

Analysis of translation errors and evaluation of pre-editing rules for the translation of English news texts into Spanish with Lucy LT

Julia Mercader-Alarcón
Felipe Sánchez-Martínez



Grup Transducens
Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant

ABSTRACT

In this paper we study the effect of pre-editing rules on the quality of the translations produced by the MT system *Lucy LT* when translating English news texts into Spanish. We carried out an error annotation of the first 200 segments of the *News Crawl: articles from 2014* corpus and devised a set of 8 pre-editing rules. The application of these rules to a different set of segments from the same corpus results in a reduction of the word error rate of about 11%.

Keywords: machine translation; patterns of translation errors; pre-editing rules; Lucy LT; machine translation evaluation; error annotation; error analysis; Multidimensional Quality Metrics

RESUM

En aquest article, s'estudia l'efecte de les regles de preedició en la qualitat de les traduccions produïdes pel sistema de TA Lucy LT en la traducció anglès-espanyol d'articles de notícies. Per a això, es va realitzar una anotació d'errors dels primers 200 segments del corpus *News Crawl: articles from 2014* i es van elaborar 8 regles de preedició. El resultat d'aplicar aquestes regles a un conjunt diferent de segments del mateix corpus va ser una reducció de la taxa d'error per paraula d'un 11% aproximadament.

Paraules clau: traducció automàtica, patrons d'errors de traducció, regles de preedició, avaluació de la traducció automàtica, Lucy LT, anotació d'errors, anàlisi d'errors, Multidimensional Quality Metrics

RESUMEN

En este artículo, se estudia el efecto de las reglas de preedición en la calidad de las traducciones producidas por el sistema de TA Lucy LT en la traducción inglés-español de artículos de noticias. Para ello, se realizó una anotación de errores de los primeros 200 segmentos del corpus *News Crawl: articles from 2014* y se elaboraron 8 reglas de preedición. El resultado de aplicar estas reglas a un conjunto diferente de segmentos del mismo corpus fue una reducción de la tasa de error por palabra de un 11% aproximadamente.

Palabras clave: traducción automática, patrones de errores de traducción, reglas de preedición, evaluación de la traducción automática, Lucy LT, anotación de errores, análisis de errores, Multidimensional Quality Metrics



1. Introduction

There are two main activities in the automation of the translation process: *post-editing* and *pre-editing*. On the one hand, post-editing consists of correcting the translation produced by a machine translation (MT) system, which usually contains errors, to make it adequate for the intended purpose. This activity increases, in most cases, the productivity of the translation process and its profits (Plitt & Masselot, 2010). Hence the number of companies that include or are considering to include post-editing in their translation processes is growing everyday. According to Sánchez-Martínez (2012), this rise is because of four main factors: the improvement of MT techniques, the increased availability of resources such as MT software and data, a change in the users' expectations about MT, and a better integration of MT systems in computer-assisted translation tools (e.g. DéjàVu¹, SDL Trados² or OmegaT³).

On the other hand, pre-editing consists of preparing the text before translation to avoid words and constructions —such as unusual grammatical constructions or the use of words between prepositions and phrasal verbs— that are prone to cause MT errors. It is therefore clear that by improving the (automatic) translatability of the source language texts, we may improve the productivity of post-editors. This may be achieved either by using a controlled language or by pre-editing the texts before their translation by means of an MT system.

Writers and pre-editors may use *The Global English Style Guide* that Kohl (2008) devised in order to improve the quality of the texts in two ways: to make them more translatable and to improve their comprehension by non-native speakers. Some of the rules that Kohl includes in his guide are: to keep the verb and the preposition of phrasal verbs together, to reduce as much as possible the use of the passive voice, to avoid subordinate clauses and to try to use short sentences.

In this paper we study the effect of pre-editing rules on the quality of the translations produced by the rule-based MT system *Lucy LT*⁴ when translating English news texts into Spanish. The pre-editing rules to be applied are devised after a careful analysis of the type of errors produced by *Lucy LT*. To this end, we annotated the errors⁵ found in the first 200 segments of the *News Crawl: articles from 2014*⁶ corpus using the open-source software *translate5*⁷ and the Multidimensional Quality Metrics (MQM; Lommel et al, 2014) framework, which we adapted to our needs. MQM provides a framework for defining task-specific translation metrics, and an open and expandable system for describing translation quality metrics using a shared vocabulary of issue types.

To measure the impact of the pre-editing rules we compared the translation produced by *Lucy LT* when the source language text is not pre-edited to the translation it produces when translating a pre-edited text. The results show that the word error rate is around 11% lower when the source language text is pre-edited following the pre-editing rules we have devised.

