TECHNICAL AND LINGUISTIC ANALYSIS OF A COMMERCIAL MACHINE TRANSLATION SYSTEM (*POWER TRANSLATOR®*)

GLORIA CORPAS PASTOR

Universidad de Málaga

ANTONIO J. MORENO ORTIZ

Universidad de Granada

1. INTRODUCTION

The demand for translation has greatly increased over the last few years. The translation services of the European Commission are currently generating a great volume of documentation so as to cope with the present level of sophistication of products.[1] Several new computational tools appear every day that aim not only to assist but, in some cases, even to replace the human translator. At present, there is a large market for machine translation (MT). However, not all MT systems produce equally satisfactory target language (TL) texts. As Neubert has rightly pointed out: «you won't get quality, quantity, automation and speed in one bargain».[2]

Lehrberger and Bourbeau[3] distinguish three different types of computerised translation according to the degree of automation achieved and the target user: *machine-aided human translation* (MAHT), *human-aided machine translation* (HAMT) and *fully automatic machine translation* (FAMT). The quality of the output is also relevant in order to establish two further subtypes within the last type:[4] *fully automatic high quality translation* (FAHQT) and *fully automatic low-quality translation* (FALQT).

The main aim of this paper is to review and assess a fully automatic MT system: *Power Translator®* by Globalink Inc. We will be mainly concerned with its technical and linguistic aspects, although some other features will also be tackled. A case study of a legal text will enable us to ascertain the usefulness of such tool, both as a stand-alone instrument for the end-user and as an assistant for the human translator.

---

[1] J. Sager: *Language Engineering and Translation: Consequences of Automation,* Amsterdam-Philadelphia, John Benjamins, 1993.

[2] A. Neubert: «Computer-Aided Translation: Where Are the Problems?» *Target* (1991), 3, 1, p. 56.

[3] J. Lehrberger and L. Bourbeau: *Machine Translation. Linguistic characteristics of MT systems and general methodology of evaluation,* Amsterdam-Philadelphia, John Benjamins, 1988.

[4] A. Gross: «Limitation of computers as translation tools», in J. Newton (ed.): *Computers in translation. A practical appraisal,* London-New York, Routledge, 1992, p. 96-130.

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

2. TECHNICAL ASSESSMENT

Platforms, versions, and features

Power Translator is available in three versions: Power Translator Standard 2.0, Power Translator Professional 5.0, and Power Translator Deluxe 1.0 on CD-ROM. Hardware requirements for Windows 3.x platforms are similar for all three versions: 386 or higher processor, 4Mb RAM (8 recommended) and 22 Mb hard disk space. Power Translator is also available for DOS, Macintosh, OS/2, and UNIX platforms.

We will analyse the Professional and Deluxe versions for the Windows 3.x platform. The two versions have similar capabilities and features, although there are some differences. The most obvious one is that with the Deluxe version, the user may choose not to install the whole program. This may save some disk space, but it also slows down the translation process. Both versions can convert text to speech providing a sound card is installed in the system, but the Deluxe version offers a wide range of choice. Finally, the Deluxe version, unlike its sibling, does not run on the network and cannot be enhanced with specific add-on lexicons. The following description applies to both versions.

Preparing the text

Power Translator can handle different document formats –ANSI, DOS, RTF, DCA– and several word processors: Word (for Windows/DOS/Mac), WordPerfect (for Windows/DOS/Mac), AmiPro, WordStar, Multimate, Interleaf Publisher, Display Write, Office Writer, FrameMaker, Digital WPS, Wang PC, Q&A Write, MacWrite II, and WriteNow. After the translation process, some document attributes are lost, namely, tables, columns, footnotes, and graphics formatting.

The program offers three different methods of translation: loading a document in the source language window, editing the text in it, or file translation. The TL window allows the user to edit the output. When the system comes across difficulties, error codes are generated in the window: the symbols @@ indicate that the word they precede has not been found in the source language (SL) dictionary (cf. *implementary* or *indistictly* in the appendix); the symbols @+ mean that the program has been unable to inflect the word (cf. *extranjera* in the appendix); and the symbols ++ indicate that no translation was available in the TL dictionary for that word. Another important convention is the «suspend translation string». It is marked as «SSTRX» by default, but it can be personalised by the user.

The translation itself is a reasonably fast process. It took 29 minutes for Power Translator Professional to translate a 39,000 word document on a Pentium 100 with 16 Mb RAM (approximately 1,344 words per minute).

