

English > Spanish translation in an MT environment

Rosana Wolochwianski

Machine translation (MT): is it a wish or a need? My interest in the topic of MT dates back to the days when I worked as an English-as-a-second-language teacher of business English for companies and professionals. Once, the owner of an insurance company called me to take some English lessons. When I asked him what had motivated him to start the course, he told me that he had entered into negotiations with an important London firm for a huge insurance contract. At a certain point, he had to send the London office a letter – in English, of course. A friend of his gave him a CD with a program that was supposed to translate his letter into fine English. The fact is that he never heard back from his contact after sending that letter, and he lost this potential business. As it is popularly said, there is no second chance at making a good first impression. It was in that moment that I became aware that for those of us who deal with language on an everyday basis, it is quite obvious that the quality of many such programs is poor, and this generates jokes and anecdotes that make us feel that we, the translators, are all-powerful and cannot be substituted.

However, it is evident that the level that technology democratization has reached in this globalized era generates consequences

that we would have never imagined. The quantity of information circulating today is so huge, and the eagerness to access it so urgent that it is almost impossible to think that only a group of qualified professionals producing flawless 300 to 400 words per hour can satisfy such a great demand.

Thus, it is necessary to admit that a good part of this huge amount of information will be processed by MT programs – especially information that, in any case, would never have reached the hands of professional translators, either for lack of time or lack of budget. An alternative to zero translation has emerged. Technology and automatization are no longer an option. They have become a need.

Different approaches to translation automatization

Questions about quality among people devoted to the study and development of automated translation systems have been recurrent ever since MT endeavors started about 50 years ago. For a long time, MT research focused on what is called rule-based machine translation (RBMT). This is the most classic approach, and it represents the type of MT program that is commercially available or accessible through the internet today. It basically tries to imitate how a human translator works, by making use of bilingual dictionaries and a set of lexical, syntactical and semantic rules for each language pair. They are “black box” systems, concerned with input/output and difficult to parse the



Rosana Wolochwianski is an ATA-accredited licentiate in translation with 15 years' experience in the industry. She directs the Center for Translation Studies in Rosario, Argentina.

inner workings of, and it is costly and complex to develop them for every new language pair.

Projects linked to artificial intelligence (AI) aim to introduce enough knowledge of the world into an MT system in order to make it “think” and interpret the way a human being does. Thus, while RBMT relies on computational linguistics, AI projects resort to knowledge representation and engineering. They also work with a set of so-called universal linguistic principles that are supposed to hold across all languages. These projects are usually university-based.

In the 1990s, researchers started to explore new alternative methods. After almost two decades of translation memory (TM) use, huge amounts of aligned bilingual material became available. The new challenge was getting this corpus of aligned material to feed an MT system and, combined with good search engines, render an MT program capable of “learning” through successive translations and of being easily open to new language pairs. That’s where statistics-based MT (SBMT) and example-based MT (EBMT) systems entered the scene, with their probability-driven and pattern-driven approaches, respectively.

There is even a project of context-based MT (CBMT) going on, for which bilingual dictionaries and monolingual extensive corpora are used in order to train a prototype MT system. And, of course, the new automatization efforts aim at combining the best of all worlds, by building hybrid systems: RBMT + SBMT + TM and so on.

Target-language proposed equivalent vs. translation

With more than 50 years of research, different applied technologies, new investments and lots of previously aligned information to feed the systems, MT still does not work so well. It is criticized and gives rise to jokes and anecdotes derived from the incorrect and sometimes funny results it renders, as well as some irresponsible, non-controlled use of its raw results. Why is it so? The preliminary answer is, simply, because translation is a difficult and very complex process. The result any MT program can produce is just a target language proposed equivalent, not a translation in its proper sense of “finding that intended effect upon the language into which [the translator] is translating which produces in it the echo of the original,” according

to Walter Benjamin in his 1921 essay “The Task of the Translator.”

