

Post-Editing Machine Translation As an FSL Exercise

MICHAEL D. KLIFFER

*Department of French,
McMaster University*

Received: 28 March 2007 / Accepted: 5 May 2007

ISSN: 1697-7467

ABSTRACT: This is the second half of a study on the advisability of having undergraduates perform an exercise in post-editing (PE) on output from machine translation (MT). After a sketch of its chequered history, I take up the various ways in which students could be introduced to MT, and conclude that PE is the most useful, mainly because it makes them aware of MT's limitations and fundamental language properties underlying these limitations. Through a review of the literature on PE practices and challenges, especially a psycho-linguistic study by Krings (2001), I highlight crucial differences between professional PE and our pedagogical experiment. An error analysis of the student PE results (English to French) is then compared with analyses of the MT input, student translations of the same text from scratch, and parallel results from the initial, French-to-English experiment. Finally, student comments on the exercise suggest that weaker students especially find it beneficial.

Key words: post-editing, teaching translation, machine-translation, French, error-analysis.

RESUMEN: Este artículo (que continúa la investigación previa) examina la utilidad de la traducción automática (TA) en la post-edición (PE) de textos de los avanzados estudiantes anglófonos de licenciatura en francés. Tras un esbozo de la turbulenta historia de la TA, echamos una mirada a las varias maneras de presentarla a una clase de traducción y concluimos que la PE es de lejos el método más útil, sobre todo por la concienciación que impone de los límites de la TA y de las propiedades fundamentales del lenguaje que subyacen a éstos. Por medio de un breve estudio de escritos relativos a las prácticas y a los desafíos de la PE, y gracias especialmente a un estudio psicolingüístico de Krings (2001), señalamos las principales diferencias entre la PE profesional y el resultado de nuestro experimento pedagógico. Un análisis de los errores estudiantiles en un ejercicio de PE (del inglés al francés) se compara con el ejercicio análogo de la fase inicial (del francés al inglés), luego con los resultados en pura TA, sin PE, del mismo texto, y finalmente con las traducciones de estudiantes hechas sin TA. Para terminar, los comentarios estudiantiles sobre el ejercicio nos llevan a creer que éste beneficia sobre todo a los estudiantes flojos.

Palabras clave: postedición, eneñanza de la traducción, traducción automática, análisis de errores, francés.

1. INTRODUCTION

This is a continuation of a study, Kliffer (2001), which explored the practice of post-editing the output from machine translation (MT) in a third-year French-to-English translation course. In this paper, I will do a comparison between the demands of post-editing (PE) in the commercial world and the conditions prevailing in an intermediate-level translation course (this time, English-to-French) in which PE could be introduced. We will see in which ways commercial and pedagogical domains are antithetical with regard to certain factors required for a viable MT program and in which respects introducing PE at a pre-professional level can help raise students' awareness of challenges facing both human and machine translation. This ambivalent verdict for classroom PE will be brought out via an analysis of a questionnaire which students filled out after completing a PE task.

1.1. Nature and history of MT

I will begin by sketching the nature and history of MT, followed by a look at the specificities of PE, especially in comparison with traditional revision. The historical summary is taken from Bouillon (1993), Hutchins and Somers (1992) and Somers (2003).

MT has had a rather chequered history. Its early years were characterized by a conviction that technology would ultimately permit completely automatic translation. Russian and French proposals in the 30's focussed on an esperanto-based universal interlanguage, while American efforts in the 40's to develop Russian/English MT derived from advances in cryptography and information theory. As early as the 50's, though, Bar-Hillel at MIT concluded that MT would probably always need human intervention. Thanks to advances in computing and formal linguistics, hopes were high at research centres in Europe (including the USSR), America and Japan. The 50's saw work on all 3 major rule-based approaches: direct (specific language-pairs, with term-by-term equivalence lists), transfer (specific language-pairs, with discrete structural and semantic analyses of source and target Ls) and interlingua (a single "universal" intermediate representation, independent of both source and target).

In 1966, the bubble burst. The ALPAC report declared that MT was twice as expensive and half as fast as human translation. Consequently, American research faced a radical slow-down for the next decade. Canada and Europe nevertheless continued to show advances, such as the *Météo* system from TAUM (Traduction Automatique de l'université de Montréal), a rare successful attempt at applying fully automatic translation. The interlingua approach encountered numerous problems in the mid-70's, such as insufficient modularity (a failure in one stage froze the whole process) and loss of surface SL forms which turned out to be crucial for selection of TL structures. Less complex, the transfer approach then appeared more promising.

