Analysing Post-Editing Performance: Correlations with Years of Translation Experience

Giselle de Almeida
Dublin City University
Dublin 9,
Ireland
giselle.dealmeida2@mail.dcu.ie

Sharon O'Brien
Dublin City University
Dublin 9,
Ireland
sharon.obrien@dcu.ie

Abstract

Post-editing (PE) is still a new activity for many translators. The lack of training, clear and consistent guidelines and international standards may cause difficulties in the transition from translation to PE. Aiming to gain a better understanding of these difficulties and using data gathered from a pilot project, this paper explores possible correlations between PE performance and previous translation experience. We test a combination of the LISA QA Model and the GALE post-editing guidelines as a typology for classifying post-editing changes implemented by six post-editors for French and Spanish (three for each target language). This enables a comparison of the types of changes made for the two target languages. We also measure speed and keyboard/mouse activity and link those to translator experience. The insight gathered may be useful for devising future PE guidelines and training programmes.

1 Introduction

With the growing use of machine translation (MT), the market for post-editing (PE) is also expanding. However, there is currently little training available for post-editors, guidelines tend to vary from company to company, and there are no internationally adopted quality measurement standards yet. All of these factors contribute to the difficulties encountered by translators when post-editing.

The over-riding objective of our research is to model post-editing behaviour for two languages which belong to the same language family in order to design specifications for computer-aided support of post-editing, as well as to design a training programme for new post-editors. As part of this research, we conducted a pilot project in June 2009. This was the initial step for a larger scale project, to be carried out in 2010. By analysing PE performance we hope to gather data that may help improve future PE guidelines and training. This paper is structured as follows: Section 2 provides an overview of the pilot project and its objectives and constraints. Section 3 discusses the profile of a good post-editor and introduces the typology we used for classifying the post-editing changes made during the project. Section 4 provides an analysis of the data, including number and types of changes made, productivity and keyboard/mouse data. Section 5 presents conclusions and recommendations for future work.

2 Pilot project

One aim of the pilot project was to try to gather more insights about the influence that previous translation experience may have on post-editing performance. This stems from the frequently-asked question: do highly experienced (and presumably highly efficient) translators also make for good post-editors? It also aimed at testing our methodology (which involved remote recording of post-editing activity) as well as the typology we customised for classifying PE changes (Section 3.2). As part of the analysis, we also tried to identify similarities and differences in PE strategies for two languages of the same language family (French and Spanish for the pilot project). It is important to note that we did not perform a quality assessment of the material, as we are in-

---

This research is funded jointly by the Irish Research Council for Science, Engineering and Technology and VistaTEC Ltd. through the Enterprise Partnership Programme.

terested in measuring post-editing activity. The pilot project was carried out within the constraints of a "live" localisation project at VistaTEC (a Localisation Service Provider). This operational constraint dictated the file format used (Idiom Workbench), as well as the MT engine (Language Weaver). All the participants had previous experience using Idiom Workbench (this was one of the requirements for the localisation project).

The source language file totalled 350 words in English. Although this is a small number of words, it was deemed to be suitable for testing our methodology in a pilot experiment. The subjects who took part in the pilot were professional translators and they were paid the standard fee for their time.

The domain of the text was IT. Again, this was dictated by the constraints of the localisation project. It is worth noting that MT is becoming increasingly common for localisation projects in this domain.

Language Weaver was trained in advance with files from previous localisation projects for Novell (the localisation client). The MT engine was trained with approximately 3,000,000 words for each language pair, using files from previous Novell projects, in order to ensure the quality of the MT output and the use of Novell terminology. Measuring the translation quality of the MT output for the two languages analysed was beyond the scope of our study. However, bearing in mind the volume of the material used to train the MT engine for French and Spanish, and also taking into account that the files used for the training had previously undergone a quality control process by Novell, we can speculate that the quality of the raw MT output would be equivalent for both languages, and satisfactory for the purposes of the live localisation project.

A total of 3 participants for French and 3 participants for Spanish (all native speakers) were invited to take part in the pilot project. They were selected from among the translators taking part in the live localisation project.

For comparison purposes in our analysis, the translators selected had different levels of professional experience (in number of years). Four of the six participants had previous experience with post-editing, while the other two had no previous experience.

We prepared in advance a computer located in VistaTEC's headquarters in Dublin, with the machine-translated file to be post-edited in Idiom Workbench. InputLog (a keylogging program used for research into text production) was used in the background, as was Camtasia Studio (to record the screen in .avi format).

