An Evaluation of the Accuracy of Online Translation Systems

Milam Aiken
University of Mississippi

Kaushik Ghosh
University of Mississippi

John Wee
University of Mississippi

Mahesh Vanjani
University of Mississippi

An Evaluation of the Accuracy of Online Translation Systems

Milam Aiken
University of Mississippi
USA
aiken@bus.olemiss.edu

Kaushik Ghosh
University of Mississippi
USA
kghosh@bus.olemiss.edu

John Wee
University of Mississippi
USA
jwee@bus.olemiss.edu

Mahesh Vanjani
Texas Southern University
USA
vanjanim@tsu.edu

Abstract

Until fairly recently, translation among a large variety of natural languages has been difficult and costly. Now, several free, Web-based machine translation (MT) services can provide support, but relatively little research has been conducted on their accuracies. A study of four of these services using German-to-English and Spanish-to-English translations showed that Google Translate appeared to be superior. Further study using this system alone showed that while translations were not always perfect, their understandability was quite high.

INTRODUCTION

Expert human translation still surpasses the best results of machine translation (MT) systems (Bar-Hillel, 2003), but it is often hard to schedule an interpreter at the spur of the moment, especially for relatively obscure languages. Several free, fully automatic, Web-based translation services are available to fill this need but at the expense of lower accuracy. However, many translations do not need to be perfect. For example, a reader of a Web page or an email message written in a foreign language might need to get only the gist of the passage before deciding whether more detailed, human translation is needed or the content is not important enough to proceed further with it. That is, poor accuracy quickly can have greater value than higher accuracy that is too late (Muegge, 2006). As a result, more words are now translated per year using MT than are translated by human translators, and the demand continues to grow (LISA, 2009).

Few studies have been conducted on the relative accuracies of these Web-based services, however. The purpose of this paper is to provide a performance overview of four leading MT systems provided on the Web and to further assess the accuracy of the best.

Prior Studies of Web-Based MT Systems

Machine translation was first proposed in 1947, and the first demonstration of a translation system was in January 1954 (Hutchins, 2003). MT became available for personal computers in 1981, and in 1997, *Babel Fish* (using *SYSTRAN*) appeared as the first, free, translation service on the World Wide Web (Yang & Lange, 1998).

Although several evaluation studies have been conducted on MT systems (e.g., NIST, 2008), based upon an extensive review of the literature, only a few have focused solely upon Web-based versions. For example, four have tested the accuracy of *SYSTRAN* (originally provided at http://babelfish.altavista.com/babelfish - now: http://babelfish.yahoo.com/):

>   **Study 1** (Aiken, Rebman, Vanjani, & Robbins, 2002): In one of the earliest studies of a Web-based MT system, four participants used *SYSTRAN* to automatically translate German, French, and English comments in an electronic meeting. After the meeting, two objective reviewers judged the overall accuracy of the translations to be about 50% while the understanding accuracy was about 95%.
>
>   **Study 2** (Aiken, Vanjani, & Wong, 2006): In another study, a group of 92 undergraduate students evaluated *SYSTRAN* translations of 12 Spanish text samples to English, and they were not able to understand only two of the 12 translations (83% accuracy). No significant differences in understandability were found based on gender, but those who reported understanding some Spanish were able to understand many of the translations to English better. Further, the accuracy did not seem to correlate with the complexity of the sentences.
>
>   **Study 3** (Yates, 2006): In a third study, 20 sentences (10 Spanish, 10 German) selected from Mexican and German civil codes and press releases from foreign ministries were translated to English with *SYSTRAN*, and the author evaluated the samples' accuracies. The system's performance was rated as poor, but it was not uniformly poor, i.e., German texts were translated less poorly than the Spanish ones.
>
>   **Study 4** (Ablanedo, Aiken, & Vanjani, 2007): In a final study, 10 English text samples were translated by an expert and an intermediate-level Spanish translator as well as *SYSTRAN*. The most fluent human was 100% accurate, and the other achieved 80% accuracy. The MT system achieved only 70% accuracy but was 195 times faster than the humans.