In other words, the meaning of an expression does not exist beyond its use in a given context. That is, there is no pre-existing translation for it that a program can just find and probabilistically deduce or decode. On the contrary, a piece of translation should be “elaborated” on the spot, since it is not a mere transfer of meaning from one language to another. It is more of a negotiation of sense, where meaning is gained and also sometimes lost. Words that have certain nuances or implications in one language, polysemy aside, will often lose them in translation. This is precisely why different translators can use different translations for the same original text or why the same original text can require different translations in different contexts, so retrieving an exact equivalent from a database is just not good enough. In such processes a translator relies on his or her culture, experience, beliefs, assumptions and, above all, interpretation skills and common sense – what we call the pragmatic level of language, our inherently human capacity to “infer” rather than “decode” meaning and produce sense.

Pragmatic processes allow us to close the gap between the semantic representation of a given text and its interpretation as a statement realized within a certain

context. What is said is not comprised only by the conventional meanings, but by the result of reference allocations, disambiguation and the enrichment of some expressions – what takes us from the level of conventional meaning, to that of communication.

Consequently, most of us would agree that in order to obtain a high-quality piece of translation, human intervention is still needed. It could be in the authoring phase, by resorting to “controlled language” rules that make the original text easier to “digest” and/or after the translation process, in a post-editing phase. In any case, it is always useful to know what points should be taken into account when dealing with MT for a specific language pair. Difficulties arise at all levels: lexical, structural, semantic and contextual.

English > Spanish translation pitfalls

Based on my own experience as a translator, the following are the most noticeable problems I have come across in the English > Spanish pair:

- lexical and structural ambiguity
- long chains of premodifiers in English
- use of idiomatic expressions
- multiple uses of the gerund (-ing form) in English

YOUR SINGLE POINT OF ACCESS TO INTERNATIONAL-LEVEL TRANSLATION SERVICES in RUSSIA and CIS countries

Services:

- >> TEP
- >> DTP
- >> QA Procedures
- >> Localization
- >> Interpretation

Fields of Expertise:

- >> Oil & Gas
- >> IT & Telecom
- >> Automotive
- >> Legal & Finance
- >> Medicine



TARGET LANGUAGES:
Russian, Kazakh, Ukrainian, Azeri and Many More

WE MAKE SENSE

+ 7-495-787 3331

MOSCOW, RUSSIA
ML@NEOTECH.RU
WWW.NEOTECH.RU



- double meaning of the verb *to be*: *ser* and *estar*
- passive voice as a discourse impersonality marker (two passive forms in Spanish)

Let's analyze some examples processed by RBMT programs so that the problems become less theoretical and easier to grasp. Although the MT-generated translations are quoted literally and include mistakes that may arise in part from imperfect editing of the original text, I will not comment on all the mistakes. I will just focus on the ones being discussed in each case.

Lexical and structural ambiguity

Language elements are highly productive in nature. Most terms can have more than one meaning, and many of their combinations (syntactical structures) can be interpreted in more than one way. Ambiguity is present all the time along the way when it comes to communication, and the efficacy required to achieve disambiguation is part of human linguistic capacity.

The following examples were taken from a bank services brochure.

English original	Each of these deposit accounts pays a high yield to your business.
MT Spanish version	Cada uno de estas cuentas de depósitos a plazo paga una alta producción a su negocio.
Professional Spanish version	Cada una de estas cuentas de depósito ofrece un alto rédito/rendimiento/beneficio a su empresa.

The noun *yield* in English has more than one meaning, even within a business context. In a context of farming exploitation, it means the amount of food obtained from the harvests or animals, that is, the *produce*. If applied to a monetary investment, it refers to the generated *return* or *profit*. In the MT Spanish version, the word *producción* was chosen, when in fact a financial term was needed – either *rédito*, *rendimiento* or *beneficio*.

What about structural ambiguity? There are endless examples, as every language has a different structural organization, which is sometimes difficult to grasp for a program, no matter how many sets of rules are loaded into it. This one is particularly funny. It comes from the same source as the previous example:

English original	What does this really mean to me?
MT Spanish version	¿Qué hace este realmente malo a mí?
Professional Spanish version	¿Qué significa esto para mí realmente?

The result here is totally misleading, and the problem is double-edged. The verbal auxiliary *does* combined with the main verb *mean* in the original was translated as if *does* were the main verb in its affirmative present form and the main verb *mean* (or *signify*) was translated as an adjective (*nasty*, *unkind*). The individual elements in the structure were allocated different functions, and consequently, the proposed target language equivalent is totally useless in this context.