The participants received instructions in advance about PE and how to remotely connect to the computer in Dublin. In individual sessions, each of the participants post-edited the same file (from English into either Spanish or French), while we recorded their actions on-screen.

We subsequently analysed the data gathered in the individual sessions.

3 Measuring Post-Editing Performance

3.1 Profile of a good post-editor

When measuring performance in any activity, it is important to know what could be classified as "good" performance versus "mediocre" or even "poor" performance. Therefore, in order to guide our analysis, we tried to summarise the skills that a good post-editor should have. Offersgaard et al. (2008: 153) propose that many of the characteristics attributed to good translators could also apply to good post-editors. However, according to the authors, there is an important skill that is specific to PE: the ability to decide quickly whether a machine-translated segment can be useful or whether it should be ignored. This has implications not only for speed, but also when changes should or should not be made.

Bearing this in mind, and also taking into account the requirements of the localisation market, our summary of the skills required to ensure good PE performance would be:

1 - The ability to identify issues in the raw MT output that need to be addressed and to fix them appropriately. We call these "Essential Changes";

2 - The ability to carry out the post-editing task with reasonable speed, so as to meet the expectations of daily productivity for this type of activity (approximately 5,000 words post-edited per day, on average);

3 - The ability to adhere to the guidelines, so as to minimise the number of preferential changes, which are normally outside the scope of PE. We call these "Preferential Changes".

3.2 Typology for classifying PE changes

At present, there is no internationally adopted and recognised model for analysing post-editing activity. Therefore, prior to the analysis of the data, we customised a typology to classify the changes made by the participants of the two languages. Our objective was that the typology should be sufficiently broad to cover the main categories of changes made. On the other hand, we opted not to include too many categories, to avoid making the analysis cumbersome and excessively time-consuming.

We surveyed the typologies suggested by different authors for classifying MT errors. Flanagan (1994, p. 65), for instance, proposes a classification with 19 categories. Marrafa and Ribeiro (2001) offer a typology with 13 main categories, many of them with subcategories. Krings (2001, pp. 264-267) and Loffler-Laurian (1996, pp. 96-97) also propose typologies of MT errors, with several categories each. What these typologies have in common is that they were conceived to classify errors in the raw MT output. On the other hand, our goal was to analyse how post-editors deal with the MT output, so our typology should enable this different goal. Also, some of these typologies were linked to a specific type of MT engine, whereas we wanted to use a typology that was not dependent on any one specific MT paradigm, insofar as possible.

Therefore, we chose to customise a typology by combining the categories from the LISA QA Model (The Localization Industry Standards Association 2009) and the GALE Post-editing guidelines (Post Editing Guidelines for GALE Machine Translation Evaluation 2007). Since we intended to analyse the PE work done on machine-translated IT texts, it made sense to use the LISA QA Model as part of our typology, as it is widely used in the localisation industry (Kelly and DePalma, 2009: 7). It is typically used for performing Quality Assessments (QAs), not for assessing post-editing work; however, we assumed that its error categories would be sufficiently broad and clear to cover the main changes implemented in the PE task. We judged the LISA QA model in itself not to be completely transferable to describing PE activity and decided to (a) leave out the severity levels and (b) supplement it with some categories from the GALE Post-Editing Guidelines. Subsequently, during the analysis of the data, we tested the applicability of our typology and reviewed the modifications we implemented.

The categories from the LISA QA Model are: Mistranslation (incorrect translation of the source text); Accuracy (missing or extra information in the translated output); Terminology (inadequate terminology/lexicon for the context); Language (issues related to grammar, semantics, spelling and punctuation); Style (non-compliance with the project's style guide); Country (incorrect country standards, such as currency and decimal separators); and Consistency (non-standardised terminology used in the text). As well as these categories, the LISA QA Model also includes severity levels to classify errors, but we did not include them in our typology, as will be discussed later in this section.

We opted to use the same typology to classify not only essential changes, but also preferential changes, as well as essential changes not implemented. Therefore, our typology has three master categories: Essential Changes, Preferential Changes and Essential Changes Not Implemented. Under each of them, we have the same set of subcategories.

The subcategories from the LISA QA Model that we included in the typology were: Mistranslation, Accuracy, Terminology, Language, Style, Country and Consistency.