All of these tests were based upon *SYSTRAN,* the system deemed most reliable at the time of the studies. However, new translation software on *Google* appeared in October 2007. Abandoning the rule-based algorithms of S*YSTRAN* which the site had used previously, *Google Translate* (http://translate.google.com/) focuses on statistical MT in which millions of words of equivalent

text are compared for matching probabilities (Geer, 2005). Statistical models of these language pairs are then used to translate new text. In a comprehensive evaluation of 20 MT systems translating Arabic to English and 17 systems translating Chinese to English (NIST, 2008), *Google Translate* was ranked the best. However, other than *Google*, free Web-based translation systems were not evaluated in the contest.

We have been able to find only three evaluations focused exclusively on free, Web-based translation services:

> **Study 1** (Bezhanova, Bezhandva, & Landry, 2005): In perhaps the first study of free, Web-based MT systems, 17 English sentences were translated into Spanish using *LogoMedia, SYSTRAN,* and *PROMT*. The authors concluded that all three of the MT systems produced usable translations, and that none has an obvious advantage. However, the *SYSTRAN* translations were generally the worst. In addition, the authors found that short sentences were translated very well, but many longer sentence translations were very difficult to understand.

> **Study 2** (Aiken & Wong, 2006): In another comparison of Web-based translation systems, a sample of 20 Spanish phrases from an introductory textbook were translated into English using four online services (*SYSTRAN*, *SDL*, *WorldLingo*, and *InterTran*). Results show that *SYSTRAN* and *WorldLingo* were the most accurate, followed by *SDL* and *InterTran*.

> **Study 3** (Kit & Wong, 2008): In a final study, *Google Translate, PROMT, SDL, SYSTRAN,* and *WorldLingo* were compared with BLEU (Papineni, Roukos, Ward, & Zhu, 2002) and NIST (Zhang, et al., 2004) scores using 13 languages (Arabic, Chinese, Dutch, French, German, Greek, Italian, Japanese, Korean, Portuguese, Russian, Spanish, and Swedish) and two text samples (Universal Declaration of Human Rights and European Union's Treaties). *SYSTRAN* was determined to be the best choice for many language pairs, especially for translations from Greek and Russian to English while *Google Translate* was better when translating Arabic and Chinese to English. *PROMT* worked better than others with Portuguese, and *WorldLingo* worked better for Swedish to English.

EVALUATION STUDIES

Four System Comparisons

Although *Google Translate* achieved outstanding results in one evaluation of MT systems using Chinese and Arabic and *SYSTRAN* was determined to be the best in three other system comparisons, we selected four online translation services listed at http://translation.langenberg.com/ (*Google Translate, Yahoo SYSTRAN, AppliedLanguage,* and *x10*) in another attempt to determine the most accurate on-line software. We restricted our analysis to German-to-English and Spanish-to-English translations only, and 10 sample phrases in German and Spanish were obtained from http://www.fodors.com/language/ for testing.

Table 1 shows the evaluations of two objective raters who reviewed 10 German-to-English and 10 Spanish-to-English translations of the same phrase with the four systems using a scale of 1 (bad) to 5 (good). Table 2 shows the evaluations per phrase, and Table 3 shows that the *Google* translations were judged as superior for Spanish, German, and overall. In addition, the German translations using all systems were generally viewed as more accurate than the Spanish.

Table 1: Evaluation of 10 sample phrases in Spanish and German translated to English (Most accurate translations shown in bold).

