Morris, Richard E. 2004. Honest compositions in a Babel-Fish generation.
Paper presented at the 37th Annual Conference of the Tennessee Foreign Language Teachers Association.
Nashville, November 6.

HONEST COMPOSITIONS IN A BABEL-FISH GENERATION*

Richard E. Morris
Middle Tennessee State University

1. Introduction

Today as educators we have no illusions about the easy availability of translation engines, and we should not be so naïve as to think that students won't use them simply because we ask them not to. Let's face it: they are readily accessible, they cost nothing, and they provide nearly instantaneous results. They have come a long way from anything you or I are likely to have played with in the 1980s or 1990s. Today they display surprising sensitivity to idiom and cultural nuance. But how accurate are they? Visit any of their internet homepages and you would think they were linguistically all-powerful. Most homepages bill their product as "free translation." You type or paste your text in one window, press a button, and within seconds the translation appears in the other window. If you look closely at the fine print, though, you will find that even the manufacturer warns against heavy reliance on the engine as a translation tool. In fact, most use the free translation tool as a lure to upsell homepage visitors to the much more expensive human translation services. It's one of the oldest tricks in the book.

The homepage of one translation portal contains a long list of "translation tips" to keep in mind as you use the free translation engine. Of course, this begs the question, why should I need translating tips at all, if the engine is so cleverly doing the translating for me? The tips include some sensible advice about correct punctuation, and also some rather nebulous suggestions about keeping sentences simple, using words that have only one meaning, and – my favorite – avoiding idiomatic expressions. In other words, to get maximum benefit from the product you must first relieve the translation engine of all of the core translation tasks.

There is a culture of consumerism evidenced in these products as well. Perhaps the most well-known homepage is called Babel Fish. Do we even remember what a Babel Fish is? The Babel Fish is a piece of cultural history in itself that had its birth in the mind of science fiction author Douglas Adams, and appeared in the popular novel *The Hitchhiker's Guide to the Galaxy*, which was later serialized for television in Great Britain. A page at BBC online defines the Babel Fish thus:

> The Babel Fish is small, yellow, and simultaneously translates from one spoken language to another. When inserted into the ear, its nutrition processes convert sound waves into brain waves, neatly crossing the language divide between any species you should happen to meet whilst travelling in space.

A more recent version of this pop culture icon is found in the Star Trek series, with the more scientific-sounding name of "universal translator." Like the Babel Fish, the universal translator deciphers and translates any language. I found a description of how it works on the webpage of a fan of the Star Trek spinoff series Deep Space 9 (DS-9):

* The author wishes to thank the audience at the 37th Annual Conference of the Tennessee Foreign Language Teachers Conference for valuable observations and insights, some of which have been incorporated into this final version of the paper.

> One DS-9 episode had a new (humanoid) species emerge from the wormhole, and at first we heard their language directly, untranslated, because the universal translator had not yet isolated any consistent patterns to translate. A short while later, we began hearing articles and conjunctions in English, a content word here or there, surrounded by bursts of unintelligible alien communication. Within hours, of course, everyone was communicating fluently, even colloquially....

The idea of a handy little device, always out of sight, that effortlessly converts meaning from one language into another epitomizes the popular view of the language gap. When in doubt, find a machine to do it for you. There *must* be a machine to do all this hard work, mustn't there? The L2 learner might view L2 writing as a time-consuming process to be sidestepped with the help of a translation engine. Indeed, a sizable percentage of students in any one of my L2 classes sees the translation engine as a sort of calculator, except instead of doing math it does language. Since using a calculator to perform advanced mathematical operations isn't considered unethical, why should it be unethical to use a translation engine?

The question begs at least one other question. Is mathematical skill comparable to L2 skill? It seems that the answer is no. Mathematical skill, especially in its most advanced applications, focuses more on distilling reality into mathematical variables – in a word problem, for example. That is why the use of a calculator isn't frowned upon (as much): because the most challenging skill is not the math itself, but the ability to see and define the world in mathematical terms. Once that heavy task is done, the computations are incidental.