The classifications from the GALE Post-editing guidelines that we used as subcategories in our typology were: Extra information in MT output, Information missing from MT output, Adjectives, Adverbs, Capitalisation, Determiners, Phrasal ordering, Prepositions, Pronouns, Proper names, Punctuation, Spelling, Verb tense, Decimal points and Quotation marks. Although the GALE guidelines encompass other categories, we included only those that we anticipated might be relevant for our pilot data set. We also tested the validity of this assumption when analysing the data.

Additionally, we included categories that were not part of the LISA QA Model nor of the GALE guidelines: the main category Format, and the subcategories Gender and Number (under the main category Language). They were added to provide a further degree of detail to our analysis.

One aspect of the LISA QA Model that we did not include was the assignment of severity levels, such as minor and major. This is used for QA purposes, but it was outside the scope of our pilot project, as we intended to analyse and classify the *types* of changes made by post-editors in order to find out more about, and eventually describe, the strategies adopted by them rather than evaluate their work. In other words, our focus was on the process rather than the final product.

4 Analysis of the data

4.1 Correlations between number of changes, total time and translation experience

One of the findings of our analysis was that the two most experienced translators (in number of years) for both languages were also the fastest post-editors, as well as the two participants who made the highest number of essential changes. These were French participant 1 and Spanish participant 2, who have 23 and 13 years of experience as translators, respectively. It is worth noting that French participant 1 had previous experience with several PE projects, while Spanish participant 2, the fastest post-editor of the two languages, only had experience with 2 PE projects prior to this pilot. Based on this, we can speculate that previous translation experience might have an even greater impact on PE performance than previous PE experience. We will need to verify this assumption when we analyse the data from our scaled-up PE project, with a higher number of participants.

On the other hand, Spanish participant 3, who had the least experience in years as a translator (four years), was the slowest post-editor of the two groups.

Although much more data will be necessary to determine if it is possible to find compelling evidence for a correlation between translation experience and PE time, the data from this pilot project hint at the possibility that being an experienced translator would be one of the prerequisites for meeting one of the criteria of a good post-editor, i.e. speed.

4.2 Productivity

By using the values recorded in the pilot project, we extrapolated the values for post-editing productivity for all the participants. Table 1 shows the extrapolated productivity values.

Participant	PE words per hour	PE words per day (8 hours)
FR post-editor 1	1074	8592
FR post-editor 2	618	4944
FR post-editor 3	971	7768
SP post-editor 1	720	5760
SP post-editor 2	1200	9600
SP post-editor 3	540	4320

Table 1. Extrapolated productivity

The productivity calculated for the two fastest post-editors was much higher than the expected average of 5,000 words per day (9,600 and 8,592 words per day, respectively). However, it is important to qualify this: such an extremely high post-editing productivity would probably not be sustainable over a full working day or over long periods, so the actual values might be closer to the average. The same might apply to the other participants: for the same reasons, their actual daily productivity might be lower than the extrapolated figures that we calculated. Still, the daily productivity for PE would be higher than the average productivity normally expected for translation, which would be between 2,000 and 2,500 words per day.

It is interesting to compare these values with the productivity values observed in the Transtype project (Macklovitch, 2006). The Transtype project involved a very different setting from our own project: the testing of an interactive MT system. Several dry runs were conducted with the participation of professional translators, and the languages included were French and Spanish. On average, using the system, the translators were able to achieve productivities of up to 10,000 words a day (this value was extrapolated from their hourly productivity). This could serve as further proof that such systems can indeed help translators gain in terms of productivity.

Regarding the second skill that we identified in our profile of a good post-editor (the ability to identify and to fix the necessary issues in the raw MT output), the two fastest post-editors corrected almost all the errors, with only a few corrections being left out (and this was also done in less time than the other participants).

The third skill included in our profile of a good post-editor (minimising the number of preferential changes), is the only one that does not seem to have been fully met by the two fastest post-editors. The number of preferential changes was high (38 and 25, respectively) and, in the case of the fastest French post-editor, the second highest among all the participants in both languages (38 preferential changes).