	Google	Rater 1	Rater 2	Yahoo-Systran	Rater 1	Rater 2	AppliedLanguage	Rater 1	Rater 2	x10	Rater 1	Rater 2
Could you help me?												
Podría ayudarme?	**It could help me?**	4	3	**It could help me?**	4	3	Podr�a to help me?	1	1	Would be able to help me?	3	1
Könnten Sie mir helfen?	**Could you help me?**	5	5	**Could you help me?**	5	5	K�nnten you me help?	1	1	**Could you help me?**	5	5
Thank you very much.												
Muchas gracias.	**Thank you very much.**	5	5	**Thank you very much.**	5	5	**Thank you very much.**	5	1	Many thanks	4	5
Vielen Dank .	**Thank you very much.**	5	5	Thank you.	5	5	Thank you.	4	5	**Thank you very much**	5	5
Do you speak English?												
Habla usted inglés?	**Do you speak English?**	5	5	Speech English you?	3	1	Speech you ingl�s?	1	1	Speak you English?	4	3
Sprechen Sie Englisch?	**Do you speak English?**	5	5	Do they speak English?	3	4	Do they speak English?	3	4	**Do you speak English?**	5	5
Let's go to the movies.												
Vamos al cine.	**We are going to the movies.**	4	4	We go to the cinema.	3	3	We go to the cinema	3	3	We go to the movies	3	4
Gehen wir ins Kino.	We go to the cinema.	3	3	We go into the cinema.	2	3	We go into the cinema	2	3	**Let's go into the movie**	5	4

										theater		
What's your name?												
Cómo se llama?	**How is it called?**	3	2	How it is called?	3	2	C�mo is called?	1	1	How it is called?	2	1
Wie heißen Sie?	**What's your name?**	5	5	How are you called?	4	5	How do you hei�en?	1	1	**What is your name?**	5	5

	Google	Rater 1	Rater 2	Yahoo - Systran	Rater 1	Rater 2	AppliedLanguage	Rater 1	Rater 2	x10	Rater 1	Rater 2
Could you help me?												
Podría ayudarme?	**It could help me?**	4	3	**It could help me?**	4	3	Podr�a to help me?	1	1	Would be able to help me?	3	1
Könnten Sie mir helfen?	**Could you help me?**	5	5	**Could you help me?**	5	5	K�nnten you me help?	1	1	**Could you help me?**	5	5
Thank you very much.												
Muchas gracias.	**Thank you very much.**	5	5	**Thank you very much.**	5	5	**Thank you very much.**	5	1	Many thanks	4	5
Vielen Dank .	**Thank you very much.**	5	5	Thank you.	5	5	Thank you.	4	5	**Thank you very much**	5	5
Do you speak English?												
Habla usted inglés?	**Do you speak English?**	5	5	Speech English you?	3	1	Speech you ingl�s?	1	1	Speak you English?	4	3
Sprechen Sie Englisch?	**Do you speak English?**	5	5	Do they speak English?	3	4	Do they speak English?	3	4	**Do you speak English?**	5	5
Let's go to the movies.												
Vamos al cine.	**We are going to the movies.**	4	4	We go to the cinema.	3	3	We go to the cinema	3	3	We go to the movies	3	4
Gehen wir ins Kino.	We go to the cinema.	3	3	We go into the cinema.	2	3	We go into the cinema	2	3	**Let's go into the movie**	5	4

											theater		
What's your name?													
Cómo se llama?	**How is it called?**	3	2	How it is called?	3	2	C�mo is called?	1	1	How it is called?	2	1	
Wie heißen Sie?	**What's your name?**	5	5	How are you called?	4	5	How do you hei�en?	1	1	**What is your name?**	5	5	

Table 2: Average Translation Accuracies, Per Phrase, Both Raters (Scale: 1-Bad; 5-Good).

1	Can you recommend a good restaurant?	Spanish	3.5	German	4.38
2	Can you call me a doctor?	Spanish	3.25	German	3.75
3	I don't feel well	Spanish	4.25	German	3.00
4	How much is it per day?	Spanish	2.63	German	4.00
5	Good morning	Spanish	4.00	German	4.50
6	Could you help me?	Spanish	2.50	German	4.00
7	Thank you very much.	Spanish	4.38	German	4.88
8	Do you speak English?	Spanish	2.88	German	4.25
9	Let's go to the movies.	Spanish	3.38	German	3.13
10	What's your name?	Spanish	1.88	German	3.88

Table 3: Average Translation Accuracies, Overall (Scale: 1-Bad; 5-Good).