The lexicons

The lexicons are located in the subfolder «Dict» and have the extension «.dic». The following lexicons are included with the full installation of the program:

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

— esmicgen.dic	English-Spanish single-word (SW) dictionary
— esmicusr.dic	English-Spanish SW user dictionary
— eesyngen.dic	English synonyms dictionary
— eesugen.dic	English-Spanish semantic units (SU) dictionary
— essuusr.dic	English-Spanish SU user dictionary
— semicgen.dic	Spanish-English SW dictionary
— semicusr.dic	Spanish-English SW user dictionary
— sssyngen.dic	Spanish synonyms
— sesugen.dic	Spanish-English SU dictionary
— sesuusr.dic	Spanish-English SU user dictionary

Subject dictionaries can be easily added in the same subfolder. For example, the setup program for the Legal/Business/Finance subject dictionary (included in the Professional and Deluxe versions) will add the following:

— esmiclbf.dic	English-Spanish SW Legal/Business/Finance dictionary (LBF)
— essulbf.dic	English-Spanish SU LBF dictionary
— esuswlbf.dic	English-Spanish SW LBF user dictionary
— esusulbf.dic	English-Spanish SU LBF user dictionary
— semiclbf.dic	Spanish-English SW LBF dictionary
— sesulbf.dic	Spanish-English SU LBF dictionary
— seuswlbf.dic	Spanish-English SW LBF user dictionary
— seusulbf.dic	Spanish-English SU LBF user dictionary

The English-Spanish version contains some 250,000 entries in the core lexicon. The user can edit and change the information contained in any of these dictionaries, except for the entries labelled «reserved» –i.e. all parts of speech except nouns, verbs, and adjectives. The user dictionaries are meant to store both modifications to already existing entries and new entries that are appended manually. When new information is entered, one of the eight user dictionaries will be updated accordingly. Lexicon updating is a totally user-transparent process. The order to access the dictionaries when searching for a translation equivalent (TE) is as follows: subject user dictionaries (SW®SU) ® subject dictionaries (SW®SU) ® general user dictionaries (SW®SU) ® general dictionaries (SW®SU).

If a TE is not found, the term will be added to the list of unknown words, which can be invoked after the translation is completed. The user then can either discard the term or append it to the lexicons through the appropriate dialogue box. When a new word is inserted the user has to provide the part of speech and, depending on the word, information related to the inflective properties of the word. The appropriate codes can be selected from the pull-down menu.

Unlike other MT systems, Power Translator does not keep track of lexicon updating. We consider this a desirable feature and one of the weak points of the system.

After a cursory examination of a good number of translated texts, both general and specialised, we have decided to restrict our analysis to a case study concerning one type of text within the legal discourse. Such exhaustive analysis would fall beyond the scope of this paper.

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

3. TRANSLATING LEGAL TEXTS: A CASE STUDY

Big companies generate a great number of multilingual documentation every day. Translating becomes an important part of the business. When companies do not include a team of professional translators among their staff, a good MT system can prove very valuable. Especially in those cases where what is needed is just a draft version in order to get the gist of a given document. In the following sections we are going to explore the potential of *Power Translator*® as an instrument for the end-user and as a tool for the human translator. Although we have tried the program on a wide range of legal texts, due to lack of space we will have to restrict our analysis to the translation into Spanish of an American power-of-attorney (see the appendix).

Terminology

According to Alcaráz Varó,[5] legal terms fall into three main classes:

a) terms which only appear within the legal register (e.g. *committal*)

b) ordinary words which also have a special sense within the legal register (e.g. *find* in *to find guilty/innocent*)

c) legal terms that have entered the ordinary language and have kept their specialised senses (e.g. *lawsuit*).

Since *Power Translator*® has access to the LBF subject dictionary, it could be assumed that problems are reduced to a minimum at this level. Nevertheless, this is precisely one of the most serious problems of the system. For instance, the very name of the document *power-of-attorney* is wrongly translated as «energice-de-abogado» and *«una autoridad especial de abogado»*. Since the system does not admit the possibility of a verbless sentence, nouns in such circumstances tend to be reinterpreted as verbs. Thus, in the first case «power» is translated as if it were a verb in the imperative form. The second TE is also wrong, as it is a word-for-word rendering of a class *a*) term instead of its standard Spanish equivalent *poder (notarial)*, or rather *poder especial para pleitos* in this case.