An interesting aspect is that the participants with less translation experience made fewer preferential changes, for the most part (for instance, Spanish post-editor 3, who has 4 years of experience as a translator, made 25 preferential changes). Different conclusions might be drawn from these results. Perhaps the more experienced translators felt less constrained by the post-editing guidelines provided for the task (they were told not to make any unnecessary stylistic changes) and, having more professional experience as translators, simply followed their instincts and their experience and implemented all the corrections they deemed fit, even if they were preferential? The less experienced participants, on the other hand, may have felt that they needed to strictly follow the guidelines and, as a result, avoided as much as possible any changes that they considered as preferential (a side effect of this may have been a number of essential corrections being left out).

4.3 Keyboard and mouse use

Table 2 shows the values for keyboard and mouse use for each participant.

Participant	Total time keyboard (seconds)	Total time mouse (seconds)	Switches to/from keyboard and mouse
FR post-editor 1	1018	835	25
FR post-editor 2	410	2523	126
FR post-editor 3	723	1158	44
SP post-editor 1	717	1743	55
SP post-editor 2	426	1835	67
SP post-editor 3	747	2346	75

Table 2. Keyboard and mouse use

The second fastest post-editor for both languages, French post-editor 1, was also the participant who used the keyboard for the longest time among all the post-editors. The fact that this participant also switched from keyboard to mouse and vice-versa fewer times than all the other participants may have been one of the factors that contributed to the fast editing time. As this is also the participant with most experience as a translator, a high level of proficiency with the keyboard may be inferred. Cut and paste actions using keyboard shortcuts instead of constantly switching between keyboard and mouse may increase the overall speed and contribute to more efficient post-editing of the text.

The slowest French translator had the highest total mouse time, the lowest total keyboard time and the highest number of switches between keyboard and mouse among all the participants for both languages. In addition, this was the translator who made the fewest essential changes and who overlooked the highest number of essential corrections among all the participants. The low keyboard time could be a result of the low number of changes made. The high number of switches between keyboard and mouse and the frequent use of the mouse might indicate that the participant felt unsure about the task and about how to proceed - although much more data would be required to back up this supposition. In general terms, however, if we compare the results of this participant with the results from the opposite extreme (the first French participant), it is possible to deduce that a constant use of keyboard actions and a low number of switches between keyboard and mouse might constitute a good strategy for efficiently handling the post-editing task. It is worth commenting that little, if any, empirical data has been published on this topic to date (although psychomotor skills have obviously received attention in general in HCI literature).

However, the results observed among the Spanish translators seem to somewhat contradict the conclusions drawn from the data gathered for the French translators. The fastest post-editor among all the participants (Spanish participant 2), who also made the highest number of essential changes among the Spanish post-editors, had the lowest total keyboard time and the second lowest total mouse time and total number of switches between keyboard and mouse among all the Spanish translators. One aspect to be taken into account is that this participant made a lower total number of changes (74) than French participant 1 (109), who also revised the whole text after finishing the PE task and made further corrections. What should also be considered is that opting to use the keyboard or the mouse is linked to personal preferences, to a certain extent. In addition, Spanish participant 2's low keyboard and mouse time might also be explained by a high level of proficiency in typing and using the mouse.

When we put together the data regarding number of changes, professional experience and keyboard/mouse usage, a few additional conclusions may be inferred.

Of interest among the French post-editors, Participant 1 made the highest number of essential changes, had the lowest PE time, and displayed efficient use of the keyboard and mouse (highly intensive in both cases, particularly for the keyboard, but with the lowest number of switches among all the participants). Among the Spanish post-editors, Participant 2 made the highest number of essential changes, coupled with the lowest number of preferential changes for Spanish and no essential changes were left out. This participant made efficient use of keyboard and mouse (much less intensive than French participant 1, and with a higher number of switches, but this has been previously explained by the higher number of total changes and the revision carried out by French participant 1).

It is also worth mentioning that these two participants had the two highest extrapolated daily PE productivities.

4.4 Comparison between the two languages

Tables 3, 4 and 5 show the absolute number and percentage of changes made in specific categories for both French and Spanish.