However, we cannot say that writing in a language is primarily about viewing the world in *linguistic* terms and that the actual act of composition is incidental. The reasons we cannot, and should not, are many.

First, as speakers of the L2 that we teach, we know from our own experience that we progressed in our own L2 writing skill only by becoming *more* self-sufficient, not *less* so. Unbridled use of a translation engine as a writing tool denies the student an opportunity to gain proficiency and self-sufficiency through sustained practice.

Second, and more significantly, this line of thinking runs counter to nearly twenty years of progress in the area of L2 pedagogy. It denies much evidence that the act of writing is itself an integral component of the learning process. Enlisting the speedy services of a translation engine removes the act of writing from the L2 writing task.

The proficiency method, in common practice since the 1990s, calls for language as communication at *every* level, and in all modalities, *including* writing. In a 2000 article titled "Writing and foreign language pedagogy: Theories and implications," Homstad and Thorson summarize the proficiency learning experience as follows:

> Students are expected to be active participants in their own learning, to be risk-takers, and to use language to create meaningful communication (p. 9).

When is writing communicative in nature? When it is undertaken as a *process*, a *means* to an end, rather than as an end in itself. Much of the literature on writing across the disciplines finds much greater learning value in the *process* than in the product. The translation engine intrudes upon this process absolutely by taking over the L2 writing process. The student is therefore relieved from the linguistic self-sufficiency that the proficiency method seeks to instill.

In this talk I will explore three issues. First, we will take a layperson's look at how a translation engine works, just to get a sense of their abilities and limitations. Second, I will show how four different translation engines handle sentences introduced during the

first three semesters of an introductory college Spanish class. The purpose of this comparison is to get a feel for the strengths and shortcomings of translation engines in a very practical sense, in order to make a meaningful statement about their effectiveness. I will not go so far as to rate the four translation engines on their ability, since it is not my intention to suggest that any one is better – or worse – than any other.

Third and finally, I will open up the floor to discussion, as I would like to hear your views on translation engines as well as any ways you have found to incorporate them into your teaching.

2. Translation engines: A brief modus operandi

Machine translation dates back to World War II, when specialized machines were needed to encode and decode secret messages quickly. After the war, proposals were made to explore the possibilities of the decoding machine as a translating device for natural language. The earliest projects focused on translating English to Russian and vice-versa. In the 1960s, Machine Translation was deemed non-cost-effective and funding went into more promising projects, such as those in the new filed of artificial intelligence (AI). In 1970 a company named Systran began doing Russian-English translations for the U.S. Air Force. In 1976, an English-French version of Systran was adopted for use within the European Community. The 1980s saw several new projects in France, Germany, Japan, and elsewhere. The 1980s also saw the first commercial machine translation systems. In the 1990s large companies, including Japanese electronics manufacturers, began to market machine translation software for PC use. In the 1990s, work also began on speech recognition technologies.

To date there have been three main architectures used in the programs that run translation engines. These are **direct**, **transfer**, and **interlingua**.

The **direct** architecture was common in older systems. It used a vast string memory to match source phrases to target phrases, and, as such, was essentially a glorified dictionary.

The **transfer** architecture has a deeper "knowledge" of language in that it is able to take apart phrases and sentences it has never encountered before and represent them accurately in the target language. The transfer architecture draws heavily upon Noam Chomsky's work in the late 1960s on generative syntax. One of the most compelling principles of Chomsky's work is that human beings possess, as part of their innate language faculty, the ability to generate an infinite number of sentences using a finite number of syntactic structures and rules of transformation. The transfer architecture is built around these structures and rules, and translation involves translating words as well as structures.

A newer and more ambitious development is the **interlingua** architecture. In this type of system, sentences are parsed not just syntactically but also conceptually. The conceptual sentence structure is supposedly the same in all languages. In this sense, the interlingua architecture more closely models the generation of language from thought. The interlingua architecture breaks sentences down syntactically *and then semantically*, assigning them a deep conceptual structure. This deep conceptual structure may then be reconstituted as meaningful language using semantic as well as syntactic construction rules. Thus the interlingua system does not contain rules for converting words and structures between languages, but rather rules for breaking down language into its conceptual atoms, and then rebuilding it using the rules of a different language.