Use of long chains of pre-modifiers

Certain fields in English, mainly the technical and scientific ones, make use of long chains of pre-modifiers, both adjectives and nouns modifying other nouns. This feature has no parallel

in Spanish. It sometimes becomes very difficult to determine if all the elements modify the last one or if they modify one another along the chain.

Translators usually have to rephrase the whole idea through post-modification or even by splitting the sentence into smaller ones. But MT programs do not seem to be able to establish the proper relationships between the pre-modifying clause and the noun at the end of the chain and process the pre-modifiers as if they were independent phrases, disconnected from one another. Here is an example taken from the medical field:

English original	A randomized, double-blind, active-controlled study to evaluate the efficacy of XXXX ointment.
MT Spanish version	Aleatorizado, el [dilema sin solución], el estudio activo-controlado, para evaluar la eficacia de ungüento de XXXX.
Professional Spanish version	Estudio aleatorizado, doble ciego, con control activo, para evaluar la eficacia del ungüento XXXX.

In this case, we come across three modifications to the noun *study* [randomized + double-blind + active-controlled]. On top of that, two of them are compound modifiers. One of them could not be handled at all by the program, which rendered a “dilemma without solution” message. The noun *study* was only recognized as the nucleus of the last pre-modifying element. Only a professional translator familiarized with this field can grasp the interrelation of all the elements and render a useful Spanish version, after inferring the meaning thanks both to knowledge and experience or even the consultation of a specialist in the field.

Use of idiomatic expressions

Even when we are not dealing with literary language, we come across idiomatic expressions, quite metaphorical in nature, which are incorporated into everyday language.

English original	When you change the options, bear in mind two important properties.
MT Spanish version	Cuándo usted cambia las opciones, el oso en tiene inconveniente en dos propiedades importantes.
Professional Spanish version	Cuando cambie las opciones, tenga en cuenta dos propiedades importantes.

In this example, the English expression *to bear in mind*, which means “to consider, to take into account,” was not detected as such by the program, which broke it down into individual components: [noun + preposition + verb] = English [*bear + in + mind*] = Spanish [*oso + en + tener inconveniente*].

The translator, however, knows by experience that this expression works as a unit, and interprets it in its idiomatic use, rendering the Spanish version *tener en cuenta*.

English original	Employees should keep themselves from getting too close to the mixer while it is running.
MT Spanish version	Los empleados deben mantener a sí mismo de obtener cierra también a la batidora mientras corre.
Professional Spanish version	Los empleados deben evitar acercarse demasiado a la mezcladora mientras ésta está funcionando.

Here, the English verbal phrase meaning “to keep somebody from doing something” is the one generating the main problem, making abstraction of other problems, like the decodification of *too* as *also/también* instead of as *excessively/demasiado*, as well as the decodification of *close* as a verb (*cerrar*) instead of as an adjective (*cerca*). The MT program broke down the phrase into individual components: [verb + pronoun + preposition + verb with *-ing* suffix + adverb + adjective] = English [*keep + themselves + from + getting + too + close*] = Spanish [*mantener + a sí mismo + de + obtener + también + cierra*].

The translator, instead, recognized the expression “to keep somebody from doing something” as an idiomatic phrase with the sense of “preventing somebody from doing something,” as well as the idea behind “getting too close,” and thus produced the Spanish *Los empleados deben evitar acercarse demasiado*.

The translator interpreted the sense of the whole expression, and proposed an equivalent that conveys a similar sense in the target language.

Multiple uses of the gerund in English

The gerund is a highly versatile element in English. Not only can it complement the verb *to be* in the continuous tenses, it can also occupy a nominal position, as we can see in the first example below, or act as an adjective, as we see in the second example. This is the cause of many difficulties for MT systems, as only the comprehension of the communicative intention behind the statement can lead to a proper translation resolution.

English original	Changing diapers can be an exhausting task.
MT Spanish version	Los pañales cambiantes pueden llegar a ser una tarea que agota.
Professional Spanish version	Cambiar pañales puede llegar a ser una tarea agotadora.