Essential changes	**French**	**Spanish**
Accuracy	30 (17%)	33 (21%)
Consistency	10 (6%)	3 (2%)
Format	24 (13%)	24 (13%)
Language	89 (49%)	73 (47%)
Mistranslation	24 (13%)	19 (12%)
Terminology	3 (2%)	4 (3%)

Table 3. Percentage of Essential Changes for French and Spanish and their sub-categories

Preferential changes	**French**	**Spanish**
Language	21 (37%)	39 (46%)
Style	13 (23%)	28 (33%)
Terminology	23 (40%)	17 (20%)

Table 4. Percentage of Preferential Changes for French and Spanish and their sub-categories

Essential changes not made	**French**	**Spanish**
Accuracy	3 (20%)	2 (29%)
Format	2 (13%)	N/A
Language	5 (33%)	4 (57%)
Mistranslation	5 (33%)	N/A
Terminology	N/A	1 (14%)

Table 5. Percentage of Essential changes not made for French and Spanish and their sub-categories

For both Spanish and French, Language was the category with the highest number of essential changes made (49% for French and 47% for Spanish). The high volume of essential changes in the Language category might hint at the possibility that further adjustments would need to be made to the MT engine in order to improve the quality of the raw MT output so that the number of essential changes (and thus PE effort) required to meet the client's quality criteria could be reduced in future projects.[1]

There were also significant percentages of preferential changes in the Language category (37% for French and 46% for Spanish). These high numbers could be interpreted as an indication that the post-editors might need more specific guidelines and examples of the changes that should or should not be implemented in post-editing projects. They could also be due to the fact that some of the translators who took part in the project did not have much previous experience with post-editing. Most likely, this is a combination of these two factors and it has further implications for the preparation of guidelines and training.

The number of preferential changes in the Terminology category was higher than the number of essential changes, both for French and Spanish (40% for French and 20% for Spanish), with post-editors replacing correct words with synonyms, thus making stylistic changes rather than correcting actual errors. This could also be related to lack of PE experience. Participants with PE experience made preferential terminology changes to a lesser degree than those with low or no PE experience. This also has implications for guidelines and training.

The number of essential changes not implemented was low for most of the participants. French participant 2 was the post-editor who overlooked the highest number of essential corrections (14 in total).

Some differences were observed in the preferential changes: there was a higher number of preferential changes in French in the Language and Terminology categories, while the Spanish post-editors made a higher number of changes in the Style category.

Regarding essential changes not made, there were differences between the two languages as well. However, it would also be important to bear in mind that the overall number of essential changes not implemented was very low for both

[1] Having said that, it is also worth noting that the client for this live localisation project was more than satisfied by the final translation quality.

languages (a total of 15 for French and 7 for Spanish).

At least in terms of essential changes made, these results might suggest that similar PE strategies can be found in two languages of the same language family. Further analysis of data from a higher number of participants will be necessary to confirm the validity of this assumption.

5 Conclusions and recommendations

The data from this pilot experiment suggest that more experienced translators are also faster and more accurate post-editors. However, experience as a translator might also lead to a propensity to implement a higher number of preferential (or stylistic) changes, which is often contrary to PE guidelines. Translators do this because they are trained to polish text, but polishing is not always required by clients of machine translation. Therefore, there is a friction between translators' practices and client expectations. We do not propose to solve this dilemma here, but by analysing what translators do while post-editing and by looking at potential connections between experience, training and behaviour, we hope to contribute towards making post-editing less of a challenge.

One of our highly experienced translators was also highly proficient on the keyboard and switched between keyboard and mouse very rarely. However, data for the second most experienced translator do not confirm this trend. This suggests that keyboard and mouse usage is a highly personal preference. Future editor support for post-editing may have to take these personal preferences into account.

Another interesting avenue of exploration might be an analysis of pauses and their correlation with PE activity. Again, this would be outside the scope of the present study, but it will be considered for the scaled-up version of our analysis, taking into account previous work done in this area by Köhn and Haddow (2009) and by O'Brien (2006).

For both languages in our pilot project, the highest percentage of changes was in the category of Language and most of these changes related to gender and number agreement and phrasal ordering. This could be indicative of similar post-editing strategies for languages from the same family. However, it could also be an artefact of the MT engine. More research is required to see if there are common post-editing strategies across languages from the same family. If similar strategies across language families can be confirmed, then it would be helpful to implement an intermediate phase in the MT workflow, which could be used to automatically correct PE issues for both languages. While Statistical Post-Editing (SPE) already ventures in this direction, it is dependent on the availability of previously post-edited material. There is, therefore, also a need for automatic post-processing techniques that make use of regular expressions, macros or a tool for automating the correction of PE issues. An example worth mentioning would be the Pan American Health Organisation (PAHO)'s work practices for MT. Vasconcellos (1986) comments on the use of language-specific macros for quickly moving portions of text or for replacing specific constructions while performing PE tasks. The macros can be useful, for instance, for changing Verb-Subject-Object constructions into Subject-Verb-Object constructions when post-editing texts machine-translated from a Romance language into English. Although this information from PAHO dates back to 1986, PAHO still use such macros in their post-editing practices today (Aymerich and Camelo, 2009).