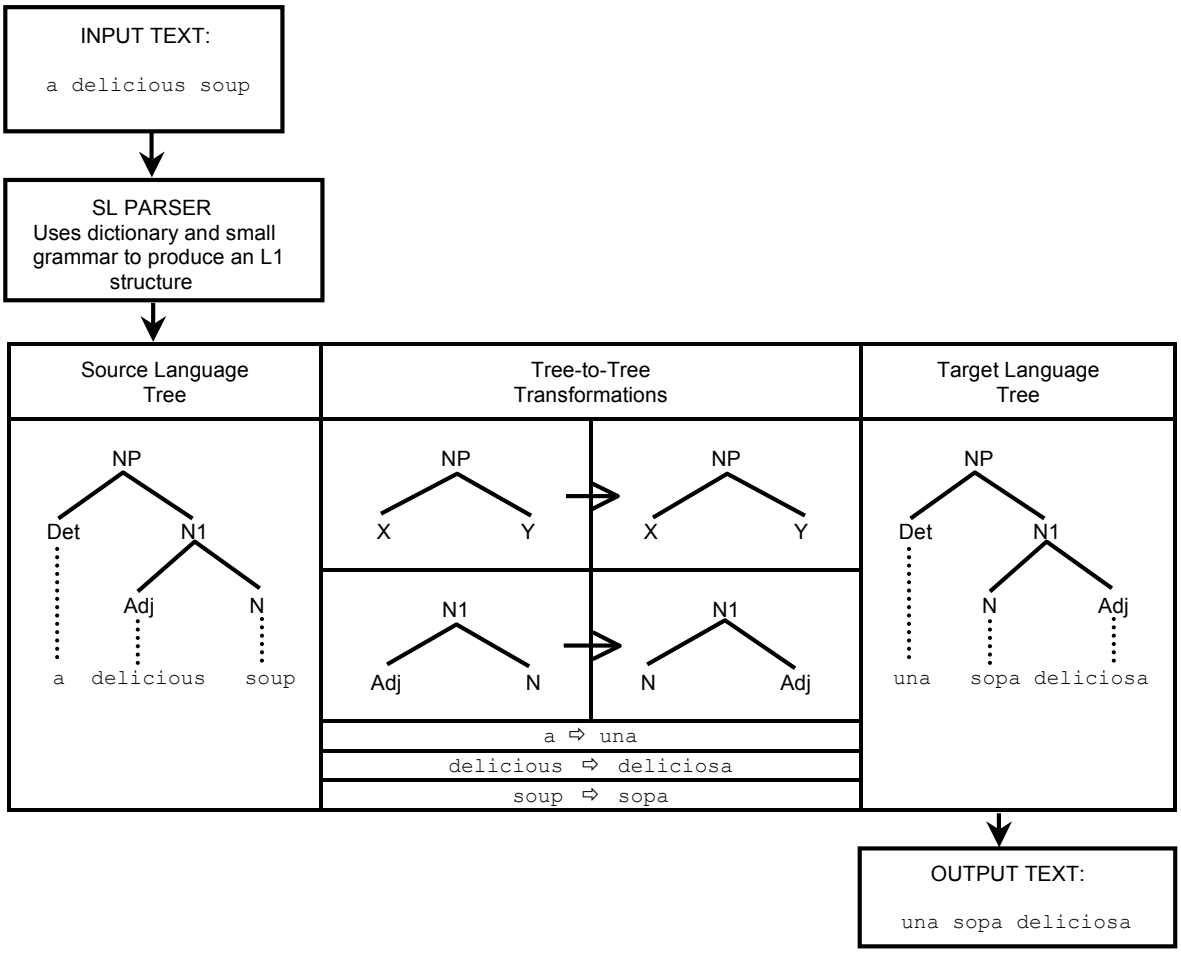
Today systems typically draw upon the most practical elements of both the transfer and interlingua architectures.