A similar study, which also proved the positive effect of pre-editing on the MT translation quality, was carried out in 2011 for the English–French language pair (Thicke, 2011) and the translation of texts from the online *Help for SAS Anti-Money Laundering Software*.⁸ However, instead of computing the word error rate, Thicke (2011) measured the post-editing

¹<http://www.atril.com/>.

²<http://www.sdl.com/cxc/language/translation-productivity/trados-studio/>.

³<http://www.omegat.org/en/omegat.html>.

⁴ <http://www.lucysoftware.com/espanol/traduccion-automatizada/kwik-translator/>.

⁵ «An error represents any issue you may find with the translated text that either does not correspond to the source or is considered incorrect in the target language» (Burchardt & Lommel, 2014: 12).

⁶<http://www.statmt.org/wmt15/training-monolingual-news-crawl/news.2014.en.shuffled.gz>.

⁷<http://www.translate5.net/>.

⁸http://www.sas.com/en_us/industry/banking/anti-money-laundering.html.



productivity, that is, the time a post-editor needs to correct an MT output to make it adequate for the intended purpose. She tried with and without a customised MT system —enriched with specific terminology— as well as with and without pre-edited texts. In all cases the use of a pre-edited text results in a productivity increase. Results, however, may vary because pre-editing and post-editing levels may change depending on the nature of the texts, the language pair and the purpose of the translation.⁹ Thicke's study differs from ours in the language pair, the nature of the texts to be translated and the measure used. However, both studies show a reduction in the post-editing effort when Kohl's and other pre-editing rules are applied to the source text.

The rest of the paper is organised as follows. Next section describes two frameworks for the annotation of translation errors: MQM and TAUS DQF. Section 3 then describes the methodology we have applied for the customisation of the framework we have used for error annotation and the design of pre-editing rules. The next two sections are dedicated to the quantitative and qualitative analysis of the errors found, and to the description of the pre-editing rules devised and their evaluation, respectively. The paper ends with some concluding remarks.

2. Frameworks for the annotation of errors

Before applying pre-editing rules to a source text, it is important for ensuring better results to determine the nature of the errors that the MT system produces when translating the type of texts in question, and the best way to achieve this is by performing a systematic error annotation on a sample text and a posterior quantitative and qualitative analysis. In this way, we will be able to determine the most recurrent errors and their nature, and to devise and select the most suitable pre-editing rules to reduce these errors as much as possible.

In literature, we can find two different frameworks for the annotation of errors: the *Multidimensional Quality Metrics (MQM)* (Lommel et al, 2014) and the Dynamic Quality Framework (DQF) by TAUS (Translation Automation User Society).¹⁰ Although these two frameworks were studied separately, since June 2015 (after this study was finished) TAUS and DFKI (German Research Center for Artificial Intelligence) have harmonized their error typology evaluations as part of the European-Union-funded QT21 project.

2.1 Multidimensional Quality Metrics

Multidimensional Quality Metrics (MQM) is a framework developed for the EU-funded QTLaunchPad project by DFKI and partners. Instead of providing a translation quality metric, MQM provides a framework for defining task-specific translation metrics and an open and expandable system for describing translation quality metrics using a shared vocabulary of issue types (Lommel et al, 2014). This set of issue types is composed of 10 general categories since June 26, 2015 (*accuracy, compatibility, design, fluency, internationalization, locale convention, style, terminology, verity and other*) and around 100 subcategories.¹¹

As MQM guidelines recommend, annotators should devise their metrics according to the type of texts they are evaluating and the purpose of the evaluation. They must select only the issue types that are useful to address the relevant research questions. Thus, metrics should not contain extraneous or irrelevant categories, must be granular enough to address the research questions posed and small enough to be kept in mind by annotators (Burchardt & Lommel, 2014).

⁹A French news article translated into Spanish generally will require a lower level of post-editing than a Chinese legal document translated into Spanish.

¹⁰<https://www.taus.net/>.

¹¹For a detailed definition of every issue type, please visit <http://www.qt21.eu/mqm-definition/issues-list-2015-06-16.html>.



Once the metric is devised, annotators should create a decision tree with all the issue types. This tree is useful to maintain the metric in mind and to distinguish between the different issue types. It also has the purpose of maintaining the consistency between annotators' own decisions and, at the same time, between the different annotators that carry out the evaluation. Finally, the annotation is done by using the free/open-source software *translate5*.

As regards the amount of text to be annotated, there is insufficient evidence to confirm that a certain amount of text is required for a complete analysis of errors, but, according to the QTLaunchPad results (Burchardt & Lommel, 2014), it is possible to detect more patterns and errors with a minimum of 100-150 segments.

