition of *henkilöllä* ‘person’ and *henkilöllä on yksi* ‘the person has one’).
- **Added relation**: TT relation not present in ST arises due to morpho-syntactic errors (Example 2: *Norton AntiVirus* as a modifier of *CD-levyltä*).
- **Mistaken participant**: Head or dependent of the relation different in ST and TT, not same entity (Example 3: *tilanneilmaisimen* ‘progress bar’ as subject of *napsahtavan* ‘to click’).
- **Mistaken relation**: Relation between two concepts different in ST and TT, changed role (Example 3: temporal relation of *seuraavaksi* ‘next’ and *napsahtavan* ‘to click’).
- **Substituted participant**: Head or dependent of the relation different in ST and TT, same entity (one TT concept, e.g. *kansalaista* ‘citizens’ or ‘nationals’, replaces two in ST, e.g. *citizens* and *nationals*).
- **Substituted relation**: Relation between two concepts different in ST and TT, same semantic roles (*Lissabonin sopimuksella muutetaan* ‘with the Treaty of Lisbon is amended’ for *Treaty of Lisbon amends*).

Classifying errors into the categories listed above was for the most part straightforward. Sometimes errors in one sentence could have been analyzed in multiple ways, in which case the analysis with the lowest total error count was selected. The most difficult classifications were related to accepting concepts and relations as substitutions. In borderline cases, unclear cases were counted as errors, but as these decisions were based on intuitive assessment, they remain somewhat subjective. It would be interesting to see how resources such as domain ontologies could support these assessments.

4 Results

The numbers of errors classified into each category are shown in Tables 2 and 3. The rule-based translation system (RBMT) made a total of 289 errors, of which 121 (42%) were related to concepts and 168 (58%) to relations between concepts. The statistical system (SMT) made nearly twice as many errors, 516 in total, of which 163 (32%) were related to concepts and 353 (68%) to relations. For the different text types, the rule-based system had similar amounts of concept errors in each text, while the statistical system had about twice the number of errors in the magazine article than in the other two translations. For relations, both systems made the least errors with the simpler sentences in the software user guide and the most in the magazine article. The difference is large especially for the statistical system.

When examined more closely, the numbers show quite different patterns for the two types of translation systems. For the rule-based system, the most typical error is mistranslating an individual concept (38% of all errors) and the second most common is omitting a relation (32%). There are hardly any omitted or added concepts and thereby hardly any omitted or added participants to a relation. For the statistical system, the most common errors by far are omitted relations (42%), while other types show more even distribution.

Table 2 Errors related to individual concepts

		Omitted concepts	**Added concepts**	**Mistranslated concepts**	**Untranslated concepts**	**Concept errors total**
RBMT	Green Paper	1	0	34	0	35
	User guide	1	1	35	6	43
	Magazine	0	0	41	2	43
	Total	**2**	**1**	**110**	**8**	**121**
SMT	Green Paper	24	8	8	0	40
	User guide	17	5	14	7	43
	Magazine	16	11	34	19	80
	Total	**57**	**24**	**56**	**26**	**163**
Human ref.	Green Paper	18	7	0	0	25
	User guide	20	21	0	3	44
	Magazine	36	6	0	0	42
	Total	**74**	**34**	**0**	**3**	**111**

Table 3 Errors related to relations between concepts

		Omitted rel.	Omitted part.	Added rel.	Added part.	Mistaken rel.	Mistaken part.	Relation errors total
RBMT	Green Paper	31	1	8	0	4	15	59
	User guide	19	1	4	0	9	6	39
	Magazine	42	0	7	0	12	9	70
	Total	**92**	**2**	**19**	**0**	**25**	**30**	**168**
SMT	Green Paper	49	21	9	3	8	17	107
	User guide	54	19	6	0	3	13	95
	Magazine	113	13	3	6	5	11	151
	Total	**216**	**53**	**18**	**9**	**16**	**41**	**353**
Human ref.	Green Paper	5	14	2	13	1	1	36
	User guide	3	21	2	23	1	1	51
	Magazine	2	45	3	11	0	1	62
	Total	**10**	**80**	**7**	**47**	**2**	**3**	**149**

The corresponding total number for the human translated texts is 260 errors, of which 111 (43%) are related to concepts and 149 (57%) to relations. Compared to the two machine translations, the pattern that emerges is again quite different. With the exception of a few untranslated terms, all concept errors relate to omitted (28% of total) or added concepts (13% of total), and the vast majority of the relation errors are a direct consequence of these omissions and additions: omitted or added participants (31% and 18% of total, respectively). Unlike the machine translation systems, the human translations show hardly any cases where both participants of the relation are present but the relation is missing, or where the participant or relation is different in source and target texts.