So what can a translation engine do? A well-made translation engine is extremely good at translating isolated words or minimal phrases, such as *John eats* or *the black cat*. Depending on grammatical complexity, it is anywhere from very good to astonishingly inept at translating larger phrases, simple sentences, complex sentences, and sentence clusters. Like a bilingual dictionary, a translation engine has an extensive vocabulary. In addition, it also has information that bilingual dictionaries do not have, such as finite verb forms. For example, when you look up *eat* in an English- Spanish bilingual dictionary, you are given the infinitive *comer* and you are expected to determine the finite forms on your own. A good dictionary usually has a verb list elsewhere to guide you in choosing the correct finite form, for example *comemos* for *we eat*. A translation engine does this step for you, and it does so not by accessing a grand verb table, but rather by building the form from its morphological components using word formation and grammar rules, in much the same way as the human brain.

How is this possible? Unlike a dictionary, which offers up individual words, the translation engine typically takes as its input a string of words or a whole sentence. This means it has access not only to a massive vocabulary but also to detailed grammar rules about how words are permitted, or likely, to relate to each other contextually.

Let's take a closer look at how the transfer engine works. First, the engine analyzes the word, phrase, or sentence that you have typed in. It begins by looking up each word in its dictionary and assigning it to a syntactic category, such as noun, preposition, etc. This information, along with the categories of words it can be in construction with, is found in the word's "listing."

An English → Spanish transfer architecture
(cf. Trujillo 1999: 123; Arnold et al. 1994: 60)



Second, it uses English syntax rules to try to parse the sentence. In other words, it builds a syntactic constituent structure for the sentence. For the sentence *I saw the cat* it is able to identify *I* as a noun phrase (NP) and grammatical subject, *saw the cat* as the verb phrase or predicate, *saw* as the head of the verb phrase, *the cat* as a noun phrase and grammatical object of the sentence, *the* as a determiner and *cat* as a noun. The result is a structure resembling a Calder mobile, from which each word "hangs" from a different node.

Third, the engine "looks up" the L1 words in the L2 dictionary. Having parsed the L1 sentence, the engine is able to make necessary decisions about word meanings; for example, it that the word *saw* is a verb rather than a noun because it has already parsed as such in step two.

Fourth, the engine applies a set of complex transfer rules. Some of these rules are general, meaning that they apply robustly, whereas others are specific, meaning they have exceptions or special conditions. How the rules are ordered or interspersed depends on the order in which they must apply to give the right result – a concern that will be familiar to anyone who has studied generative syntax or generative morpho-phonology. The result is a new structure that reflects the word order for the target language.

The transfer system typically operates in only one direction, since the transformation rules are not symmetric. So a different program is needed not only for each language pair, but also for each direction.

3. A comparison of the systems

What about the reliability of translation engines within sentences? How able are they to render idiomatic expressions or subtle nuances of meaning? To answer these questions, I first selected four different translation engines that provide free online translation. I screened each one to ensure that it was actually a separate entity. This step was necessary because some engines are licensed to more than one so-called "portal." For example, the free translation engines found at Google, AOL, Babel Fish, Compuserve, and Lycos all use the same Systran engine. In the selection of portals, I simply chose the one with the most user-friendly interface. The four engines chosen were:

Engine name	Manufacturer	Portal used
Logomedia	Language Engineering Co. Belmont, MA, USA	www.1-800-translate.com
Systran	Systran San Diego, CA, USA	www.dictionary.com
Reverso	Softissimo Paris, France	www.reverso.net
Promt	Promt, Ltd. St. Petersburg, Russia	www.translation2.paralink.com

Next, I compiled a list of sentences based upon those found in a beginning college Spanish text. I used the first thirteen chapters of the second edition of *VISTAS: Introducción a la lengua española*. Some of the sentences I used were taken directly from the book. Others were formulated based on material in the text. All were intended to target specific structures and idioms that students are routinely called upon to learn and as a part of their Spanish coursework.

