An Evaluation of Machine Translation Software

Emily Wehby

June 19, 2000

1. Introduction

Hutchins and Somers (1992) point out that, in an increasingly global economy, the relatively limited number of professional translators cannot meet the growing demand for rapid translations. Developers of machine translation systems aim to fill this need by offering computer-generated translations. They are quick to indicate, however, that large gaps exist between human translation and machine translation. While humans remain the only truly reliable translators, machine translation provides the advantage of immediate turnaround, and although the resulting translation may be less than accurate, it is generally enough to give users the "gist" of a foreign language document or Web page.

Hutchins and Somers (1992) indicate that there is much more to translation than knowing two languages. Even those fluent in two languages may have difficulty with the translation process. Beyond the ability to communicate in both languages, effective translation requires consciousness of the syntactic rules of each language and how these correspond. In a handbook for students of translation, Pérez Róman and Ferrié (1985) provide some brief theory of translation. The successful translator must have a broad cultural background and sensitivity to language nuance. Authors often cite the amount of word knowledge involved in the translation process as a major obstacle to accurate machine translation (Budiansky 1998). Pérez Róman and Ferrié describe three types of translation: literal translation, literary translation, and free translation. Literal translation involves providing word for word equivalents of the original text. This is not necessarily limited to word for word substitution but includes the application of word order and syntactic rules to render the translation comprehensible in the target language. Literary translation is more complex in that although the translation follows the meaning of the original closely, the process includes steps to ensure that the translation reads idiomatically in the target language. Finally, in free translation the translator attempts to relate the important ideas of the original without following the language and style of the source text. Considered the most difficult type of translation, free translation is more a process of rewriting the original in the target language than translation. Currently, researchers and users expect machine translation (MT) systems to be capable of only the first type of translation, literal translation.

Given the difficulty of incorporating common sense and stylistic sensitivity into computer systems, what can users of machine translation expect? As Dorr et al. (1999) indicate, many existing evaluations of MT software represent the perspective of the researcher in the process of developing a particular MT system, and no general evaluation methodology exists. Articles on MT written for popular audiences tend to focus on back and forth translations or translations of idioms to test MT software (Budiansky 1998). The back and forth method, which consists of feeding the system text in language A, requesting a translation into language B, and then translating the results back into language A in order to compare it with the original text, may underestimate the capabilities of an MT system. Translation rarely involves one-to-one correspondence, and it is unlikely that even professional translators will produce the exact source text when translating to and from languages several times. While mistranslations of idioms have provided entertainment for many users, such tests allow little insight into the software's ability to handle syntactic structures. They serve rather as a test of whether the system incorporates dictionaries of idioms.

In order to provide a more detailed evaluation of selected MT software, this study focused on four samples selected for representativeness of different knowledge domains. The procedure consisted of requesting translations of each sample from each of three internet sites. Error analysis of the resulting translations allowed inferences regarding the underlying principles driving the translation. In order to narrow the focus, the study was limited to Spanish to English and English to Spanish translation and to three internet sites that provide such translations online and free of charge.
2. MT Architectures

The literature on machine translation categorizes systems according to the architecture used. Systems use direct or indirect architectures, with indirect architectures further classified as either transfer or interlingua (Hutchins and Somers 1992; Whitelock and Kilby 1995; Dorr et al. 1999).

2.1 Direct

Direct systems, the first to be developed in the history of MT and often referred to as first generation systems, (Whitelock and Kilby 1995; Hutchins and Somers 1992) are closest to word for word translation. Many of these systems preserve the word order of the source text, even if such word order is unacceptable in the target language (Dorr et al. 1999). Hutchins and Somers (1992) characterize the direct approach as one in which intermediate stages are absent in the translation process; source language analysis leads to the generation of target language output. In general the system first performs morphological analysis on the source language input in order to eliminate inflections. This allows the system to use the root to look up words in the bilingual dictionary. The system then replaces words with their target language counterparts and may apply some local reordering rules to lead to more comprehensible word order in the target language output. Direct systems are the most limited of the three types, and translation output may be very difficult to understand for users unfamiliar with the source language. (Dorr et al. 1999).