To further illustrate the different patterns produced by the human translators and machine translation systems, we can examine the cases of omission and addition that were reclassified as substitutions (Table 4). The human translations contained considerably more substitutions of both concepts and participants than either of the two machine translation systems. Substitutions of any kind were very rare for the rule-based system and somewhat more common for the statistical. However, about half of the instances of substituted concepts and substituted participants in the translations made with the statistical translation system come from the same text (the software user guide) and involve one verb (*click* → *valita* discussed above). For the human translator, substitutions were more frequent in the magazine article and European Commission text while the software guide appeared to be translated more literally.

Table 4 Substitutions

		Substituted concepts	Explicitated concepts	Substituted relations	Substituted participants	Total substitutions
RBMT	Green Paper	0	3	0	0	3
	User guide	1	0	2	0	3
	Magazine	0	4	0	2	6
	Total	**1**	**7**	**2**	**2**	**12**
SMT	Green Paper	7	3	3	6	19
	User guide	12	0	2	24	38
	Magazine	1	1	0	1	3
	Total	**20**	**4**	**5**	**31**	**60**
Human ref.	Green Paper	23	4	7	44	78
	User guide	10	1	7	29	47
	Magazine	28	6	4	52	90
	Total	**61**	**11**	**18**	**125**	**215**

5 Discussion

The results of the error analysis reveal different patterns for the two machine translation strategies, and the effect on the translations are different. With the statistical system, the most common error of omitting the relation between two concepts even when both concepts are present in the translation as well as omitting and adding concepts may produce incomprehensible "word salad" consisting of unconnected concepts. With the rule-based system, on the other hand, the most common error of mistranslating an individual concept as well as mistaken relations may produce superficially more convincing but misleading sentences. The different patterns also mean that while the statistical system made more errors overall, it is difficult to state conclusively which translation preserved more of the source text semantic content. This is because this analysis treats all errors as equally critical even though, as noted in the introduction, some errors may be more and some less destructive to the transfer of meaning and the reader's ability to recover it. A deeper analysis of the effect and criticalness of different error types would be needed for assessing the preservation of semantic content.

In contrast to human translators, the errors typical of machine translation systems are mistranslations as well as various errors related to the relations between concepts even in cases where the concepts themselves may be translated correctly. In the human translation, omitted and added relations seem to result directly from omitting or adding concepts. Although it is not evident in the numbers, a closer examination of the texts also reveals that the cases classified as omissions and additions in the human translations differ from those in the machine translations. Most noticeably, none of the concepts added by human translators are completely unrelated to the source text in the way of the machine translated Example 1 in Section 3. Furthermore, machine translated sentences may contain several unconnected omissions and additions, whereas in human translated sentences, multiple omissions or additions generally result from the omission or addition of an entire clause. Often this involves information reordering or compensation across sentence limits, explicitation (explicitly stating information left implicit in the source text) or implicitation (leaving information to be inferred from context). The error definition ("semantic component not shared by source text and target text") and the sentence-level unit analysis make it difficult to account for such strategies. However, since machine translation systems do not utilize them, deviations from the source text should at least be flagged as errors in machine translations.

6 Conclusions

The error analysis presented in this paper reveals interesting differences between human translations and machine translations as well as two different types of machine translation systems. In this way, it serves as a first step in developing an error classification with a view to developing semantic quality criteria. While the error analysis succeeds in uncovering differences, it is only a preliminary step. Firstly, the analysis does not yet account for the real effect of different error types, and a deeper analysis is needed to assess how different types of errors affect preservation of semantic content. Secondly, differentiating true omissions from implicitation, true additions from explicitation and mistakes from substitutions calls for a more refined and semantically-oriented analysis and will be an important aspect in assessing semantic quality.