Then I submitted each sentence for translation to the four different engines. For each sentence, I decided upon a correct translation and allowed for alternate vocabulary and phrasings. Deviations would be considered ungrammatical. The table [see Appendix] shows the English input sentence in the first column, and the desired Spanish translation output in the second column. The actual translation outputs for each engine are found in the next four columns. The far right column summarizes whatever grammatical task or tasks are being targeted in that particular series.

The table also identifies error weight by means of cell shading. Unshaded boxes match a correct translation exactly. Progressively darker boxes identify increasing quantity and severity of errors in the translation. For the purposes of this project, an error was defined as any mismatch between the correct translation and the engine translation. Mismatch could exist between words, word clusters, or a sentence or sentence cluster.

	Sentence is grammatical but is not a reasonable translation of the source sentence.
	Sentence is ungrammatical; it contains 1 error **un**related to the target task(s)
	Sentence is ungrammatical; it contains 1 or more errors related to the target task(s) and/or 2 or more errors **un**related to the target task

Referring to the Appendix, we see that the severity of the errors gradually increases with increasing grammatical complexity of the source phrase or sentence. By the bottom of the table, none of the translation engines is capable of producing consistently grammatical output. Keep in the mind that the sentences follow the increasing difficulty of a Spanish textbook, thus the sentences at the bottom of the table correspond to the skill level of a student who has completed chapter 13 of the textbook. This is roughly two-thirds of the way through the third college semester, which at Middle Tennessee State University is identified as Intermediate Level II.

We also notice how some translation engines perform better in some grammar areas and worse in others. This is because the ability of each engine to translate accurately is only as strong as the computational model that comprises its operating program.

The tables show how effectively each engine carries out specific grammar tasks, and also points to the difficulty in assessing one particular engine for overall quality or effectiveness, or in ranking them with respect to each other.

Since a detailed discussion of the strengths and weakness of each engine is beyond the scope of today's talk, I would like to point out just one significant problem that became apparent over the course of the comparison. None of the engines fared particularly well in making inferences about subjects or objects that referred back to subjects or objects mentioned previously. For example, in fairly straightforward sentence pairs, only Systran (Babel Fish) correctly recovered number and gender assignments from a previous sentence. Here are some examples:

Source sentence(s)	**Official translation**	**Logomedia**	**Systran**	**Reverso**	**Promt**
Does he have the apples? Yes, he has them.	¿Tiene [él] las manzanas? Sí, [él] **las** tiene.	¿Tiene las manzanas? Sí, **los** tiene.	¿Él tiene las manzanas? Sí, él **las** tiene.	¿Tiene él las manzanas? Sí, él **los** tiene.	¿Tiene él las manzanas? Sí, él **los** tiene.
Does she want to buy the apples? Yes, he wants to buy them.	¿Quiere [ella] comprar las manzanas? Sí, quiere comprar**las**.	¿Quiere comprar las manzanas? Sí, quiere comprar**los**.	¿Ella desea comprar las manzanas? Sí, ella desea comprar**las**.	¿Quiere ella comprar las manzanas? Sí, ella quiere comprar**los**.	¿Quiere ella comprar las manzanas? Sí, ella quiere comprar**las**.

Systran was also the only engine able to correctly infer the gender of a pronoun based on a noun mentioned elsewhere in the same sentence:

Source sentence	Official translation	Logomedia	Systran	Reverso	Promt
Take the apple and eat it.	Tome la manzana y cóma**la**.	Tome la manzana y cóma**lo**.	Tome la manzana y cóma**la**.	Tome la manzana y cóma**lo**.	Tome la manzana y cóma**lo**.
Take the apple but don't eat it.	Tome la manzana pero no **la** coma.	Tome la manzana pero no **lo** coma.	Tome la manzana pero no **la** coma.	Tome la manzana pero no cóma**lo**.	Tome la manzana, pero no cóma**lo**.