2.2 Transfer

The limitations of direct architectures led to the development of indirect architectures. Transfer systems are the current focus of MT research. Unlike MT systems that employ a direct architecture, both transfer and interlingua approaches involve intermediate representations of a language. Dorr et al. (1999) point out that transfer architectures do not represent a discrete class but lie on a continuum from direct to interlingua approaches. Some transfer systems incorporate only syntactic transfer, while some involve semantic transfer as well. The distinguishing characteristic of the transfer approach, however, lies in the incorporation of language-dependent intermediate representations. Analysis of the source language leads to an abstract representation of the source text. A bilingual transfer module, specific to the two languages in question, then converts the source language representation into a target language representation. Finally, analysis of the target language representation leads to the generation of target language output. (Hutchins and Somers 1992).

2.3 Interlingua

Researchers explored the third type of MT architecture, interlingua architectures, in the 1960's, before transfer architectures (Whitelock and Kilby). The difficulty of implementing interlingua systems turned researchers' attention to transfer systems. There are currently few commercially available MT systems based on the interlingua approach (Dorr et al. 1999). The idea of the interlingua architecture rests on the presence of language-independent intermediate representations. The incorporation of intermediate representations makes the interlingua approach similar to the transfer approach, but all representations in transfer architectures are language-specific. Researchers at first hoped to develop a universal interlingua which could serve as an intermediate representation between any two languages, but the difficulty of this endeavor led to less ambitious goals for interlingua systems (Hutchins and Somers 1992). Interlingua architectures are especially difficult to implement since many such systems attempt to use a deep semantic analysis in order to arrive at appropriate meaning representations. Such an analysis requires programming the computer to employ extensive use of world knowledge, or common sense, a major challenge for researchers currently working in machine translation and other applications of artificial intelligence (Dorr et al. 1999).

3. Linguistic considerations and predicted errors

Most descriptions of machine translation technology include a categorization of linguistic considerations that may cause problems for MT systems. (Hutchins and Somers 1992; Whitelock and Kilby 1995; Dorr et
al. 1999). While the authors vary in the number of categories listed and level of detail of their descriptions, all include the primary categories of lexical ambiguity and syntactic or structural ambiguity.

### 3.1 Lexical Ambiguity

Lexical ambiguity involves the presence of one-to-many mappings of meaning between languages. As the translation tests in this study all involve English to Spanish or Spanish to English translations, anticipated errors in the area of lexical ambiguity may involve the mistranslation of an English word having two meanings into a Spanish word which carries the unintended meaning. For example, the English word *book* may be used as a noun, and correctly translated into Spanish as *libro*. Alternatively, the English word *book* may be used as a verb, meaning *to reserve*, such as in *to book a flight*. This meaning should be translated into Spanish as *reservar*. Lexical ambiguity also exists in the other direction, as there are Spanish words which have two meanings translated by separate words in English. For example, the Spanish *pluma* may be either English *pen* or *feather*. The correct choice for translation depends on context. Thus, in order to deal effectively with problems of lexical ambiguity, MT systems must incorporate the ability to determine the context of a word. Because of the existence of several such examples for English and Spanish, a majority of the errors found in the translation tests should involve lexical ambiguity.

### 3.2 Syntactic Ambiguity

Syntactic ambiguity, another major area that may cause problems for translation, consists of ambiguities present at the level of the sentence rather than at the level of the word. Hutchins and Somers (1992) cite the famous example of syntactic ambiguity in the English sentence, *Flying planes can be dangerous*, which can either mean *It can be dangerous to fly planes* or *Planes which are flying can be dangerous*. Dorr et al. (1999) point out that resolving such ambiguities in the translation process may be unnecessary if the syntactic ambiguity transfers to the target language. Hutchins and Somers (1992) refer to this option, usually available only if the source and target languages are closely related, as a *free ride*. Because Spanish and English are fairly closely related syntactically, relatively few of the errors found in the translation tests should result from cases of syntactic ambiguity.