Averages	Google	Rater 1	Rater 2	Yahoo Systran	Rater 1	Rater 2	Applied Language	Rater 1	Rater 2	x10	Rater 1	Rater 2
Spanish		4	3.9		3.9	3.5		2.6	2		3.3	2.9
German		4.5	4.5		3.9	4.3		2.5	2.9		4.7	4.5
Overall		4.25	4.2		3.9	3.9		2.55	2.5		4	3.7
Both raters: Spanish	3.95			3.7			2.3			3.1		
Both raters: German	4.5			4.1			2.7			4.6		
Both raters: Overall	4.23			3.9			2.5			3.9		

In this study, *Google Translate* probably performed better than the other systems because it is the only one that uses a statistical-learning rather than a rule-base approach. In comparison, the Kit & Wong (2008) study found *Google* and *SYSTRAN* had nearly identical BLEU and NIST scores for both German and Spanish. However, *Google Translate's* performance has evolved considerably since their test was conducted in January 2008. Further, our results are based upon human judgment that is superior to automated techniques (Snover, Dorr, Schwartz, Micciulla, & Markhoul, 2006). Finally, even though some studies (e.g., Coughlin, 2003) indicate BLEU is a reliable measure of translation accuracy and correlates well with human judgements, during the 2005 NIST MT evaluation, for example, BLEU failed to correspond to the scores produced in the human evaluations (Callison-Burch, Osborne, & Koehn, 2006). Thus, our study is the first to evaluate *Google Translate* and *SYSTRAN* based upon human judgment.

Further Testing with Google Translate

Although *Google Translate* was judged the best of the four systems, its translations were not always the best. Further, the analyses were based upon the subjective opinions of only two raters. In an attempt to test the system more thoroughly, we surveyed 32 students in an introductory MIS course to determine their understanding of the 10 phrases translated by *Google Translate* from Spanish or German into English. Each student was given a survey showing the translations from Spanish (Appendix A) or German (Appendix B) in which they were asked to record which of five possible sentences was the best match in meaning.

Students rated their overall English fluency as high (mean = 6.19, std dev = 1.17: 7-point scale), but Spanish and German fluency was low, as expected (mean = 1.87, std dev = 1.52). Survey answer frequencies are shown in Table 4, and the percentages of correct responses per question are shown in Table 5. Although accuracies ranged as low as 21.4%, some answers were nearly identical (e.g., sentence #9), and thus, the overall meaning of the sentence might still have been understood regardless of the incorrect answer being chosen. One notable exception was Spanish #4 in which most students thought the meaning was "How long is a day?" instead of "How much is it per day?"

Table 4: Survey Frequency of Responses
(Spanish N=14, German N=18: Options no students chose are omitted).

Answer	Spanish	German	% Spanish	% German
1c	10	17	71.4	94.4
1d	3	0	21.4	0
1e	1	1	7.1	5.6
2a	1	1	7.1	5.6
2b	9	17	64.3	94.4
2c	3	0	21.4	0
2e	1	0	7.1	0
3a	1	0	7.1	0
3b	0	2	0	11.1
3d	13	14	92.9	77.8
3e	0	2	0	11.1
4a	10	0	71.4	0
4b	4	16	28.6	88.9
4c	0	2	0	11.1
5a	2	2	14.3	11.1
5c	8	11	57.1	61.1
5d	2	1	14.3	5.6
5e	2	4	14.3	22.2
6a	8	1	57.1	5.6
6b	4	15	28.6	83.3
6c	1	1	7.1	5.6

6d	1	1	7.1	5.6
7a	10	13	71.4	72.2
7c	0	1	0	5.6
7d	2	2	14.3	11.1
7e	2	2	14.3	11.1
8a	2	0	14.3	0
8c	0	1	0	5.6
8d	1	2	7.1	11.1
8e	11	15	78.6	83.3
9a	3	8	21.4	44.4
9d	2	3	14.3	16.7
9e	9	7	64.3	38.9
10a	3	4	21.4	22.2
10c	5	12	35.7	66.7
10d	4	0	28.6	0
10e	2	2	14.3	11.1

Table 5: Percentages of correct responses
(Spanish N=14, German N=18).