2.2 TAUS Dynamic Quality Framework

Dynamic Quality Framework (DQF) has been developed by TAUS (Translation Automation User Society) for the human evaluation of translation quality (Görög, 2014). This quality is conceived as dynamic given that it changes depending on the characteristics of the text (i.e. content, intent, target readers, etc.) and it is achieved when the client is satisfied with the results. DQF allows for selecting the more adequate evaluation model according to the specific quality requirements by a flexible framework that is based on three parameters: *utility*, *time* and *sentiment*. Utility refers to the importance of the functionality of the translated content; time, as the name suggests, is the time needed for translating the text; and sentiment is the importance of the impact on the brand image, i.e. how damaging might it be to a client if the content is badly translated.

DQF tools enable users to carry out an adequacy and fluency evaluation, to compare translations, to measure post-editing productivity and to score translated segments based on an error typology.

Error typology-based evaluations are done on a segment basis and errors are annotated according to five categories (*accuracy*, *country standards*, *language*, *style* and *terminology*). Once the error is classified, the evaluator assigns a severity level to each one (*critical*, *major*, *minor* and *neutral*) and the quality of the translation is assessed according to the quantity and severity of the errors found.

Adequacy/Fluency evaluations are also segment-level evaluations and they are recommended when time and resources are not enough to make an error typology-based evaluation. Adequacy is measured by comparing the target sentence with a reference sentence. However, if this is not available, evaluators should use the source text to determine how much of the meaning in the source text is expressed in the translation and then select an adequacy level according to a four-point scale (*everything*, *most*, *little*, *none*).¹² Fluency evaluation has more to do with grammar, linguistics and locale aspects of the translation and does not need a reference sentence. It is also measured in a four-point scale (*flawless*, *good*, *disfluent* and *incomprehensible*).

For the purpose of this paper, we discarded the use of TAUS DQF and used MQM because its set of issue types—with 10 general categories (7 at the time of conducting this study) and around 100 subcategories¹³—is wider and more flexible than that of TAUS DQF. In this way, we were able to select the more suitable issue types according to the needs of our study and perform a more detailed evaluation. Moreover, the free/open-source software *translate5* allows the user to perform a more complete annotation by annotating at the word level, rather than at the segment level, with the possibility of marking the errors on the

¹²This is based on the five-point scale of Przybocki et al. (2008): *all*, *much*, *half*, *little*, *none*.

¹³Please note that the MQM issue types used in this study are part of the version 0.2.0 (2014-08-19), before MQM and DQF harmonisation, where local convention was a fluency subcategory: <http://www.qt21.eu/mqm-definition/issues-list-2014-08-19.html>



translated text, to assign them a level of severity, to add comments, and to extract statistics on the annotation when it is finished.

3. Methodology

Before error annotation it is necessary to define the set of issue types that will be used during the analysis of errors and development of pre-editing rules. We decided to focus on linguistic aspects such as grammar and spelling to devise pre-editing rules, since we are interested in having a machine translation that is correct from both the lexical point of view and the grammatical point of view. Therefore, general categories such as *verity* (for those statements that contradict the world of the text), *internationalisation* (for problems related to the internationalisation of the content), and *design* (for problems related to graphic aspects, vs. linguistic aspects, of the content) were eliminated. The general categories used for error annotation are *accuracy* and *fluency*.

Regarding *accuracy*, we eliminated the subcategories related to terminological aspects because we did not have any termbase to be taken into account. As for *fluency*, we gave priority to *grammatical* and *spelling* categories, so we omitted subcategories such as *content* (for issues related to the presentation of the information), *style guide*, *character encoding*, *non-allowed characters*, *pattern problem*, *sorting*, *corpus conformance*, *broken link/cross-reference* or *index-TOC*, since they are destined to mark issues related to text encoding, order of elements, corpus construction or external links.

The list below shows all categories used for the annotation:

[Accuracy](#)

[Mistranslation](#)

[Omission](#)

[Untranslated](#)

[Addition](#)

[Fluency](#)

[Spelling](#)

[Typography](#)

[Grammar](#)

[Word-form](#)

[Part-of-speech](#)

[Agreement](#)

[Tense-mood-aspect](#)

[Word-order](#)

[Function-words](#)

[Locale-convention](#)

[Date-format](#)

[Time-format](#)

[Measurement-format](#)

[Number-format](#)

Once we determined the issue types relevant to our study, we developed a decision tree, which is a hierarchical categorization of the issue types, to guide the annotator during the error-annotation process. The decision tree we used is shown in Appendix A, which is based on the one described by Burchardt & Lommel (2014: 19).