However, Systran was unable to do this consistently. The following examples show how Systran failed to assign correct feminine gender to a demonstrative pronoun that refers to a noun in the same sentence:

Source sentence	Official translation	Logomedia	Systran	Reverso	Promt
I like this shirt but I'm going to buy that one.	Me gusta esta camisa pero voy a comprar **ésa** (**aquélla**).	Me gusta esta camisa pero voy a comprar **ese**.	Tengo gusto de esta camisa pero voy a comprar **aquél**.	Me gusta esta camisa pero voy a comprar **esto un**.	Me gusta esta camisa pero voy a comprar **aquel**.

None of the translation engines was able to infer correct gender in sentence pairs containing both a direct and an indirect object pronoun:

Source sentences	Official translation	Logomedia	Systran	Reverso	Promt
Do they give her the apples? Yes, they give them to her.	¿[Ellos] le dan las manzanas? Sí, [ellos] se **las** dan.	¿Le dan las manzanas? Sí, se **los** dan.	¿Le dan las manzanas? Sí, le **los** dan.	¿Le dan ellos las manzanas? Sí, ellos se **los** dan.	¿Le dan ellos las manzanas? Sí, ellos se **los** dan.

The outermost limitation of the translation engine is therefore clear: although it may be anywhere from rather poor to excellent at translating isolated phrases or sentences, it cannot be relied upon to make basic sentence- or discourse-level inferences. For this reason, it should be obvious that using a translation engine to translate an entire themed paragraph, essay, or composition, is an invitation for disaster.

5. Can you identify a composition that has been "engined?"

Although it might be possible to identify telltale signs of the translation engine's handiwork, one should be cautious about making assertions to students that one is absolutely able to identify work that has been assisted by a translation engine. If we let our suspicion be roused whenever student work is "too good," we will send the signal – perhaps even subliminally - that excellent work is somehow suspicious and therefore not to be pursued. This is not a signal that as educators we should wish to send.

That said, there are a few characteristics that do *seem to suggest* a translation engine error rather than a student error. Note that students may also be capable of these same mistakes in their writing, and we have no real way of knowing for sure one

way or the other. However, my own teaching experience suggests that these mistakes are unusual for students working unaided. One reason is that the mistakes display strong morphological and syntactic competence alongside rather poor lexical competence. Students typically display the opposite: they tend to have, at least in the initial learning stages, stronger lexical competence and weaker morphological or syntactic competence.

Hint #1: Morpho-syntactically solid yet semantically awkward

Since translation engines seem to be stronger at translating syntactic and morphological structures and weaker at discerning meaning, one often encounters odd sentences that are structurally sound but have little or no meaning in the L2. The following sentences, all of which are from the comparison tables, illustrate this phenomenon:

**Conseguí vestido rápidamente.*	*I got dressed quickly.*
**Nos caímos dormidos en la hierba.*	*We fell asleep on the grass.*
**Pedro pone encendido su camisa.*	*Pedro puts on his shirt.*
**A partir de tiempo al tiempo visito España.*	*From time to time I visit Spain.*
**Conseguimos adelante bien.*	*We get along well.*
**Nos ponemos a lo largo bien.*	*We get along well.*
**Los niños eran seis años de viejo.*	*The children were six years old.*
**Ana se da una lluvia.*	*Ana takes a shower.*

Hint #2: Unexpected misread

On a smaller scale than the blatant semantic misrenderings shown above, unexpected misreads show a high level of morphological or syntactic accuracy within a short phrase but also contain an unlikely interpretation of one or two isolated words. Since the translation engine is incapable of rejecting an output on the basis of probability, it is capable of really preposterous judgments of meaning. The misjudgments illustrated by the following examples are are most likely the result of an information flaw in the transfer system itself.