Future work will therefore be focused on semantic aspects and the connection between errors and semantic content. As first steps, this will involve evaluating the criticalness of different types of errors with the help of reading comprehension tests performed on machine translated texts. This test will help to assess the "error tolerance" of the reader and the relationship between different types of errors, argumentation structure and reading comprehension. We are also investigating ways to operationalize our semantic quality criteria and exploring the use of computational resources such as automatic parsing and WordNet as a way to make semantic evaluation more objective and less labor intensive.

Acknowledgements

This work has been funded through the LANGNET Finnish Graduate School in Language Studies.

References

Bensoussan, Marsha & Judith Rosenhouse 1990. Evaluating student translations by discourse analysis. *Babel* 36:2, 65–84.

Callison-Burch, Chris, Miles Osborne & Philipp Koehn 2006. Re-evaluating the Role of BLEU in Machine Translation Research. *EACL-2006: 11th Conference of the European Chapter of the Association for Computational Linguistics*, 249–256.

Commission of the European Communities 2009. Green paper on a European Citizens' Initiative [online]. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0622:FIN:EN:PDF [accessed 30 August 2010].

Euroopan yhteisöjen komissio 2009. Vihreä kirja eurooppalaisesta kansalaisaloitteesta [online]. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0622:FIN:FI:PDF [accessed 30 August 2010].

Giménez, Jesús & Lluís Màrquez 2007. Linguistic features for automatic evaluation of heterogenous MT systems. *StatMT '07: Proceedings of the Second Workshop on Statistical Machine Translation,* 256–264.

Hall, Stephen S. 2008. Last of the Neanderthals [online]. Available at: http://ngm.nationalgeographic.com/2008/10/neanderthals/hall-text [accessed 30 August 2010].

Hall, Stephen S. 2009. Viimeinen neandertalilainen [online]. Available at: http://natgeo.fi/neandertalilaisen-periman-kartoitus [accessed 30 August 2010].

Jones, Douglas, Martha Herzog, Hussny Ibrahim, Arvind Jairam, Wade Shen, Edward Gibson & Michael Emonts 2007. ILR-based MT comprehension test with multi-level questions. *NAACL '07: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers on XX,* 77–80.

LDC 2005. Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Translations Revision 1.5, January 25, 2005 [online]. Available at: http://projects.ldc.upenn.edu/TIDES/Translation/TransAssess04.pdf [accessed 30 August 2010].

Padó, Sebastian, Daniel Cer, Michel Galley, Dan Jurafsky & Christopher Manning 2009. Measuring machine translation quality as semantic equivalence: A metric based on entailment features. *Machine Translation* 23:2–3, 181–93.

Papineni, Kishore, Salim Roukos, Todd Ward & Wei-Jing Zhu 2002. BLEU: a method for automatic evaluation of machine translation. *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics,* 311–318.

Secară, Alina 2005. Translation evaluation: A State of the Art Survey. *Proceedings of the eCoLoRe/MeLLANGE Workshop,* 39–44. Available at: http://ecolore.leeds.ac.uk/ [accessed 30 August 2010].

Symantec Corporation 2006a. Norton AntiVirus Käyttöohje [online]. Available at: ftp://ftp.symantec.com/public/finnish/products/norton_antivirus/ver2007/manuals/nav2007_user_guide.pdf [accessed 30 August 2010].

Symantec Corporation 2006b. Norton AntiVirus User Guide [online]. Available at: ftp://ftp.symantec.com/public/english_us_canada/products/norton_antivirus/2007/ manuals/navr1ug.pdf [accessed 30 August 2010].

Williams, Malcolm 2001. The application of argumentation theory to translation quality assessment. *Meta* 46:2, 326–44.

Wu, Dekai & Pascale Fung 2009. Can Semantic Role Labeling Improve SMT? *Proceedings of the 13th Annual Conference of the EAMT,* 218–225.

---

[1] Available at: http://www.sunda.fi/eng/translator.html. The version used to generate the translations for this study is an online demo and does not include all the functionalities of the commercial Sunda translation system.

[2] http://translate.google.com