### 4. Software

Translation tests run on three online MT systems served as a preliminary evaluation of the effectiveness of widely available free of charge machine translation services. These systems included Freetranslation.com, Altavista's Babelfish translator, and yupi.com. All three systems supported translation from Spanish to English and from English and Spanish as well as translation between other language pairs.

#### 4.1 Freetranslation.com translator

Transparent Language provides Freetranslation, located at http://www.freetranslation.com, as "an easy-to-use site for rapid translation" that gives "the 'gist' of foreign language text and web pages". The company provides this free service in order to showcase its TranscendRT Machine Translation technology. (http://www.freetranslation.com/faqs.htm). Transparent's website includes a brief, vague explanation of how the TranscendRT technology works, classifying it as a transfer method of machine translation. First the source document undergoes segmenting, division into paragraphs, sentences, and words. Individual words then undergo morphological analysis, and the system looks up words in its dictionary. Then, the system performs functional analysis in order to determine how each word functions within the source text sentence. A process of syntactic analysis determines the grammatical structure of each sentence, and then the translator reorders the elements of the sentence to form acceptable target language syntax. Generation of the target text occurs after an analysis, which, in effect, reverses the steps used to analyze the source text (http://www.transparentlanguage.com/ets/about/-MTWhitePaper.htm).

#### 4.2 Altavista's Babelfish translator

Altavista's Babelfish translator, located at http://babelfish.altavista.com, was a result of a joint effort between Altavista and Systran and appeared online in December of 1997
Systran is one of the oldest MT systems, with over twenty years of operational service (Hutchins and Somers 1992). According to Yang and Lange (1998), the Babelfish translator receives over 500,000 requests for translations per day. The Systran website describes its translation methodology as "a sentence by sentence approach, concentrating on individual words and their dictionary data, then on the parse of the sentence unit, followed by the translation of the parsed sentence". The Systran webpage describes the roles of the parser, target language translation modules, and the synthesis module. Although the MT literature describes Systran as an example of a first generation or direct system (Hutchins and Somers 1992; Whitelock and Kilby 1995), Systran's webpage description hints at some characteristics of the transfer approach, in that the parser attempts to build up representations of the source sentences and the synthesis module is source language independent.

4.3 Yupi.com translator

The third translator tested, located at http://www.yupi.com/Traductor, is part of a Spanish language page, yupi.com, which provides news articles, a search engine, and other services. E-lingo, a company which "develops and hosts real-time translation solutions for portals, e-tailers, email providers, and content sites" provides the translation service. E-lingo does not provide any information about the workings of its MT system but indicates that it uses MT technology provided by Transparent Language and by Lernout and Hauspie.

5. Test samples

The four samples selected in order to test the MT systems described in the preceding section each represent a different genre, or semantic domain. As Dorr et al. (1993:13) observe, machine translation generally works best for scientific and technical documents, as MT systems have difficulty with texts that incorporate more metaphor and world knowledge, such as literature. The samples included a selection from a VCR manual, a short research article abstract, a paragraph from a New Yorker article, and a paragraph from a short story. For each sample a human-produced translation was available for comparison to the machine translation results. The first sample came from a manual for a Panasonic VCR model PV-M1346/PV-M1356W, which included both English and Spanish instructions. While this sample included the most limited vocabulary, it was possibly the most ambiguous syntactically, as the English instructions consisted mostly of incomplete sentences. The second sample consisted of a research abstract for an article by Margaret G. McKeown, "Creating effective definitions for young word learners" in the journal Reading Research Quarterly 28(1). The article, originally written in English, included a brief abstract in both English and Spanish as well as a few other languages. The third sample came from an article by Jon Lee Anderson entitled "The Power of García Márquez" published in the September 27, 1999 issue of The New Yorker. The Spanish translation of this article appeared in the October 2, 1999 edition of the Colombian magazine Semana. The final sample, chosen to represent literary language and predicted the most difficult to translate, consisted of the opening paragraph of Colombian author Gabriel García Márquez's short story La siesta del martes (Tuesday Siesta), translated into English by J.S. Bernstein. See Appendix A for complete text of each of these samples in the original and in human-produced translation.