Changing diapers was taken by the MT program as a phrase comprised by a noun with a pre-modifying element [modifier + noun]. In fact, *changing* occupies a nominal position, and this is usually expressed in Spanish by means

of an infinitive form, in *cambiar pañales* [infinitive + noun]. Most titles and subtitles in manuals and instructions use this structure headed by the gerund in English (*Changing the battery, Ordering parts* and so on), and they are more than often translated as gerunds in Spanish, producing an unnatural and incorrect target equivalent.

English original	He is charming.
MT Spanish version	Él está encantando.
Professional Spanish version	Él es encantador.

Here, *charming* was taken as the *-ing* complement in a continuous structure and was translated as such, when in fact



TRANSLATION AND INTERPRETING

Translation, traducción, 翻譯 Übersetzung, tradução...

The need for skilled translators and interpreters is greater than ever, here in New York and around the world. NYU's School of Continuing and Professional Studies (SCPS) offers professional level courses and certificates in the world's most-spoken languages, both on-site and online.

TRANSLATION/INTERPRETING PROGRAMS INCLUDE:

- Certificate in Translation
 - Online and on-site: English to Spanish, French to English, Spanish to English, General Translation
 - Online only: Arabic to English, German to English, English to French, English to Portuguese
- Certificate in Medical Interpreting in Spanish/English and Chinese/English
- Certificate in Court Interpreting in Spanish/English
- Professional development opportunities for practicing translators and interpreters
- Foreign Language Proficiency Testing in Arabic, Spanish, Chinese, and many more

For information on all our programs, visit our website or call to receive our new Bulletin.

scps.nyu.edu/x73
1-800-FIND NYU, ext.73

NYU SCPS
SCHOOL OF CONTINUING & PROFESSIONAL STUDIES
At the center of the center of it all.

New York University is an affirmative action/equal opportunity institution. ©2008 New York University School of Continuing and Professional Studies

Language Focus

it is in an adjectival position and denotes an attribute of the subject. Neither of the resolutions in these two examples has proved to be effective.

The double sense of the English verb *to be*: *ser* or *estar*?

One of the most significant difficulties when it comes to English > Spanish automatic translation, as any English speaker who has attempted to learn Spanish can probably tell you, is the Spanish differentiation between the verbs *ser* and *estar*. In general terms, these two verbs make a distinction between permanent and transitory states/attributes. The lack of such a differentiation in English leads to numerous difficulties, as it is not that simple to establish which attributes belong to each of the two categories. Sometimes the attributes can work on both categories, depending on the context.

English original	He is completely drunk.
MT Spanish version	Él es bebido completamente.
Professional Spanish version	Él está completamente ebrio.

In this example, the verb *ser* was chosen, when in fact the idea was that of a transitory state (being drunk), not of a permanent attribute (being a drunkard). Even the presence of the adverbial form *completely*, which indicates a perfective aspect and thus a non-permanent attribute, was not enough of a signal for the program to render a correct solution.

English original	We are extremely tired.
MT Spanish version	Somos cansados muy.
Professional Spanish version	Estamos muy cansados.

Even more noticeable is the problem here, as the adjective *tired* cannot be used as a permanent attribute. We might think that in this case the program lacks enough information for attributes, in connection to permanent/transitory features. Counting on this kind of information could help to reduce mistakes, at least, in non-ambiguous cases.

Use of passive voice as a discourse impersonality marker

Technical discourse is usually strongly marked by an impersonality feature, derived from the fact that the agent that performs the action tends to be irrelevant. The resource used in English to achieve this end is the passive verbal form (*be* + past participle), which can be complemented or not by an explicit agent afterwards. From the Spanish perspective, the use of the passive construction with *se* (which does not admit an agent) is one of the instrumental elements to achieve this effect (besides the use of impersonal forms such as *uno* and *usted*). The periphrastic passive phrase, which parallels the English one, is only used when the agent is relevant or explicit.

English original	Best performance is achieved by using round duct instead of rectangular.
MT Spanish version	Mejor desempeño es logrado utilizando redondea conducto en vez de rectangular.
Professional Spanish version	Se logra un mejor rendimiento utilizando un conducto redondo en lugar de uno rectangular.

