

MEASURING THE ACCURACY OF SPANISH TO ENGLISH TRANSLATIONS

Milam Aiken, University of Mississippi, aiken@bus.olemiss.edu
Mahesh B. Vanjani, Texas Southern University, vanjanim@tsu.edu
Zachary Wong, Sonoma State University, zachary.wong@sonoma.edu

ABSTRACT

In the past few years, free and readily available Web-based machine translation (MT) services have appeared that can meet the needs of fast, informal translation needs. However, few if any studies have been conducted on these services to determine their accuracy. This study used perhaps the most successful, commercially available, online MT system to translate sample Spanish text of varying difficulty into English, which was then evaluated by 92 undergraduate business students for understandability. Results show that only two of the 12 text samples were not understood.

Keywords: Machine Translation, Understandability, Accuracy

INTRODUCTION

Multi- and bi-lingual translation services are in great demand due to the increasingly global nature of the business environment, but human translation services are often in short supply. While human translation usually is quite accurate, it can be fairly expensive and/or not readily available. For example, a busy executive might find a Web page written in Spanish that seems to contain important information, but if a Spanish-speaking employee is not on hand, the executive might not be able to read the page.

Several Web-based translation services allow users to translate samples of text, or entire Web pages, at any time, quickly and for free. However, it is not clear how accurate these translations are. Even if a translation has very few word errors, even one error can have a dramatic affect on its meaning. For example, a translation might be "The legal contract we discussed must be placed in the senior executive office by five o'clock Wednesday." If, however, the correct translation should have been "... junior executive ..." the difference is potentially quite important, even though there is only one error in 17 words (a 94% accuracy rate). On the other hand, even if a translation is ungrammatical and has a few incorrect words, it may be understood. In fact, much everyday conversation is not absolutely perfect

grammatically, and yet, the meaning is usually clear. For example, even a garbled sentence such as "How you are?" is often understood by most English speakers. The required accuracy of a translation is dependant in large part upon the consequences of errors. A legal document's translation should probably be nearly perfect, while an informal email message's translation could probably be allowed a few errors.

BACKGROUND

Several studies have focused on the accuracy of commercially available MT systems. For example, some studies of Spanish and English translation have used *SYSTRAN* software (<http://www.systransoft.com/index.html>), one of the best-known and most comprehensive translation systems in the world. *SYSTRAN* reports that it offers 36 language pair translations and that it also has the largest dictionaries of any MT system. In addition, several Web-based translation services rely on *SYSTRAN*.

There are no universally accepted and reliable measures of machine translation accuracy [1, 3], and only a few comprehensive tests have been conducted. For example, one study [5] compared three ARPA-supported systems (Pangloss, Candide, Lingstat) with the 13 commercial systems from *Globalink*, *PC-Translator*, *Microtac*, *Pivot*, *PAHO*, *Metal*, *Socatra XLT*, *SYSTRAN*, and *Winger*. Further, one study [2] concluded that the *SYSTRAN* provided poor translations, in general. In the study, short sentences were translated very well, but many longer sentence translations were very difficult to understand (see Table 1). However, efficiency is quite good with a comment of 20 words usually taking less than 0.5 seconds to translate.

Relatively low accuracy is the major barrier to increased use of MT systems. Accuracy often is expressed in terms of errors per hundred words or as a percentage, and 20 errors per 100 words are not uncommon with MT. However, word arrangement errors, insertions, and deletions can have an effect on the intelligibility of a translation, and these errors are

not reflected in simple word error rates. Therefore, focusing on the understandability of a translation rather than on its word error rate may be more appropriate.

Table 1. A Comparison of three MT systems [2]

System:	Logo Media	SYSTRAN	PROMT
Total number of sentences translated	27	27	27
Translated perfectly	2	3	5
Translated well	17	9	14
Translated imperfectly but, in general, comprehensibly	8	10	7
Not quite comprehensible sentences	0	5	1
Unrecognized words	16	10	7

Comparative evaluation: number of sentences

EXPERIMENT

In an attempt to measure the quality of translations from a common MT system (*SYSTRAN*), we used the Web-based version available at <http://babelfish.altavista.com/>. Spanish-to-English translations were chosen because of the shortage of subjects available to evaluate Spanish sentences.