Another example of class *a*) is *principal*. It is wrongly translated as *jefe*. This polysemous term means «a person on whose behalf a second person acts»[6] in this context. Therefore *poderdante* is the right translation in such lexico-syntactic environment.

Some archaic terms characteristic of the legal discourse are sometimes wrongly translated (cf. «hereby» as *por lo presente*). Although there are no examples in the sample text, other typical terms such as *doth*, *witnesseth*, or *thereon* are not included in the SL subject and general lexicons.

The class *b*) of legal terms pose some difficulties as regards collocations. For instance, *submit* is another polysemous word. It has ordinary and legal senses, but in this case the second option is selected. In the TL text it is rendered as «someter». However,

[5] E. Alcaraz Varó: *El inglés jurídico, textos y documentos*, Barcelona, Ariel, 1994.

[6] S. Deebank, A. Grogan, R. Innocent, R. Leng and H. Timms: *Harrap's Dictionary of Law & Society*, Kent, Harrap, 1989.

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

as collocator of bases such as *writs*, *documents*, or *appeals*, it should have been translated as *presentar*.

And finally, class *c*) of legal terms can also be a problem. For example, *corporation* is translated by a false friend, «corporación», instead of *compañía*. The Spanish cognate is applied to institutions such as the Chamber of Commerce or the association of solicitors.[7]

Polysemy

The legal discourse is highly polysemous and full of synonyms. Thus it is astonishing that, despite the fact that a synonyms dictionary is available —rather poor for English, practically non-existent for Spanish—, absolutely no account at all is taken of polysemy. This means that there is just one translation for each part of speech of a given lemma (cf. *principal* above). Translating a given term without taking into account any contextual clues runs the risk of producing erroneous matches: *profits* as «utilidades» instead of *ganancias* or *commision* as «junta» instead of *nombramiento*. This is usually the case with prepositions: *above* is always *sobre*, *at* is always *a*, and *before* is always *antes de*. This originates translations errors such as «con oficinas a 182 Park Avenue», instead of «con oficinas en Park Avenue (nº) 182».

The treatment of polysemy is one of the weakest points in the structure of *Power Translator's* lexicons. Actually the program advises the user to «use unambiguous words». This means that the program might produce half-satisfactory results when the text produced by the user is to be translated into a foreign language, but it will be of little or no use for those who need a translation of a given TL text into their mother tongue.

We believe that rather than «advising» the user on how to prepare the text, the program's designers might have made a more extensive use of this feature, giving it the kind of facilities found in «dialogue MT» systems.[8]

Phraseology

Translating specialised phraseology with this MT system can turn out to be a hilarious experience. Let us illustrate our point with the opening sentence: *Know all men by these presents* that... It is a customary opening in some writs which appears in the target text (TT) as *sepa todos los hombres por estos regalos que*... The system has not recognised the archaic syntactic pattern, nor the plural subject. Each word in the clause has been rendered literally and out of context. An equivalent construction in Spanish would be *sépase por la presente* or *conste/conózcase por el presente documento*.

In favour of has also been translated literally as «en el favor de», instead of the equivalent complex preposition *a favor de*. In the same line, *with reference to* is trans-

[7]Cf. M. A. del Arco Torres (coord.): *Diccionario básico jurídico,* Granada, Comares, 1989².

[8]Cf. H. L. Sommers: «Current Research in Machine Translation», in J. McNaught (ed.): *Proceedings of a Workshop on Machine Translation,* Manchester, UMIST, 1990.

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

lated as *con* la *referencia a*. The insertion of the article destroys the semantic unit of the TL equivalent (*con referencia a*). This is a recurrent error in *Power Translator*.

Collocations also give rise to a number of specific problems as regards the renderings of collocators. For instance, when *grant* collocates with *a power-of-attorney*, the correct TE is *otorgar*, not «conceder», as it appears in the text sample. Another common collocation within the legal discourse is *to enact a law*. It appears redundantly as «legislar una ley» in the TT, even though in Spanish the corresponding TE is *promulgar una ley*. Similarly, the correct rendering of *expire* in Spanish is the cognate *expira*, not «vence», as it appears in the TT. *Vencer*, on its turn, typically collocates with *plazos* ('terms') in the legal discourse.