Question No.	No. Correct	Correct %	Spanish	German	% Spanish correct	% German correct
1	27	84.4	10	17	71.4	94.4
2	26	81.3	9	17	64.3	94.4
3	27	84.4	13	14	92.9	77.8
4	20	62.5	4	16	28.6	88.9
5	19	59.4	8	11	57.1	61.1
6	19	59.4	4	15	28.6	83.3
7	23	71.9	10	13	71.43	72.2
8	26	81.3	11	15	78.6	83.3
9	11	34.4	3	8	21.4	44.4
10	17	53.1	5	12	35.7	66.7

Because many options had the same meaning, and the selection was rather arbitrary, we also asked students at the end of the survey to write what they thought translations from Spanish or German into English meant. Many of the translations of phrases from http://www.fodors.com/language/ were identical or nearly the same, so we chose five sentences that had poor translations. The actual correct answers were:

1. I work as a manager for a large European company.
2. The doctor recommended this dandruff shampoo for me.
3. You do not look well, you should see a doctor.
4. The variety of products in the nearest department store is very broad.
5. The road is closed due to construction.

Two objective evaluators judged on a scale of 1 (bad) to 7 (good) how close the students' equivalent sentences matched the actual meanings. For example, "There is a broad selection of local goods stores" and "there are many good stores" were rated lower because their meanings are different. In addition, "I manage a large European company" has a slightly different meaning. However, "Due to construction, the road is blocked" and "You seem sick, you should see a doctor" were judged to have the same meaning.

The evaluators determined ratings of 6.58, 6.44, 6.61, 4.82, and 6.25, respectively, for each of the five sentences. That is, except for sentence 4, the meanings were conveyed, even if the translation grammar was not good. Some students thought (Spanish: "The extension of products that are in the nearest department stores is very broad") and (German: "The imperative of local goods stores is very broad") meant a variety of stores instead of a variety of goods in the stores. Some didn't seem to have any idea what the German translation meant, writing "The imperative of local goods is very broad" and "The meaning of local stores is very broad."

There were no significant differences between male and female answers, or between the German and Spanish evaluations.

Complex Text Translations

A more detailed illustration of *Google Translate's* performance is shown with more complex text samples in Table 6 and 7. Although no rigorous evaluation of these translations was conducted, the resulting text indicates what kinds of errors can occur. For example, determining the gender of the subject was difficult in some cases. Not recognizing that "Daisy" is a female name, both languages had problems. The word "sie" can be "she" or "they" in German, and one German word ("unbändigem" - unrestrained) was not translated. However, the third German text translation was not entirely incorrect because "fresh cream buns" is not the same as "fresh raisin bread" in German.

Table 6: Sample translations from Spanish to English.

Source	Spanish Equivalent	Google Translation
Daisy picked up the telephone and tried to get through to her client again.	Daisy cogió el teléfono e intentó ponerse en contacto de nuevo con su cliente.	Daisy grabbed the phone and tried to make contact again with his client.
Her client, a certain Frank Baccini, who had a warehouse of electrical goods, had not paid for her two days' work.	Su cliente, un tal Frank Baccini, que tenía un almacén de electrodomésticos, no le había pagado por sus dos días de trabajo.	His client, a certain Frank Baccini, who had a warehouse appliances, he had not paid for their two days of work.
Daisy had got up early that spring morning because she was working on a case in the nearby town.	Daisy se había levantado temprano esa mañana de primavera porque estaba trabajando en un caso en una ciudad vecina.	Daisy was lifted early this morning spring because he was working on a case in a nearby town.