For the annotation and posterior evaluation of the impact of the pre-editing rules on the quality of the translations produced by *Lucy LT* when translating English news texts into Spanish, we took the first 400 segments from the publicly available *News Crawl: articles from 2014* corpus.¹⁴ The first 200 segments were used as development set for error annotation, analysis and development of pre-editing rules. The second 200 segments were used as test set to evaluate the impact of the pre-editing rules devised on the quality of the translations performed by the MT system. Table 1, provides additional data about these two set of segments.

	# segments	# words
Development	1-200	3,779
Test	201-400	4,092

Table 1: Number of segments and total amount of words in the corpora used for error analysis and development of pre-editing rules (development) and for evaluating the impact of the pre-editing rules on the translation quality (test)

4. Error analysis

The error annotation was carried out on the development set of segments and the error statistics for this set are the following:

- Accuracy: 24
 - Mistranslation: 295
 - Omission: 3
 - Untranslated: 45
 - Addition: 45
 - TOTAL: 412
- Fluency: 39
 - Spelling: 27
 - Typography: 50
 - Grammar: 1
 - Word form: 4
 - Part of speech: 28
 - Agreement: 57
 - Tense/mood/aspect: 119
 - TOTAL: 208
 - Word order: 97
 - Function words: 262
 - TOTAL: 568
 - Locale convention: 2
 - Date format: 1
 - Time format: 0
 - Number format: 4
 - Measurement format: 0
 - TOTAL: 7
 - TOTAL: 691

¹⁴<http://www.statmt.org/wmt15/training-monolingual-news-crawl/news.2014.en.shuffled.gz>.



- TOTAL: 1103

It is worth noting that during the annotation we always tried to assign the most specific issue type to an error. In those cases in which an error could not be assigned a specific subcategory, it was assigned a general category; in our case, 24 errors were assigned the (general) accuracy issue type, as it happens with other general categories.

The most recurrent errors of the MT system *Lucy LT* in these 200 segments regarding accuracy were chiefly mistranslations, followed by untranslated words or phrases and additions. Regarding fluency, the main issues were related to grammar, especially to agreement, tense, mood and aspect of verbs, word order and function words (articles, auxiliary verbs and prepositions). The local convention category has not been very relevant since only 7 errors have been annotated as a locale convention problem.

After this quantitative analysis, we extracted the following 15 patterns of translation errors:

1. Articles are not included in the translation between verbs and nouns in those cases in which it is necessary in Spanish.
2. Past simple is always translated as an imperfect tense and it is incorrect in some cases.
3. Acronyms are not always translated (some exceptions: EU=UE; U.S.= EE.UU.).
4. The second person pronoun (*you*) is always translated as *usted*. Here context would be necessary to determine if it is an issue or not.
5. The Spanish preposition *a* is not included in the translation between verbs and nouns.
6. The verb *to be* is not always translated correctly; it is sometimes translated as *ser* when it should be translated as *estar*, and vice versa.
7. Words composed of prefix + hyphen + noun are not translated correctly.
8. The possessive remains in the translation in cases in which it is not correct in Spanish (e.g. parts of the body).
9. The comma before the conjunction *and* remains in the translation and it is not included before *but*, so the punctuation of the source text remains in the target text, whereas in Spanish it is the other way round.
10. Single quotation marks remain in the translation in cases that angle quotation marks should be used in Spanish.
11. Capital and small letters remain in the translation in cases that they are incorrect in Spanish (e.g. Duchess=duquesa).
12. There are word order errors in the translation of long noun phrases because there is more than one adjective or modifier before a noun.
13. Spanish connective *que* is not included in the translation of completive clauses because it is not present in the source sentence.
14. There are gender and number disagreements in the translation when the referent of the pronouns *this*, *that*, *these* and *those* cannot be determined.
15. There are disagreements in the target language caused by the lack of elements after connectives (i.e. pronouns, auxiliary verbs) when there is a lexical ambiguity in juxtaposed sentences because the system cannot determined the subject.

5. Pre-editing rules and their evaluation

We devised a set of pre-editing rules to minimise as much as possible the translation errors made by *Lucy LT* when translating English news texts into Spanish and described in the previous section. To devise these rules we consulted the guidelines by Kohl (2008). It is important to highlight that pre-editing rules must not conflict with the grammatical rules of the



source language, so it is necessary to find a set of rules that make the source text more automatically translatable, as well as grammatically correct.