6 Error Analysis and Evaluation

Analysis of each translation produced by the MT systems allowed the identification and classification of errors. Most errors resulted from failure to translate a particular word, ambiguity in the original text, lexical ambiguity due to one-to-many mappings of meaning onto words across the two languages, or a misanalysis of the structure of sentences in the source text. See Appendices B through D for the translation results for all test samples.

6.1 Failure to translate

A failure to translate a particular word shows simply that the word is not incorporated in the system's dictionaries. Almost all such errors occurred, not surprisingly, in the translation of the final sample, the
short story excerpt. All three translators failed to translate the English word *oxcarts*. The human translator rendered this as *carretas de bueyes*. All three translators also failed to translate the Spanish *trepidante* into the English *quivering* or an English word with a similar meaning. Atlavista's Babelfish translator failed to translate English *roses* into Spanish *rosales* and Spanish *bermejas* into English *sandy*. Freetranslation showed a failure to translate in other samples as well. For the VCR manual, Freetranslation was unable to translate Spanish *grabación* into English *recording*, and for the research abstract Freetranslation did not translate Spanish *implicancias* into English *implications*. Of the three systems, Freetranslation had by far the greatest number of errors in this category and for the Spanish to English translation of the short story sample left thirteen words untranslated. This indicates that Freetranslation's dictionaries are not as complete as those incorporated in the Babelfish and Yupi systems.

### 6.2 Ambiguity in the original

A few errors resulted from ambiguous or unclear language in the source text. This differs from lexical or syntactic ambiguities in that meaning is unclear due to ineffective writing rather than ambiguities inherent in the nature of the language. The main example of this type of error came from Freetranslation's output for English to Spanish translation of the VCR manual. The manual instructions include a fragment, "Program Timer indicator lights on the front panel". Without a disambiguating *the*, this phrase may be interpreted as either an imperative (*Program the Timer indicator lights*) or as a description of an event (*The Program Timer indicator lights*). The context of the manual makes it clear that the author intends the latter meaning. In this case, the correct translation into Spanish, that incorporated in the manual, is *La luz indicador de la grabación programada se enciende en el panel delantero*. The output from Freetranslation, however, reads as the imperative, *Programe las luces indicador del reloj...*, as a result of applying the first analysis.

### 6.3 Errors due to Lexical Ambiguity

Errors due to lexical ambiguity are the easiest to identify and accounted for a large proportion of the errors found in the translations. In translating the VCR manual, both Yupi and Babelfish mistranslated the phrase *power comes on* in *The Combination VCR power comes on automatically*. Instead of translating this idiomatically as *se enciende*, Yupi gave a more literal translation, *el poder de VCR viene*, meaning the power, in the sense of capability, of the VCR comes, in the sense of moving toward. Babelfish translated this phrase with a similar structure. Also in the VCR manual, both Yupi and Freetranslation translated *key*, referring to a key on a remote control, as *llave*, a key used to open a door, rather than *botón*. For *channel*, Freetranslation gave *cauce*, referring to a body of water, rather than *canal*, referring to a television channel. Freetranslation translated the English word *play*, used in the VCR manual to refer to the length of a recording, as *obra*, a play produced in the theater. Babelfish translated *play as juego*, play as a diversion, as in playing a game. In the VCR manual Spanish to English translation all three translators gave *bellboys* as the English translation of *botones*. Although *botones* can translate as *bellboys*, the correct translation in this case was *buttons*, referring to the buttons on a remote control.

For the second sample, the research article abstract, Yupi translated *arose* in the phrase *principles that arose from the analysis as se levantaron* rather than the appropriate *surgieron*. *Se levantaron* is a form of the verb *levantarse*, used to refer to a person arising in the morning. The Spanish version of the research abstract used the word *oraciones*, which may translate into English as *sentences* or as *prayers*. The context of the article makes it clear that the meaning intended is that of *sentences*. While Yupi translated this correctly, Freetranslation gave *prayers*, and Babelfish rendered it *oraciones*. A final example of an error arising from lexical ambiguity in the translation of the research abstract, Freetranslation translated *efectivas* in *las definiciones revisadas fueron más efectivas*, (English *the revised definitions were more effective*), as *cash* rather than *effective*.