She arrived at her office with a paper bag in her hand containing fresh cream buns at a quarter to eight and was dying for a cup of coffee.	Llegó a la oficina a las ocho menos cuarto con una bolsa de papel en la mano que contenía bollos de nata y se moría por una taza de café.	He came to the office to the eight quarter with a paper bag in hand containing rolls of cream and was dying for a cup of coffee.
Today 10% of children in developing countries die before the age of five. This is down from 28% fifty years ago.	Hoy en día, un 10% de los niños de los países en desarrollo mueren antes de cumplir cinco años. Esto representa una reducción de un 28% hace cincuenta años.	Today, 10% of children in developing countries die before their fifth birthday. This represents a decrease of 28% fifty years ago.
Many hunger experts believe that ultimately the best way to reduce hunger is through education. Educated people are best able to break out of the cycle of poverty that causes hunger	Muchos expertos en el tema del hambre opinan que, a fin de cuentas, la educación constituye la mejor manera de reducir el hambre. La gente que tiene acceso a la educación cuenta con los mejores medios para salir del círculo de pobreza que causa el hambre.	Many experts in the field of hunger believe that, ultimately, education is the best way to reduce hunger. People who have access to education is the best way to break the cycle of poverty that causes hunger.

Sources:
- http://www.lonweb.org/daisy/ds-spanish-surprise.htm
- http://www.lonweb.org/daisy/ds-spanish-nightwatch.htm
- http://www.lonweb.org/hunger/hung-spa-eng.htm

Table 7: Sample translations from German to English.

Source	German Equivalent	Google Translation
Daisy picked up the telephone and tried to get through to her client again.	Daisy hob den Hörer ab und versuchte erneut, ihren Klienten zu erreichen.	Daisy lifted the handset and tried again to reach their clients.
Her client, a certain Frank Baccini, who had a warehouse of electrical goods, had not paid for her two days' work.	Ihr Klient, ein gewisser Frank Baccini, Besitzer eines Lagerhauses voller Elektrogeräte, hatte sie für zwei Tage Arbeit noch nicht bezahlt.	Your client, a certain Frank Baccini, owner of a warehouse full of electrical equipment, had them for two days work has not yet been paid.
Daisy had got up early that spring morning because she was working on a case in the nearby town.	Daisy war an diesem Frühlingstag früh aufgestanden, weil sie an einem Fall in der Nachbarstadt arbeitete.	Daisy was at this early spring day, because they at one case in the neighboring city worked.
She arrived at her office with a paper bag in her hand containing fresh cream buns at	Sie traf um viertel vor acht in ihrem Büro ein, mit unbändigem Kaffeedurst und	They met for seven forty-five in her office, with thirst unbändigem coffee and a

a quarter to eight and was dying for a cup of coffee.	einer Papiertüte frischer Rosinenbrötchen in der Hand.	paper bag of fresh raisin bread in his hand.
Today 10% of children in developing countries die before the age of five. This is down from 28% fifty years ago.	Heutzutage sterben 10% aller Kinder in den Entwicklungsländern vor dem fünften Lebensjahr. Vor fünfzig Jahren waren es noch 28%.	Nowadays, dying 10% of all children in developing countries before the fifth year of life. Fifty years ago there were still 28%.
Many hunger experts believe that ultimately the best way to reduce hunger is through education. Educated people are best able to break out of the cycle of poverty that causes hunger	Viele Hungerexperten sind der Ansicht, daß man Hunger letzten Endes nur durch Bildung reduzieren könne. Menschen mit Bildung sind am ehesten in der Lage, aus dem Armutskreislauf auszubrechen, durch den Hunger verursacht wird.	Many hunger experts are of the view that hunger is ultimately only through education can reduce. People with education are in the best position, from out of the cycle of poverty, hunger caused by the will.

Sources:

- http://www.lonweb.org/daisy/ds-german-surprise.htm
- http://www.lonweb.org/daisy/ds-german-nightwatch.htm
- http://www.lonweb.org/hunger/hung-ger-eng.htm

CONCLUSION

In this study, two objective raters evaluated common phrases translated from German and Spanish into English with four online translation services. *Google Translate* was the most accurate, and further study of this system with 32 students choosing among multiple choices and writing their own understanding of translations also showed good results. That is, even in cases where the grammar becomes garbled, the meaning can often be ascertained.

Based upon these results, we believe that *Google Translate* is the most accurate online service for translations of German and Spanish text to English. However, these services continue to evolve, and future research should compare the free, Web-based systems using more language pairs with a greater variety of text for a better understanding of which is most accurate.