The pre-editing rules to reduce the errors made by the MT system *Lucy LT* on the corpus *News Crawl: articles from 2014* are described below: 5 of these rules have been extracted from Kohl (2008) and the other 3 have been specifically devised as a result of this study:

#	Pattern of translation error	Example (before)	Pre-editing rule	Example (after)
1	Articles are not included in the translation between verbs and nouns in those cases in which it is necessary in Spanish	<p>EN¹: <i>The Scottish government announced * details of the pilot scheme in September.</i> (seg. 34).</p> <p>ES¹: <i>El gobierno escocés anunciaba * detalles del *esquema de piloto* en septiembre.</i></p> <p>EN²: <i>But the ruthless competition between them has helped drive a dynamic industry with * broadening international appeal.</i> (seg. 77).</p> <p>ES²: <i>Pero la competencia *competición* cruel entre ellos ha ayudado a conducir una industria dinámica con * ampliar* *atractivo* llamamiento internacional.</i></p>	<p>To include <i>the</i> or <i>a/an</i> before the substantive (Kohl, 2008: 68):</p> <p>¹<i>The Scottish government announced the details of the pilot scheme in September.</i></p> <p>²<i>But the ruthless competition between them has helped drive a dynamic industry with a broadening international appeal.</i></p>	<p>¹<i>El gobierno escocés anunciaba los detalles del *esquema de piloto* en septiembre.</i></p> <p>²<i>Pero la competencia *competición* cruel entre ellos ha ayudado a conducir una industria dinámica con un *atractivo* llamamiento internacional que se *ensancha*.</i></p>
2	Acronyms are not always translated	<p>EN: <i>On January 1, the people of Romania and Bulgaria will have the same right to work in the UK as other EU citizens.</i> (seg. 52).</p> <p>ES: <i>El 1 de enero, la gente de Rumanía y</i></p>	<p>Spell out all the acronym:</p> <p><i>On January 1, the people of Romania and Bulgaria will have the same right to work in the United Kingdom as other EU</i></p>	<p><i>El 1 de enero, la gente de Rumanía y Bulgaria tendrá el mismo derecho al trabajo en el Reino Unido que otros ciudadanos de * UE.</i></p>



		<i>Bulgaria tendrá el mismo derecho al trabajo en el *UK* que otros ciudadanos de * UE.</i>	<i>citizens.</i>	
3	Words composed of prefix + hyphen + noun are not translated correctly	EN: Co-defendant <i>LaFonse Dixon Jr. was convicted at trial.</i> (seg. 124). ES: *Co-acusado* <i>LaFonse Dixon Jr. era condenado *en prueba*.</i>	To eliminate the hyphen: Codefendant <i>LaFonse Dixon Jr. was convicted at trial.</i>	Coacusado <i>LaFonse Dixon Jr. era condenado *en prueba*</i>
4	The comma before the conjunction <i>and</i> remains in the translation and it is not included before <i>but</i> , so the punctuation of the source text remains in the target text when in Spanish it is the other way round.	EN ¹ : <i>Smith says the devil is in the details, and that the EPA appears to be (...).</i> (seg. 54). ES ¹ : <i>Smith dice que el diablo está en los detalles*, y* que el EPA parece estar (...).</i> EN ² : <i>The conservation officers check on them twice a day but they don't have a vet, (...).</i> (seg. 148). ES ² : <i>Los oficiales de preservación comprueban *en sobre ellos* dos veces * un día pero no tienen un *veterano* veterinario, (...).</i>	To omit the comma before <i>and</i> ¹ and include it before <i>but</i> ² : ¹ <i>Smith says the devil is in the details and that the EPA appears to be (...).</i> ² <i>The conservation officers check on them twice a day, but they don't have a vet, (...).</i>	¹ <i>Smith dice que el diablo está en los detalles y que el EPA parece estar (...).</i> ² <i>Los oficiales de preservación comprueban *en sobre ellos* dos veces * un día, pero no tienen un *veterano* veterinario, (...).</i>
5	There are word order errors in the translation of long noun phrases because there is more than one adjective or modifier before a noun	EN: <i>15 November 2013 Last updated at 14:01 By Robert Pigott Religious affairs correspondent, BBC News.</i> (seg. 27).	Do not use compounds but a modification with "of": : substantive + of + complements (Kohl, 2008: 56): <i>15 November 2013 Last updated at</i>	<i>15 de noviembre de 2013 Último actualizado* a las 14:01 Por Robert Pigott, corresponsal de asuntos religiosos, Noticias de BBC.</i>