English original	Thermal breaks should be used in areas of extreme cold.
MT Spanish version	Interrupciones térmicas deben ser usadas en áreas de frío extremo.
Professional Spanish version	En áreas de frío extremo deben usarse interruptores térmicos.



Translations into Spanish and Portuguese

At D&M Language Services we provide Spanish and Portuguese translations and DTP services only of the highest quality, paying special attention to detail and timeliness. We rely on a wide-ranging team of professional translators and desktop publishers selected following strict criteria based on experience, skill and quality of work. We take pride in fulfilling all of our clients' needs consistently and promptly. We handle every field of expertise and work with SDLX, Trados and WorldServer.

D&M Language Services

Lima, Peru
contact@dmlanguageservices.com
www.dmlanguageservices.com



#1 for Spanish and Brazilian Portuguese

Idea Factory Languages (IFL) is the Latin American market leader in revenues, number of in-house staff, and industry recognition, as well as the only independent language service provider with dedicated production offices in Argentina and Brazil. We are the natural choice for clients with high service and quality expectations, who seek a regional supplier with scalability, flexibility and reliability for a mutually rewarding partnership.

- The right result.
- The right cost.
- The IFL experience.

Idea Factory Languages, Inc.

Buenos Aires, Argentina
iflinfo@idea-factory.net • www.iflang.com

As we can see in the two examples above, the periphrastic passive form was used by the MT program for the Spanish version, paralleling the English, but rendering an unnatural result in Spanish. The MT program could not solve the underlying “discourse genre” problem that could have allowed the selection of the more adequate of the two alternate options in Spanish.

Let's not forget to ask ourselves 'What for?'

Demonizing MT tools is not my intention. It is important to understand that they are just another tool aiding the translation process by resorting to rule application and/or matching efforts. They produce a target language proposed equivalent that, in most cases, needs to be polished or reformulated to become a translation, as it is only by resorting to the human pragmatic capacity of making inferences and generating sense that a translation can be obtained. Even the advocates of MT admit today that, at the

current state of affairs, only fully automatic usable translation can be achieved, and fully automatic high quality translation is still under development.

When used informally, however, for the sake of gisting (cybernauts, chat room users), when devoted to the task of prioritizing localization needs (knowledgebases, intranets, newsletters, patents) or when applied consciously and responsibly, for the right purpose and within a comprehensive translation automatization solution (for example, hybrid systems that combine TM leverage with automatic translation for untranslated text, in a quality-oriented environment), MT tools can become highly productive and cost-effective tools, as several success stories in the localization industry are starting to show.

People who are actually involved in MT post-editing efforts could certainly benefit from systematic studies aiming at highlighting the specific problems each language pair poses. That might help

them be aware of what types of errors to expect, leading to more efficient results in terms of time and quality.

And let us not forget that it is our responsibility, as professionals in this industry, to raise awareness in the market – both among language professionals/companies and among the final users – with respect to the pros and cons, as well as the wise use of these undeniably fascinating tools. **M**

References

Benjamin, Walter. "The Task of the Translator." In *Theories of Translation. An Anthology of Essays from Dryden to Derrida*. Ed. John Biguenet and Rainer Schulte. London: The University of Chicago Press, 1992. 71-82.

Weber, Nico. "Automatic Translation: Background, Problems and Perspectives." In *Multilingualism and Electronic Language Management*. Ed. Walter Daelemans, Theo du Plessis, Cobus Snyman, and Lut Teck. Studies in Language Policy in South Africa 4. Hatfield, Pretoria: Van Schaik Publishers, 2005. 41-60.

The advertisement features a photograph of the modern, illuminated bridge in Buenos Aires at night, with city lights in the background. In the top right corner, there is a logo for 'LOCALIZATION LATIN AMERICA' consisting of a stylized 'X' made of blue and orange geometric shapes. In the bottom left corner, there is a list of event details:

- ◆ November 18-20, 2008
- ◆ Buenos Aires, Argentina
- ◆ www.localizationlatinamerica.com

Join us in Buenos Aires!

Brought to you by the organizers of Localization World.