REFERENCES

Ablanedo, J., Aiken, M., & Vanjani, M. (2007). Efficacy of English to Spanish automatic translation. *International Journal of Information and Operations Management Education*, 2(2), 194-210.

Aiken, M., Rebman, C., Vanjani, M., & Robbins, T. (2002). Meetings without borders: A multilingual Web-based group support system, *America's Conference on Information Systems*, August 9-11, Dallas, Texas.

Aiken, M., Vanjani, M., & Wong, Z. (2006). Measuring the accuracy of Spanish to English translations. *Issues in Information Systems*, 7(2)125-128.

Aiken, M., & Wong, Z. (2006). Spanish-to-English Translation Using the Web. *Proceedings of the Southwestern Decision Sciences Institute*, March 9 – March 13, 2006, Oklahoma City, Oklahoma.

Bar-Hillel, Y. (2003). The present status of automatic translation of languages. In Nirenburg, S., Somers, H., and Wilks, Y. (eds.) *Readings in Machine Translation*. MIT Press: Boston, 45-77.

Bezhanova, O., Byezhanova, M., & Landry, O. (2005). *Comparative analysis of the translation quality produced by three MT systems*. McGill University, Montreal, Canada.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. *Eleventh Conference of the European Chapter of the Association for Computational Linguistics: EACL 2006*, 249–256.

Coughlin, D. (2003). Correlating automated and human assessments of machine translation quality. *MT Summit IX, New Orleans, USA*, 23–27.

Geer, D. (2005). Statistical machine translation gains respect. *IEEE Computer*, 38(10), 18-21.

Hutchins, W. (2003). Commercial systems: The state of the art. In E. Somers (ed.) *Computers and Translation: A Translator's Guide*, Amsterdam: John Benjamins, 161-174.

Kit, C., & Wong, T. (2008). Comparative evaluation of online machine translation systems with legal texts. *Law Library Journal*, 100(2), 299-321.

LISA (2009). Machine Translation. LISA Best Practices Guide. Retrieved June 22, 2009 from http://www.lisa.org/Best-Practice-Guides.467.0.html.

Muegge, U. (2006). Fully automatic high quality machine translation of restricted text: A case study. *Proceedings of the 28th International Conference on Translating and the Computer*, November, London.

NIST (2008). National Institute of Standards and Technology: Machine translation evaluation official results. Retrieved June 22, 2009, from http://www.itl.nist.gov/iad/mig//tests/mt/2008/doc/mt08_official_results_v0.html

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. In *ACL-2002: 40th Annual meeting of the Association for Computational Linguistics* pp. 311–318.

Shankland, S. (2008). Google's translation center: Language lessons for the Googlebot? August 4, http://news.cnet.com/8301-1023_3-10005605-93.html

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA-2006)*, 223–231, Cambridge, MA, August.

Yang, J., & Lange, E. (1998). SYSTRAN on AltaVista: A user study on real-time machine translation on the Internet. *Proceedings of the 3rd Conference of the Association for Machine Translation in the Americas*, 28-31 October, Langhorne, PA, 275-285.

Yates, S. (2006). Scaling the tower of Babel Fish: An analysis of the machine translation of legal information. *Law Library Journal*, 98(3), 481-500.

Zhang, Y., Vogel, S., & Waibel, A. (2004). Interpreting BLEU/NIST scores: How much improvement do we need to have a better system? *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, 2051-2054.

APPENDIX A

SPANISH SURVEY

Sex: M   F

My English grammatical fluency is:
1 very poor   2   3   4 Neutral   5   6   7 very good

My Spanish grammatical fluency is:
1 very poor   2   3   4 Neutral   5   6   7 very good

Please circle what you think is the most correct meaning of each Spanish-to-English translation.

1. You can recommend a good restaurant?
    a. Can he recommend a good restaurant?
    b. Can they recommend a good restaurant?
    c. Can you recommend a good restaurant?
    d. Can you give me a good restaurant?
    e. How can you recommend a good restaurant?