		ES: <i>15 de noviembre de 2013 por *Última Vez informaba* a las 14:01 Por *asuntos de Robert Pigott Religious corresponsal*, Noticias de BBC.</i>	14:01 By Robert Pigott, correspondent of religious affairs , BBC News.	
6	Spanish connective <i>que</i> is not included in the translation of completive clauses because it is not present in the source sentence	EN: <i>But he said it was possible * they had simply seen officers responding to the initial call.</i> (seg. 80). ES: <i>Pero decía que era posible * *habían* visto sencillamente * los oficiales que responden a la inicial *llaman*.</i>	To add the connective <i>that</i> (Kohl, 2008: 117): <i>But he said it was possible that they had simply seen officers responding to the initial call.</i>	<i>Pero decía que era posible que hubieran visto sencillamente *que* los oficiales *que respondían* a la inicial *llaman*.</i>
7	There are gender and number disagreements in the translation when the referent of the pronouns <i>this, that, these</i> and <i>those</i> cannot be determined	EN: <i>Women refugees from Syria are being sexually harassed (...), adding another layer of suffering for those who have fled their homes in search of safety, (...).</i> (seg. 30). ES: <i>*Los refugiados de* mujeres de Siria están *estando* sexualmente *acosados* (...), añadiendo otra capa de sufrimiento para *esos* que han huido de sus casas en busca de seguridad, (...).</i>	To include the referent after the pronouns transforming them into determinants (Kohl, 2008: 105): <i>Women refugees from Syria are being sexually harassed (...), adding another layer of suffering for those women who have fled their homes in search of safety, (...).</i>	<i>*Los refugiados de* mujeres de Siria están *estando* sexualmente *acosados* (...), añadiendo otra capa de sufrimiento para esas mujeres que han huido de sus casas en busca de seguridad, (...).</i>
8	There are disagreements in the target language caused by the lack of elements after connectives (i.e. pronouns, auxiliary	EN: <i>Denis, who currently has 2,600 followers and *presumably found the photo by searching keywords,</i>	To repeat the subordinate connective in the coordinate clause <i>Denis, who</i>	<i>Denis, que actualmente tiene 2.600 seguidores y que probablemente *encontraba* la</i>



	<p>verbs) when there is a lexical ambiguity in juxtaposed sentences because the system cannot determine the subject</p>	<p><i>commented on it: '(...)' (seg. 173).</i> <i>ES: Denis, quien actualmente tiene 2.600 seguidores y probablemente *encontrado la foto por palabras clave *inquisitivas*, *comentadas* sobre ello: '(...)'.</i></p>	<p><i>currently has 2,600 followers and who presumably found the photo by searching keywords, commented on it: '(...)'.</i></p>	<p><i>foto por palabras clave *inquisitivas*, *comentadas* sobre ello: '(...)'.</i></p>
--	---	---	---	---

The sentences that need to be pre-edited could be automatically detected, and even the pre-editing rules could be automatically applied, if a specific module similar to those used by transfer-based MT systems for lexical and structural transfer were developed; in particular a shallow-transfer module like those used by the Apertium free/open-source MT platform (Forcada et al., 2011) would work for most of them. However, for the experiments in this work, the sentences to be pre-edited were manually identified, and the pre-editing rules manually applied to the source text, because the development of such module falls out of the scope of this work.

The issues that are not included in the previous list must be solved by post-editing the MT output, since it was not possible to devise pre-editing rules to prevent those errors without making grammatical mistakes in the source text.

In addition to the annotation of errors, we post-edited the MT output. These post-edited translations were used as a reference translation to which the MT output is compared by computing the word error rate, i.e. the percentage of words that need to be inserted, removed or replaced to convert the translation being evaluated into the reference translation. Table 2 shows, for the development and test corpora, the word error rate of the translation performed by *Lucy LT* when the text to be translated has been manually pre-edited by applying the pre-editing rules described above, and when the text to be translated has not been pre-edited in any way.

Word error rate (%)	Non pre-edited text	Pre-edited text
Development	37.9%	33.4%
Test	38.5%	34.2%

Table 2: Word error rate on the development and test corpora both when the source text is pre-edited and when it is not.

As can be seen, by applying the pre-editing rules we devised to the source text in the development corpus a word error rate reduction of 4.5 percentage points is achieved, i.e. a relative improvement of 12%. The results on the test corpus confirm the positive impact of the pre-editing rules on the translation quality: a word error rate reduction of 4.3 percentage points and a relative improvement of 11% is obtained. The fact that the improvement on the test corpus is almost equal to the improvement on the development corpus indicates that the



pre-editing rules we devised generalise common transformations to be made to the input texts, and that they are not merely solving specific problems that only shows on the development corpus.

It is worth noting that, as we used as reference translation a post-edited translation of the MT output being evaluated, the reduction in the word error rate is directly related to an improvement in the productivity of post-editors, since the word error rate calculated in this way approximates the post-editing effort needed to correct the MT output.

As regards the dependence of the pre-editing rules on the target language to which the English texts are translated, most of them would also be useful when the target language of the translation is not Spanish. This is the case of rule #1, which introduces missing articles between verbs and nouns; rule #2, that expands acronyms; rule #5, which avoids the use of compounds; rule #6, that enforces the introduction of the connective *that* to improve the translation of completive clauses; rule #7, used to include the referent after a pronoun to transform it into a determinant; and rule #8, which repeats the subordinate connective in the coordinate clause to help the MT system determine the subject in juxtaposed sentences.