Syntax

Wrong sentence segmentation of the ST gives rise to many mistakes in the output. For instance, premodification in complex noun groups is not always detected. Thus, *administrative* or *judicial proceedings* appears as «administrativo o diligencias judiciales». In this case, *administrative* also refers to the noun: *procedimientos administrativos o judiciales*.

In cases where the noun group includes several determiners, the program usually interprets one of them as pronoun or noun, translating the noun on its own. Thus, *all other acts* becomes «todos otros actas». Incidently, *acts* has only one TE: *actas*.

Gender agreement also poses difficulties in translation. Let us consider the following translated sentence: *tales mandamientos, documenta, instrumentos particulares o públicos, aplicaciones, las instancias, las apelaciones como pueden ser necesarias...* Gender agreement is done on the basis of the last noun «apelaciones», which is feminine, notwithstanding the fact that in Spanish gender agreement is in the masculine when both feminine and masculine nouns are involved. Other parsing mistakes concerns the phrase *to be notified*. Since the program DOS not identify the right subject (the attorney), a gender and number agreement is made as if *to be notified* was just a postmodifier of *appeals*.

Leaving aside obvious translation errors, such as «documenta», the previous sentence also illustrates problems with articles and verbs. This is a common feature of the program. Definite articles abound throughout the text. The system appears to have a rule to insert a definite article whenever an English noun is translated into Spanish: «el *notario público*», «los *Estados Unidos de América*»; and «las *instancias*», «las *apelaciones*» (but just «*mandamientos*» or «*instrumentos*»). Verb tenses, on the other hand, tend to be mistranslated when the Spanish subjunctive mood is involved. This is the case of modal verbs (*may*, *can*), infinitives in purpose clauses and other subordinate clauses which require the main verb in the subjunctive in Spanish. So, the indicative forms «puede» and «puedan» in the TT should be *pueda* and *puedan*.

Other

Additional problems include the translation of dates, encyclopaedic information and style. Dates are translated literally without paying any attention to the style conventions

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

of the TL: *...June 22, 1997* is simply «...Junio 22, 1994» instead of «el 22 de julio de 1994». Even the English suffixes for ordinal numbers are not recognised by the program and carried over into the TT: *this 18th day of October, 1994* is rendered as «este 18th día de Octubre, 1994», instead of «a 18 de octubre de 1994».

Institutional bodies and names of laws, decrees or acts are not usually included in the LBF subject dictionary. Therefore, such segments are usually mistranslated in a word-for-word fashion (e.g. *Authority of Application* or *Foreign Investment Law* Nº 20.557).

On the other hand, TL style conventions are also frequently lost. For instance, *by* is used in English to introduce the signature, whereas the customary convention in Spanish is *firmado*, not «por».

4. CONCLUSION

The analysis carried out has revealed serious weaknesses both in the case of general and specialised texts. Actually, it did not make much difference for the final output whether the LBF subject dictionary was used or not.

As we have attempted to demonstrate, this MT system has proved to have serious shortcomings at the morphological, syntactic, lexical-semantic, terminological, phraseological and stylistic levels. Although it is true that Computational Linguistics, as its present stage, does not have a definite answer for some of these aspects, others could easily have been solved. For instance, the program shows silly mistakes at the graphemic level such as translating the English personal pronoun *I* as *Yo*, always with a capital letter.

The account of polysemy is obviously inadequate. An MT system that aims to translate natural language must not ignore this linguistic phenomenon, as it will unfailingly lead to disappointing results. It is clear that semantic disambiguation is not an easy task, but a serious MT application should at least provide partial solutions to the problem. A very simple one would have been to perform a basic context checking which would eliminate a number of possible translation equivalents and then present the user with a list of possible candidates. In this way, basic interactive capabilities would be added to the system.

However, the system does have certain positive features. As we have already mentioned, it is possible for the user to add and define lexical information, although more flexibility in this respect would be desirable. Its speed of execution is fairly reasonable. We also find it to be a good point the inclusion of a Spanish grammar in hypertext format (Windows «.hlp» file). On the other hand, we miss an English grammar. This is due to the fact that the application is primarily aimed at the English-speaking market.

As for the usefulness of the program our estimation is that it is of little help for the professional translator, as the TT needs to be dramatically changed in order to achieve a satisfactory rendering. These modifications affect both lexical items and the overall textual structure. This means that postediting takes a long time, and the trade-off is

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

generally not worthwhile. Nevertheless, Power Translator may prove fairly helpful for those end-users whose needs are as simple as getting the gist of the text.