The third and fourth samples, the selections from the *New Yorker* article and the García Márquez short story, did not lead to quite as many errors due to lexical ambiguity. This is probably because these samples were shorter than the selections from the VCR manual and the research abstract. A few interesting errors did arise, however. For the *New Yorker* excerpt, Yupi translated the English *charge him* in the

---

1 I would like to thank Carlos Restrepo for help in interpreting the Spanish language examples listed in this and the following section.
phrase *arrest him and charge him with having links to the M-19 guerillas, as cobrarlo.* The Spanish verb *cobrar* does not include the English meaning indicated by this context; it translates as English *charge* in the sense of charge someone money for goods or services. The human translator for this selection rendered the intended meaning with the phrase *arrestarlo bajo la acusación,* which translates literally into English as *arrest him under the accusation.* For the same selection, Freetranslation translated *flight in flight into exile* as *vuelo,* an airplane's flight, rather as *huida,* a flight as in escape, such as that of a fugitive. Finally, for the last sample, the short story excerpt, Yupi translated *fans in offices with electric fans,* as *entusiastas,* meaning people who are fans of a sport, etc., rather than as *ventiladores,* fans used to cool.

### 6.4 Errors due to Structural Analysis

Another class of errors consists of those that arose from a misanalysis of the syntax or structure of sentences in the source text. For example, Yupi translated the English phrase *24 fifth graders* into Spanish *24 quintos alumnos.* This translation is interesting in that it includes the term *alumnos,* English *students,* which the original only implies, but does not come up with the appropriate translation *estudiantes de quinto grado.* In some of the translations, the MT systems interpreted adjectives as applying to phrases other than those intended in the original. For the final sample, the excerpt from the García Márquez story, Babelfish translated the English phrase *electric fans, red-brick buildings,* and *residences with chairs and little white tables* into Spanish *los ventiladores, los edificios del rojo-ladrillo,* y *las residencias eléctricos.* This interprets *electric* as modifying fans, buildings, and residences rather than just fans. The Yupi translation of the same sample gave the English *palms and powdery rosebushes* for the original Spanish *palmeras y rosales polvorientos.* The human translation gives *dusty palm trees and rosebushes,* translating *polvorientos* as describing only rosebushes. This aspect of the translation results from the syntactic ambiguity in the original text rather than a misanalysis of the structure by the translation system.

The misinterpretation of verb morphology also led to errors in translation. As Whitley (1986) points out, in English many verbs have past participle forms that are homophonous with the past tense form. For example, the form *murdered* may be used as a participle, as in *the murdered man,* or as the past tense form, as in *He murdered a man.* (Whitley 1986:96). Spanish verb morphology does not follow this pattern; there are two distinct forms for the past participle and the past tense. In the English to Spanish translation of the VCR manual from Freetranslation, the phrase *at a preset time becomes en un fijó tiempo.* *Fijó* is the third person past form of the verb rather than the participle form, *fijada.* The human-produced translation reads *a una hora fijada.*

All of the English to Spanish translations from Freetranslation showed a failure to recognize the infinitive of the English verb. Rather than analyzing the 'to + verb' form as infinitive, Freetranslation always interpreted *to* as a preposition. The translation system always analyzed the verb form as the imperative rather than the infinitive. For the VCR manual sample, Freetranslation translated *to choose as* *a escoge* rather than as the infinitive *escoger.* For the research abstract, Freetranslation translated *to understand as* *a entiende* rather than *entender* and rendered *to use as* *al uso* rather than *usar.* For the short story sample Freetranslation translated the English infinitive to *cross* as the Spanish prepositional phrase *a la cruz.* Finally, in the translation of the sample from the *New Yorker* article, the phrase *a plan to arrest him* became *un plan al arresto él* instead of the correct *un plan para arrestarlo.* In addition to misinterpreting the infinitive as a prepositional phrase, Freetranslation translated *him* as the nominative *él* rather than as the direct object *lo.*