6. Concluding remarks

In this paper we have presented a study of the effect of pre-editing rules on the quality of the translations produced by the rule-based MT system *Lucy LT* when translating English news texts into Spanish. Before devising the pre-editing rules to apply to the source text we conducted a thorough analysis of the errors made by this MT system. To this end we adapted the MQM framework to the needs of our study and performed an error annotation of the translation errors made by the MT system when translating a development corpus.

We have measured the impact of the pre-editing rules on the quality of the translations produced by translating a test corpus, independent from the one used to devised the set of pre-editing rules, and comparing the translation produced to a reference translation. This reference translation was obtained by post-editing the MT output obtained when no pre-edting rules were applied. The results show a clear reduction in the number of edit operations that need to be performed to turn the MT output into the reference translation.

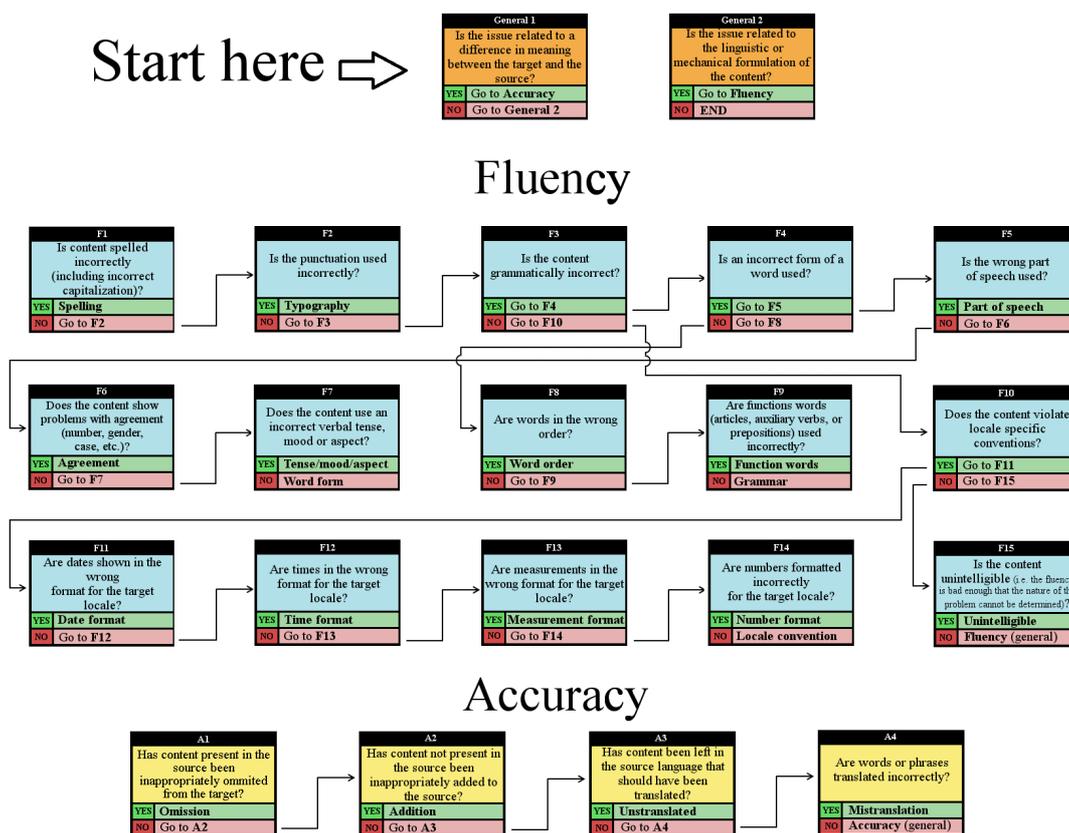
As a future work, it remains to be studied the effect of the pre-editing rules devised on the translations of texts from different domains.

The corpora used for the study, as well as the annotated corpus and reference translations, can be downloaded from <http://www.dlsi.ua.es/~fsanchez/resources/tradumatica/corpora.zip>.



Appendix A. Decision tree

The decision tree we have used for error annotation is shown below; this decision tree is based on the one by Burchardt & Lommel (2014: 19).



Acknowledgements

We thank Marc Mittag (MittagQI) and the *translate5* team for their support and help on the annotation tool and Jaap van der Meer (TAUS) for granting us access to TAUS DQF and the rest of TAUS documentation. Finally we are thankful to Mikel L. Forcada (Universitat d'Alacant) for comments and suggestions on an early version of this paper.