Translation accuracy is highly dependent on the quality of the source. Therefore, we obtained random Spanish sentences from two introductory Spanish textbooks, and two Web sites (believed to have no spelling or grammatical errors). In addition, source text difficulty is an important factor in translation accuracy [4]. Consequently, along with the source text (shown in Appendix 1), we added measurements of reading ease (0=very difficult, 100=very easy) and reading grade level (1=very easy, 12=very difficult), both calculated automatically by *Microsoft Word* based upon sentence length, word length, and other factors. Source text difficulty ranged from a low of reading ease = 75.8 and a grade level = 3.6 for text sample Number 4, to a high of reading ease = 0 and a grade level = 12 for text sample Number 11.

A sample of 92 undergraduate students (38 female) were asked to evaluate the understandability of the translations shown in Appendix 1, using a scale of 1

= "I have no idea what this means" to 7 = "I am sure what this means." In addition, they recorded their Spanish- and English-speaking abilities using a scale of 1 = "none" to 7 = "excellent."

RESULTS

Subjects reported not being able to speak Spanish well (M = 2.7, SD=1.5), but they were able to speak English fairly fluently (M = 6.4, SD = 0.9). As shown in Table 2, translation Number 4 was understood the best, while Number 3 was understood the least. A difference-of-means T-test showed that only Translations 3 and 9 were significantly (at $\alpha = 0.05$) below the median on the understandability scale (not understood), although Translations 8 and 11 were not significantly different from the median. Only Translations 7 and 10 had no grammatical or word choice errors, although their phrasing is still awkward.

Table 2. Understandability Results

Translation	Mean	SD
1	6.065	1.316
2	5.402	1.644
3	2.978	1.791
4	6.424	1.019
5	5.946	1.304
6	5.087	1.560
7	6.370	1.002
8	3.750*	1.720
9	3.620	1.759
10	6.261	1.137
11	4.141*	1.726
12	5.077	1.522

* = not significantly different from median = 4 at $\alpha = 0.05$

There were no significant differences between males and females, except for Translation 11, where females reported not being able to understand it (M = 3.6) while males reported they did (M = 4.5). But, there were significant differences in understandability between those who reported knowing Spanish and those who didn't for Translations 3, 9, 10, and 11. Those who know Spanish and English can recognize how words can be translated literally, thus giving them an extra clue as to the real meaning of a sentence. In addition, Number 10 included the untranslated word 'resfriado,' which only Spanish speakers are likely to know (the correct English translation is 'cold' as in the illness). There were also significant differences between those who reported knowing English well and those who didn't for Translations 7 and 10. In these two translations, the

sentence structures are somewhat convoluted, making the understandability a little uncertain. Thus, subjects who know English better might be able to make better sense out of a complicated sentence structure. However, there was no significant correlation between understandability and source reading ease ($R^2=.24$, $p=0.42$)

CONCLUSION

A group of 92 undergraduate students evaluated Web-based SYSTRAN translations of 12 Spanish text samples that varied in reading difficulty. The students reported being able to not understand only two of the 12 translations (83% accuracy), but the understandability of translations did not seem to depend upon the complexity of the source text.

A more comprehensive analysis is needed with many more text samples, but this study is perhaps the first to evaluate the understandability of perhaps the most popular, free, and readily available machine translation system. In addition, more study is needed to judge the adequacy of translations. For example, what accuracy level is sufficient for a variety of documents? In this study, is 83% good or bad? To our knowledge, few if any studies have addressed the relative importance of accurate translations.

REFERENCES

1. Balkan, L., Netter, K., Arnold, D., & Meijer, S. (1994). Test suites for natural language processing. *Translating and the Computer*, London: Aslib, 51-58.
2. Bezhanova, O., Byezhanova, M., & Landry, O. (2005). *Comparative analysis of the translation quality produced by three MT systems*. McGill University, Montreal, Canada
3. Falkedal, K. (1991). *Proceedings of the Evaluators' Forum*, April 21-24, Les Rasses, Vaud, Switzerland. Geneva: ISSCO.
4. Hale, S. & Campbell, S. (2002). The interaction between text difficulty and translation accuracy. *Babel*, September, 48(1), 14-33.
5. White, J. & O'Connell, T. (1994). The ARPA MT evaluation methodologies: Evolution, lessons, and future approaches. In: *Technology Partnerships for Crossing the Language Barrier. Proceedings of the First Conference of the Association for Machine Translation in the Americas*, (5-8) October, Columbia, Maryland.