The 80's saw further theoretical work on both rule-based and statistical programs, the latter based on large corpora. Commercially available systems emerged (e.g. SYSTRAN, LOGOS, METAL for European languages), while Japanese electronic firms Fujitsu, Hitachi and Toshiba produced direct or transfer systems handling mainly Japanese and English, requiring human intervention before and after the actual MT phase, and generally restricted to sub-domains like finance and information technology. The Japanese also funded a multilingual research project funding work in Korea, Mainland China, Taiwan and Malaysia. Researchers

also marshalled Artificial Intelligence and knowledge-based semantic approaches in efforts to improve MT.

This period also saw the appearance of translation tools, including bilingual corpora concordancing, on-line dictionary creation/access, and the Translator's workstation, which bundled multilingual word processing, optical character recognition, terminology management software, facilities for concordancing, and in particular "translation memories" (side-by-side alignment of source text and its translation for future referencing). These tools enable the translator to take full control of the process or have partial input when used along with MT.

Since 1990, corpus-based approaches have proven to be both a viable alternative and a complement to the rule-based approaches that dominated in the 80's in both the transfer and interlingua models. Utilizing advances in stochastic techniques for speech recognition, IBM's Candide was based on the French and English corpora of the Canadian parliamentary record (Hansard). By calculating the probability that a given SL word corresponds to a given TL word aligned with it, Candide produced results purportedly as good as competing rule-based models.

Another corpus-based approach assumes that how an expression has been translated before can be a reliable guide to future translations of it. This 'example-based' MT selects an equivalent phrase from a parallel bilingual corpus, where the alignment has been done through statistical or rule-based methods. The weak spot is the recombination of phrases taken from the corpus, but the advantage is that the output reflects previous good quality translations by professionals.

By the end of the 90's, hybrid efforts were gaining favour. Carnegie-Mellon's CATALYST, for instance, combined a knowledge-based interlingua with a statistical analysis of corpora. With interlingua approaches especially, the trend has been to enrich lexical entries, which address not only syntactic constraints formally handled just via an autonomous syntax but also encyclopedic, non-linguistic information.

1.2. Rational for PE

In spite of the considerable efforts and resources directed over recent decades toward making MT automatic, the most commercially successful systems still rely on human intervention, usually at the pre- and post- MT stages. Pre-editing is appropriate especially for texts of a technical or purely information-conveying nature, where style considerations are limited to clarity, as opposed to literary and even journalistic texts, in which the simplification inevitably resulting from pre-editing would often entail a loss of crucial stylistic features vital to an interpretation which goes beyond the bare-bones message. A psycholinguistic study of PE, Krings (2001) was done without controlled language, i.e. pre-editing, but Krings argues that even if the latter improves PE, it adds significant time and effort to the overall translation task. Post-editing, on the other hand, means that stylistic features are retained and therefore amenable to translation, depending on the time available and the wishes of the person who ordered the translation.

1.3. Why PE is nearly always necessary and is likely to remain so

According to Melby and Warner (1995), commercially viable MT is feasible only with domain-specific, i.e. technical, subjects, which manifest **superficial ambiguity**, wherein a

finite number of concepts are mapped to a lexical-terminological unit, which means exactly one of those concepts pertains in a given situation. (cf. *Météo*, the one example where MT without PE gives high-quality results). With dynamic, general language, MT has always given poor results. Melby & Warner claim this is because non-technical language allows for unpredictable stretching of word and sentence meaning, a condition they call **fundamental ambiguity**, i.e. the semantic indeterminacy of general language rules out a rigorous pre-specification of all translation equivalents within a text. According to Melby and Warner, one consequence of this far-reaching distinction between general and technical texts is that the former will always lie beyond the reach of fully automatic MT, no matter how sophisticated MT software becomes. To quote Vasconcellos (1988a), «Rapid post-editing yields optimum results if texts are checked for MT suitability before they are translated and post-edited.»

1.4. PE training

How is PE different from the traditional revising of human translation (RHT)? According to Krings (2001), PE is typically concerned with recurring, predictable errors, while RHT deals with inadvertent misunderstandings and omissions. PE focuses on words/expressions while RHT often involves correcting overall sentence- and discourse-level interpretations (Vasconcellos (1988a)). Quality expectation is invariably high with RHT, but variable with PE, so the post-editors must know when to avoid over-editing, i.e. be able to differentiate between style- and content-correction.

In particular, PE results from weighing of certain factors. Among these, Allen (2004) mentions user expectations (e.g. is mere gisting acceptable, or is the text to be published?), nature of the content (is it semantically and/or grammatically limited?), the cost/effort issue (are there time or money constraints?), the degree of integration with the MT program and the effect of any pre-editing.