References

- Burchardt, A. & Lommel, A. (2014). *Practical Guidelines for the Use of MQM in Scientific Research on Translation Quality* <<http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf>> [Download on 26 February 2015].
- EMNLP 2015: Tenth Workshop on Statistical Machine Translation. (2015). *News Crawl: articles from 2014* [corpus] <<http://www.statmt.org/wmt15/training-monolingual-news-crawl/news.2014.en.shuffled.gz>> [Download on 8 March 2015].



- Forcada, M. L. et al. (2011). "Apertium: a free/open-source platform for rule-based machine translation". *Machine Translation*, 25(2), 127-144.
- Görög, A. (2014). "Quantifying and benchmarking quality: the TAUS Dynamic Quality Framework". *Tradumática*, 12, 443-454
<<http://revistes.uab.cat/tradumatica/article/view/n12-gorog2/pdf>> [Download on 8 March 2015].
- Hutchins, J. (1996). *Evaluation of machine translation and translation tools*. Norfolk: University of East Anglia.
- Koby, G. S. et al. (2014). "Defining Translation Quality". *Tradumática*, 12, 413-420.
<<http://revistes.uab.cat/tradumatica/article/view/n12-koby-fields-hague-et-al/pdf>> [Download on 8 March 2015].
- Kohl, J. R. (2008). *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market*. Cary (NC): SAS Institute Inc.
- Lommel, A. & Melby, A. K. (2014). *Multidimensional Quality Metrics (MQM) Issue types. Version 0.2.0 (2014-08-19)* <<http://www.qt21.eu/mqm-definition/issues-list-2014-08-19.html>> [Download on 26 February 2015].
- Lommel, A. et al. (2014). "Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics". *Tradumática*, 12, 455-463
<<http://revistes.uab.cat/tradumatica/article/view/n12-lommel-uzskoreit-burchardt/pdf>> [Download on 12 March 2015].
- Lommel, A. et al. (2015a). *Multidimensional Quality Metrics (MQM) Definition. Version 0.3.0 (2015-01-20)* <<http://www.qt21.eu/mqm-definition/definition-2015-01-20.html>> [Download on 26 February 2015].
- Lommel, A. et al. (2015b). *Multidimensional Quality Metrics (MQM) Definition. Version 0.9.3 (2015-06-16)* <<http://www.qt21.eu/mqm-definition/definition-2015-06-16.html>> [Download on 30 June 2015].
- Lommel, A. et al. (2015c). *Multidimensional Quality Metrics (MQM) Issue types. Version 0.5.1 (2015-06-16)* <<http://www.qt21.eu/mqm-definition/issues-list-2015-05-27.html>> [Download on 30 June 2015].
- Plitt, M. & Masselot, F. (2010). "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context". *The Prague Bulletin of Mathematical Linguistics*, 93, 7-16 <<https://ufal.mff.cuni.cz/pbml/93/art-plitt-masselot.pdf>> [Download on 2 November 2015]. See also <<http://langtech.autodesk.com/productivity.html>>.
- Sánchez-Martínez, F. (2012). "Motivos del creciente uso de la traducción automática seguida de posesición". *Tradumática*, 10, 150-156
<<http://revistes.uab.cat/tradumatica/article/view/24/PDF>> [Download on 8 March 2015].
- TAUS. (2011a). *Adequacy/Fluency Evaluation*
<https://www.taus.net/knowledgebase/index.php?title=Adequacy/fluency_evaluation> [Download on 4 March 2015].
- TAUS. (2011b). *Adequacy/Fluency Process Guide*
<https://www.taus.net/knowledgebase/index.php?title=Adequacy/fluency_process_guide> [Download on 4 March 2015].
- TAUS. (2011c). *Dynamic Quality Framework*
<https://www.taus.net/knowledgebase/index.php?title=Dynamic_Quality_Framework> [Download on 4 March 2015].



- TAUS. (2011d). *Error Typology* <https://www.taus.net/knowledgebase/index.php?title=Error_typology_evaluation> [Download on 4 March 2015].
- Thicke, L. (2011). "Improving MT results: a study". *Multilingual* (Jan/Feb 2011), 22(1), 37-40 <http://lexworks.com/cms/wp-content/uploads/2012/07/Improving_MT_Results_Multilingual_01-11.pdf> [Download on 15 March 2015].
- Van der Meer, A. (2015). "DQF and MQM Harmonized to Create an Industry-Wide Quality Standard" <<https://www.taus.net/think-tank/news/press-release/dqf-and-mqm-harmonized-to-create-an-industry-wide-quality-standard>> [Download on 30 June 2015].