2. You can call a doctor?
    a. Can I visit a doctor?
    b. Can you call me a doctor?
    c. Can he call me a doctor?
    d. Can they call me a doctor?
    e. Can we call a doctor?

3. I do not feel well.
    a. I do not feel it.
    b. I do not feel.
    c. I do not feel the well.
    d. I don't feel well.
    e. I do not feel a well.

4. How much is a day?
    a. How long is a day?
    b. How much is it per day?
    c. How much is it today?
    d. How cold is it today?
    e. How hot is it today?

5. Good morning.
    a. Morning is good.
    b. Good mornings.
    c. Good morning.

       d. Mornings are good.
       e. Good day.

6. It could help me?
       a. Could it help me?
       b. Could you help me?
       c. Could I help you?
       d. Could he help me?
       e. Could we help you?

7. Thank you very much.
       a. Thank you very much.
       b. Thank them.
       c. They are thankful.
       d. We are thankful.
       e. You are thankful.

8. Do you speak English?
       a. Do they speak English?
       b. Does she speak English?
       c. Do they understand English?
       d. Does he understand English?
       e. Do you speak English?

9. We are going to the movies.
       a. Let's go to the movies.
       b. They are going to the movies.
       c. They go to the cinema.
       d. We go to the cinema.
       e. We are going to the movies.

10. How is it called?
       a. What are you called?
       b. How is your name spelled?
       c. What's your name?
       d. How is it called?
       e. How are you called?

Please write down in grammatically correct English what you think the following Spanish-to-English translations mean:

6. Working as a manager for a large European company.
7. The doctor recommended me for this dandruff shampoo.
8. You do not have good appearance – you should go to see a doctor.
9. The extension of products that are in the nearest department store is very broad.
10. The road is closed due to workers.

APPENDIX 2

GERMAN SURVEY

Sex: M   F

My English grammatical fluency is:
1 very poor   2   3   4 Neutral   5   6   7 very good

My German grammatical fluency is:
1 very poor   2   3   4 Neutral   5   6   7 very good

Please circle what you think is the most correct meaning of each German-to-English translation.

11. Can you give me a good restaurant?
    a. Can he recommend a good restaurant?
    b. Can they recommend a good restaurant?
    c. Can you recommend a good restaurant?
    d. Can you give me a good restaurant?
    e. How can you recommend a good restaurant?

12. You can call me a doctor?
    a. Can I visit a doctor?
    b. Can you call me a doctor?
    c. Can he call me a doctor?
    d. Can they call me a doctor?
    e. Can we call a doctor?

13. I am not well.
    a. I do not feel it.
    b. I do not feel.
    c. I do not feel the well.
    d. I don't feel well.
    e. I do not feel a well.

14. How much does it cost per day?
    a. How long is a day?
    b. How much is it per day?
    c. How much is it today?
    d. How cold is it today?
    e. How hot is it today?

15. Good morning.
   a. Morning is good.
   b. Good mornings.
   c. Good morning.
   d. Mornings are good.
   e. Good day.

16. Could you help me?
   a. Could it help me?
   b. Could you help me?
   c. Could I help you?
   d. Could he help me?
   e. Could we help you?

17. Thank you very much.
   a. Thank you very much.
   b. Thank them.
   c. They are thankful.
   d. We are thankful.
   e. You are thankful.

18. Do you speak English?
   a. Do they speak English?
   b. Does she speak English?
   c. Do they understand English?
   d. Does he understand English?
   e. Do you speak English?

19. We are going to the movies.
   a. Let's go to the movies.
   b. They are going to the movies.
   c. They go to the cinema.
   d. We go to the cinema.
   e. We are going to the movies.

20. What's your name?
   a. What are you called?
   b. How is your name spelled?
   c. What's your name?
   d. How is it called?
   e. How are you called?

Please write down in grammatically correct English what you think the following German-to-English translations mean:

1. I work as a manager for a large European company.
2. The doctor gave me this dandruff shampoo is recommended.
3. You can see from sick, you should consult a doctor.
4. The imperative of local goods stores is very broad.
5. The road is blocked due to construction works.