A final example of error due to misanalysis of structure appears in many of the translations of sentences involving noun-to-noun relationships. Whitley (1986) discusses the differences in the expression of noun-to-noun relationships in English and Spanish in order to highlight problems that may arise for the native speaker of English learning Spanish as a second language. According to Whitley, "a major source of interference in learning Spanish is the fact that nonpossessive N1 of N2 relationships in English may be changed to N2 N1." For example, the phrase *a metal furniture screw manufacturer* corresponds to *a manufacturer of screws of furniture of metal.* (Whitley 1986:153). Spanish expresses such noun-to-noun relationships using the pattern N1 de N2. For example, the English *sea breeze* translates into as Spanish *brisa del mar* (*breeze of the sea).* Overgeneralization of this rule led to translation errors. For the research abstract Freetranslation rendered the English to *address problems* as *problemas de dirección.* This translation compounds two different errors, interpreting the *to* of the English infinitive as a preposition (Spanish *a*), and interpreting *address* as a noun. The translation reordered the two nouns *address* and *problems* into *problems of address* to fit the Spanish pattern N1 de N2. This contrasts with the correct
translation, para abordar los problemas, which results from the analysis of to address as a verb and problems as the direct object. Also, in the translation of the research abstract, the phrase in helping students became en estudiantes de porción. In this case, the translation system analyzed helping as modifying students and translated helping as a noun, a helping of food. As a result, the translation does not give the meaning intended by in helping students but rather conveys the idea of in students who study a helping of food. Both Freetranslation and Babelfish made similar errors in translating the short story excerpt. In the phrase by his friend President Mitterand, both systems analyzed friend in a manner similar to helping in the previous example. Freetranslation gave the translation por su Presidente de amigo Mitterand, and Babelfish gave the translation por su presidente Mitterand del amigo. In the second instance, Babelfish analyzed President Mitterand together as one noun, N1 in the N1 de N2 pattern, while in the first example Freetranslation separated Mitterand from President, and President alone fills the role of N1. These errors, however, may have arisen because the author of the original English text did not include a comma between friend and President. When the original text included a comma, the MT systems produced accurate translations. Freetranslation gave por su amigo, el Presidente Mitterand, and Babelfish gave por su amigo, presidente Mitterand.

7. Conclusions: How Good is Free MT?

While the errors pointed out in the preceding section show that none of the translations were entirely accurate, some of the translations were fairly effective in conveying the basic ideas of the original text. Overall, the Yupi translations were the most accurate, and the translations from Freetranslation were of the poorest quality, with the Babelfish translations falling in the middle. While all of the translators showed some failures to translate and errors due to lexical ambiguity, Yupi had fewer structure-related errors than the other two translators. The translations provided by Freetranslation showed the most errors overall, including the greatest number of failures to translate. The inability of Freetranslation to recognize the infinitive in English led to some incomprehensible text in its English to Spanish translations. The errors due to the misanalysis of noun-to-noun relationships in translations from both Freetranslation and Babelfish may cause additional confusion for readers of the target text. Also, the many errors due to lexical ambiguity in translations from all three translators cause difficulty in understanding for readers unfamiliar with the source text language. The type of text translated does seem to have some effect on translation accuracy. The excerpt from the short story led to the greatest number of failures to translate. The translations of the short excerpt from the New Yorker article were relatively more understandable overall. The less literary language used in the VCR manual and in the research abstract did not lead to significantly greater accuracy for the translations of these texts, though they did include less instances of failures to translate. Overall, the MT output bore little resemblance to the human-produced translations, and exact matches of translated phrases occurred very rarely. In conclusion, while any of the three MT systems examined, Yupi, Babelfish, and Freetranslation, may serve as a useful translation aid for users who have a working knowledge of both source and target languages, the output may prove very puzzling for a reader unfamiliar with the source text.

References


e-lingo: Internet Translation Solutions http://www.e-lingo.com


SYSTRAN's MT Architecture http://www.systransoft.com/mt_arch.htm

Transparent Language Machine Translation and the Power and Purpose of "Gisting" in the Internet Era: A White Paper from Transparent Language. August 1999
http://www.transparentlanguage.com/ets/about/MTWhitePaper.htm