APPENDIX 1

Spanish Source Text with Translations

(Source: *Getting Along in Spanish*, M. Pei and E. Vaquero, Bantam, New York, 1957.)

1. ¿Dónde puedo tomar el autobús para Madrid?
[Translation included in the book: 'Where can I get a bus to Madrid?']
words: 7, reading ease: 42.6, grade level: 9.
SYSTRAN translation: **Where I can take the bus for Madrid?**
 2. ¿Quiere usted indicarme dónde está Calle Barcelona?
[Translation included in the book: 'Will you direct me to Barcelona Street?']
words: 7, reading ease: 30.5, grade level: 10.7
SYSTRAN translation: **Love you to indicate to me where is Barcelona Street?**
 3. Deseo apearme en Calle Barcelona.
[Translation included in the book: Please let me off at Barcelona Street.]
words: 5, reading ease: 15.6, grade level: 12
SYSTRAN translation: **Desire to lower to me in Barcelona Street.**
 4. Tengo hambre y sed.
[Translation included in the book: I'm hungry and thirsty.]
words: 4, reading ease: 75.8, grade level: 3.6
SYSTRAN translation: **I am hungry and thirst.**
 5. ¿Cuánto vale eso?
[Translation included in the book: How much is that?]
words: 3, reading ease: 49.4, grade level: 7.6
SYSTRAN translation: **How much it is worth that?**
- (Source: *Mastering Spanish*. L. Turk & A. Espinosa, Heath & Company, Lexington, MA, 1973.)
6. Buenos días, señorita Flores. ¿Puede decirme si el profesor Valdés está en su oficina?
(Chapter 1)
words: 14, reading ease: 42.6, grade level: 9
SYSTRAN translation: **Good morning, Flores young lady. She can say to me if professor Valdés is in his office?**

7. Las comidas constituyen una parte especial de la cultura de los pueblos. En los siguientes párrafos trataremos de dar una idea de algunas comidas típicas de los países al sur del Río Grande. **(Chapter 4)**
words: 33, reading ease: 36.2, grade level: 12
SYSTRAN translation: **The meals constitute a special part of the culture of the towns. In the following paragraphs we will try to give an idea of some typical meals from the countries to the south of the Grande River.**
8. ¿Tiene la universidad una función política? Se trata de una de las cuestiones más debatidas de nuestro tiempo. **(Chapter 6)**
words: 18, reading ease: 27.9, grade level: 11.7
SYSTRAN translation: **Has political university one function? One is one of the debated questions more of our time.**
9. Espero que no sea más que un resfriado. Si es algo grave, tendremos que llamar a la madre de Carlos, la cual vive en San Agustín. **(Chapter 8)**
words: 26, reading ease: 61.5, grade level: 8
SYSTRAN translation: **I hope that it is not more than resfriado. If he is something burdens, we will have to call to the mother of Carlos, who lives in San Agustín.**
10. Con las reformas de los últimos años ha habido un aumento notable en las responsabilidades y obligaciones de los gobiernos hispanoamericanos. **(Chapter 12)**
- words: 22, reading ease: 0, grade level: 12
SYSTRAN translation: **With the reforms of the last years there have been a remarkable increase in the responsibilities and obligations of the Hispano-American governments.**
11. En Aeroméxico valoramos su tiempo, y trabajamos día con día para ofrecerle excelentes promociones y nuevas opciones de vuelos que cubran sus necesidades de viaje, además, claro, de la puntualidad y servicio que nos caracterizan. **(Source: Aeromexico airline Web site: <http://www.aeromexico.com/mex/spanish/index.html>)**
words: 35, reading ease: 0, grade level: 12
SYSTRAN translation: **In Aeroméxico we valued its time, and we worked day with day to offer excellent promotions and new options to him of flights that cover their necessities with trip, in addition, sure of the puntualidad and service that characterize to us.**
12. ¿Conoce los servicios que ofrecemos en el aeropuerto? Consiga su tarjeta de embarque sin pérdidas de tiempo en las máquinas de autocheck-in. Con el ciberticket todo es más sencillo. **(Source: Iberia airline Web site: <http://www.iberia.com/>)**
words: 29, reading ease: 17, grade level: 12
SYSTRAN translation: **It knows the services that we offer in the airport? Obtain its boarding pass without losses of time in the machines of autocheck-in. With ciberticket everything is simpler.**