Jaime is boring.	misread:	*Jaime está **agujereando***.	[boring with a drill]
	correct:	*Jaime es aburrido.*	
lost in the woods	misread:	*perdido en las **maderas***	[plural of wood]
	correct:	*perdido en el bosque*	
in the living room	misread:	*en la sala **viva***	[the room that is alive]
	corrected:	*en la sala de estar*	

Hint #3: Inference failure

As we have already observed, morphosyntactic features such as number and gender hold unreliably both within sentences and across sentence boundaries. This failure becomes particularly apparent with gender-neutral pronouns like *they* and *them*, or *it*. Of the four engines studied, only Systran is able to infer the gender of a pronoun referred to elsewhere in the same sentence (across an independent clause boundary) or in a different sentence. What is peculiar about the following examples is that they

contain a correct agreement within the first clause, and an incorrect agreement in the second clause:

Source sentence	Official translation	Logomedia	Systran	Reverso	Promt
I have a modern house, and it's beautiful.	[Yo] tengo una casa moderna, y es hermos**a**.	Tengo una casa moderna, y es hermos**o**.	Tengo una casa moderna, y es hermos**a**.	Tengo una casa moderna, y es hermos**o**.	Tengo una casa moderna, y es hermos**o**.
I have a modern house. It's beautiful.	[Yo] tengo una casa moderna. Es hermos**a**.	Tengo una casa moderna. Es hermos**o**.	Tengo una casa moderna. Es hermos**a**.	Tengo una casa moderna. Es hermos**o**.	Tengo una casa moderna. Es hermos**o**.

However, as we have seen, Systran is unable to handle other types of agreement, such as object pronoun agreement. Thus none of the engines is able to make correct discourse-level inferences more than part of the time.

Hint #4: Some English words are simply not translated

Most translation engines are designed to "give up" on source words that are not in their lexicon, or on syntactic structures that, for whatever reason, their rules cannot parse. In these instances, the engines simply cut-and-paste the source word(s) untranslated. Of the systems compared, Logomedia seems to get stuck most easily, and often with words and structures that are not particularly rare complex. When it encounters such items, it simply leaves them untranslated, as shown in the following six examples:

Source sentence	***Logomedia target sentence***
I got dressed quickly.	**Get dressed** rápidamente.
I have as many books as you.	Tengo **as many** libros como usted.
This exam is my worst one.	Este examen es mi peor **one**.
Thank you for the gift.	**Thank you for** el obsequio.
We get along well.	**Get along** bien.
I doubt there is life on planet Mars.	Dudo que hay vida en el planeta **Mars**.

In a similar vein, the translation engine's lexicon will also fail to recognize English words that have been misspelled in the source text, and will simply leave them intact (and misspelled) in the target text. Although detail-oriented students may catch these slip-ups, others will assume that the engine must be right and accept its erroneous output without question.

6. Conclusion

In conclusion, it can be said that the common online translation engines Logomedia, Systran, Reverso, and Promt are unequal in their ability to translate from English to Spanish. No single engine performed consistently better than any other in any one area. It was additionally observed that none of the surveyed engines was capable of consistently inferring gender and/or number of a pronoun that referred back to a noun in a previous sentence, and in many cases, the engines were altogether incapable of making correct number and gender agreements within a single sentence.

Taken together, all of these shortfalls reveal a fundamental ineptitude in effectively translating at the discourse level, a problem which makes the regular use of translation engines by students to formulate L2 compositions an extremely precarious gamble.

Bibliography

Arnold, D.J., Lorna Balkan, Siety Meijer, R. Lee Humphreys & Louisa Sadler. 1994. *Machine Translation: An Introductory Guide*. London: Blackwell.

Blanco, José A. & Philip Redwine Donley. 2004. *VISTAS: Introducción a la lengua española*. Boston: Vista Higher Learning. 2nd ed.

Homstad, Torild & Helga Thorson. 2000. Writing and foreign language pedagogy: Theory and implications. In *Writing across Languages*, ed. by Gerd Bräuer. Stamford, CT: Ablex Publishing. 3-14.

Trujillo, Arturo. 1999. *Translation Engines: Techniques for Machine Translation*. London: Springer Verlag.

The Babel Fish (*Hitchhiker's Guide to the Galaxy*):
http://www.bbc.co.uk/cult/hitchhikers/guide/babelfish.shtml

The Universal Translator (*Star Trek*):
http://www2.truman.edu/~mshapiro/trek.html