APPENDIX

1. Source language text

Power-of-Attorney

KNOW ALL MEN BY THESE PRESENTS that SSTRX Python Chemicals Ltd. SSTRX, a corporation with offices at SSTRX 182 Park Avenue, West Cauldron, SSTRX New Mexico, United States of America, hereby grants a special power-of -attorney in favour of SSTRX Pedro F. Pérez SSTRX, acting alternately, separately or indistinctly, to act on behalf of the principal before the Authority of Application of the Foreign Investment Law N° 20.557 and its implementary decree N° 413/91 or of any such law on Foreign Investments as may be enacted in the future and before the Registry of Foreign Investments, intervening in all kind of administrative or judicial proceedings pending at present; or that may arise in the future with reference to the investment of the principal in SSTRX «Buenos Aires (Argentina) S.A. Comercial e Industrial» SSTRX, to the registration of the original capital and to the reinvestment of profits, being empowered for the above purposes to sign and submit such writs, documents, private or public instruments, applications, petitions, appeals, as may be necessary, to be notified of resolutions and to withdraw copies thereof, and to perform all other acts which may be necessary or convenient for the better fulfilment of this power-of-attorney.

SSTRX Python Chemicals Ltd. SSTRX

By _____
SSTRX Frederick J. Simpson SSTRX

Subscribed and sworn to before me
this 18th day of October, 1994

SSTRX MARY J. GORDON SSTRX
NOTARY PUBLIC OF NEW MEXICO
MY COMMISSION EXPIRES JUNE 22, 1997

2. Target language text

Energice - de - Abogado

SEPA TODOS LOS HOMBRES POR ESTOS REGALOS que Python Chemicals Ltd. , una corporación con oficinas a 182 Park Avenue, West Cauldron, Nuevo Méjico, los Estados Unidos de América, por lo presente concede una autoridad especial - de - abogado en el favor de Pedro F. Pérez, actuando alternativamente, separadamente o @@indistinctly, para actuar en nombre del jefe antes de la Autoridad de Aplicación del Derecho de inversión @+extranjera N° 20.557 y su @@implementary el decreto N° 413/91 o de cualquier tal ley sobre inversiones extranjeras como puede legislarse en el futuro y antes del Registro de inversiones extranjeras, interveniendo en todo el tipo de administrativo o diligencias judiciales pendientes en la actualidad; o que poder provenir en el futuro con la referencia a la inversión del jefe en Buenos Aires (Argentina) S.A. Comercial e

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes

Industrial, al registro del capital original y a la reinversión de utilidades, siendo apoderado para los sobre propósitos para firmar y someter tales mandamientos, documenta, instrumentos particulares o públicos, aplicaciones, las instancias, las apelaciones, como pueden ser necesarias, para ser notificadas de acuerdos y para retirar copias de eso, y para cumplir todos otros actas que pueden ser necesarios o convenientes para el cumplimiento mejor de esta autoridad - de - abogado.

Python Chemicals Ltd.

Por _____
Frederick J Simpson

Suscrito y declarado bajo juramento ante mí
este 18th día de Octubre, 1994

MARY J. GORDON
EL NOTARIO PUBLICO DE NUEVO MEJICO
MI JUNTA VENCE JUNIO 22, 1997

BIBLIOGRAPHY

Dictionaries

ALCARAZ VARÓ, E. and B. HUGHES: *Diccionario de términos jurídicos. Inglés-Español, Spanish-English,* Barcelona, Ariel, 1993.

Diccionario jurídico Espasa, Madrid, Espasa Calpe, 1993.

GARNER, A. B.: *A dictionary of modern legal usage,* New York, Oxford UP, 1987.

LOZANO IRUESTE, J. M.: *Diccionario Bilingüe de Economía y Empresa, Inglés-Español, Español, Inglés,* Madrid, Pirámide, 1989.

MUÑIZ CASTRO, E. G.: *Diccionario Terminológico de Economía, Comercio y Derecho, Inglés-Español / Español-Inglés,* Madrid, Fontenebro, 1990.

ROBB, L. A.: *Diccionario de términos legales. Español-Inglés/Inglés-Español,* Mexico, Limusa, 1988.

La Palabra Vertida. CORPAS, Gloria; MORENO, Antonio J. Technical and Linguistic ...
Centro Virtual Cervantes