A full discussion of the PE activity typology was beyond the scope of this paper. However, we were satisfied that our customised version of the LISA QA and GALE Post-Editing Guidelines allowed us to adequately categorise and analyse the PE activity in this pilot project and we plan to use it for the scaled-up study in 2010.

As previously mentioned, we observed a tendency among the very experienced translators to make a high number of preferential changes. On the other hand, less experienced translators seemed to make fewer changes overall. It would also be important to take these findings into account when providing feedback and guidelines to potential post-editors. The level of previous translation experience could be considered one of the factors that would shape the type of training and guidelines provided.

We also observed that, in a few cases, specific issues in the raw MT output were corrected in different ways by each post-editor. An in-depth analysis of the degree of agreement or disagreement in the corrections would be beyond the scope of our research (possibly involving a comprehensive study involving Choice Network Analysis, cf. Campbell 2000), but this aspect will be examined in more detail in the scaled-up version of our pilot project.

As we expand our research and gather data from a higher number of participants in future PE experiments, it will be possible to delineate with more precision these and other trends, in order to help us identify ways to improve PE guidelines and training and minimise the overall PE effort for linguists, thus contributing to job satisfaction.

References

Aymerich, Julia and Hermes Camelo. 2009. *The Machine Translation Maturity Model at PAHO*. Proceedings of the Machine Translation Summit XII, Ottawa, Canada, 403-409.

Campbell, Stuart. 2000. *Choice Network Analysis in Translation Research*. Maeve Olohan (ed.) Intercultural Faultlines: Textual and Cognitive Aspects – Research Models in Translation Studies I, Manchester: St. Jerome, 29-42.

Flanagan, Mary. A. 1994. *Error Classification for MT Evaluation.* Proceedings of the First Conference of the Association for Machine Translation in the Americas, 1994, Columbia, Maryland, USA: 65-72.

Kelly, Nataly and Don DePalma. 2009. *Buyers Step Up Their Quality Measurement Efforts,* Common Sense Advisory Online Report.

Koehn, Philipp and Barry Haddow. 2009. *Interactive Assistance to Human Translators using Statistical Machine Translation Methods.* Proceedings of the Machine Translation Summit XII, Ottawa, Canada, 403-409.

Krings, Hans P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes.* Kent, Ohio: The Kent State University Press, edited/translated by G. S. Koby.

Macklovitch, E. 2006. *Transtype2: The last word.* Proceedings of the 5th International Conference on Languages Resources and Evaluation. Genoa, Italy: 167-172.

Marrafa, Palmira and António Ribeiro. 2001. *Quantitative Evaluation of Machine Translation Systems: Sentence Level.* Proceedings of the Machine Translation Summit VIII, Santiago de Compostela, Spain (no page numbers).

O'Brien, Sharon. 2006. *Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output.* Across Languages and Cultures, Volume 7, Issue 1. Akadémiai Kiadó, Eötvös Loránd University¸ Budapest, Hungary: 1-21.

Offersgaard, Lene, Claus Povlsen, Lisbeth Almsten and Bente Maegaard. 2008. *Domain specific MT in use.* Proceedings of the 12th European Association for Machine Translation conference, Hamburg, Germany: 153-154.

Post Editing Guidelines For GALE Machine Translation Evaluation. [Online]. Version 3.0.2. 2007-May-25. National Institute of Standards and Technology Information Access Division/Speech Group and Linguistic Data Consortium. Available from: http://projects.ldc.upenn.edu/gale/Translation/Editors/GALEpostedit_guidelines-3.0.2.pdf [Accessed 31 May 2009]

Statistical Machine Translation Gets Real: LanguageWeaver Profile. 2006. *Web Globalization Intelligence, December 2006*. Available from: www.mt-archive.info/Global-By-Design-2006-LanguageWeaver.pdf [Accessed 30 March 2010]

The Localization Industry Standards Association. (Homepage). [Online]. Available from: http://www.lisa.org/ [Accessed 31 May 2009]

Vasconcellos, Muriel. 1986. *Post-editing On-screen: Machine Translation from Spanish into English.* Proceedings of Translating and the Computer 8, London (no page numbers).