According to McElhaney and Vasconcellos (1988) and Vasconcellos (1988a), PE requires certain strategies non-existent with RHT. These include keeping revision to a minimum, using linguistic “quick fixes” (e.g. exploiting the theme/rheme distinction to avoid excessive word order changes) and word-processing automation (e.g. search and replace), as well as providing feedback to the MT software so that recurring errors can be prevented the next time. To these O’Brien (2002) adds knowledge of tools and terminology coding for both source and target languages in MT dictionaries. O’Brien recommends that PE training be introduced in the last part of an undergraduate programme, because considerable confidence and experience are indispensable. Mossop (2001) states that self-revision and revising others’ work should not be done until students have completed a practicum.

As any translation instructor within a traditional language program will have noticed, the difference between expectations in the classroom and the commercial world is vast. Given issues of cost and efficiency, the painstaking approach which language students are trained to adopt can be a liability in the business world. Indeed, Vasconcellos found that even seasoned translators new to MT have difficulty in accepting stylistic incongruities and prioritising comprehensibility. According to Povlsen and Bech (2001), the ability to work swiftly and methodically proved to be a fundamental skill for achieving both time-savings and via MT

with PE, but their experiment also showed that “rapid post-editing should remain a part-time task for translators to avoid excessive work pressure”.

In spite of such caveats, PE has proven its economic worth: Krings (2001) found that on average, 2/3 of MT output required changes, even though the texts were technical manuals, a type long considered to lend itself well to MT. He also reported that on-screen PE resulted in a 20% time saving over 100% human translation. PE allowed a reduction in cognitive effort for target-text processes, although this was partially offset by an increase in parallel effort for source-text processes, the latter increase due to comprehension obstacles in the MT text and the need to constantly shift attention back to the source. The variable of MT quality showed a skewing: high-quality and low-quality MT both required less cognitive effort than medium-quality because the latter’s degree of dependability is the least predictable. Similarly, Vasconcellos and León (1988b) claim that MT + PE is 2-3 times faster than 100% human translation, although a trainee requires about a month (100,000 words) before feeling comfortable with PE.

A final argument for PE is that Krings’ study found nothing to justify translators’ fear that PE makes them mere proofreaders. He found PE to be very similar to translation from scratch in terms of cognitive processes needed and the frequency distribution of these processes. Improvements to MT programs (e.g. better built-in dictionaries, more link-ups with Translation Memory systems, i.e. customized data-bases of previous translations) and on-screen editing have reduced drudgery for translators assigned to PE.

2. PE: AN EFFECTIVE WAY TO TEACH LANGUAGE MAJORS ABOUT MT

PE, then, makes sense in the business world, but it is also the best avenue for introducing language students to MT. First, it gives them a taste of conditions under which much professional (commercial) translation now takes place. MT is now to be found in many large corporations, largely because human translation for the huge and every-increasing volume of work would be uneconomical. Treating MT at this pre-professional stage demystifies it and drives home its limitations, which MT specialists are gradually reducing but will likely never be able to eliminate.

In principle, an introduction to MT could focus on several areas, including pre-editing, parsing, terminology handling, translation memory and finally post-editing. Parsing is really a domain for the language engineers constructing MT software, terminology is taken up in an advanced course, and translation memory is arguably too specialized a domain for an introduction to MT. There remain pre- and post-editing. The former in many respects goes against the general objectives of the course: controlled language entails ridding the source text of SL peculiarities with a view to making a direct, often word-by-word correspondence between SL and TL, so as to simplify the MT stage. The course in question, however, strives to make students aware, via real data, of how wide the gulf between French and English normally is, despite the abundance of cognates and an often parallel word order. Pre-editing, moreover, needs to be taught in a structured manner that requires considerable time and effort and, to reiterate, sends a message that appears to suppress the language-particular traits which students at this stage need to concentrate on.

Post-editing seems the ideal phase by which to demonstrate the nature of MT. Rather than simply observing an aspect of MT such as parsing or lexicon makeup, the student must complete a task, one which not only puts to the test active translation capability, but also calls on powers of judgment. As Niño (2004a) puts it, PE fosters awareness of the need for self-correction, promotes the students' autonomy in the process of learning a language and encourages the learner to produce a coherent target language text. For Belam (2003), PE is a good way for students to evaluate an MT system and drives home the differences between human and machine translation.

As the analysis of student reactions in the next section will show, the experience gives insight into the natural language challenges that MT researchers have been struggling with for decades. We should stipulate that this exercise is not meant to constitute any kind of rigorous training; its purpose is rather to engender awareness of MT's promise and limits, since our language majors are not to be confused with trainees in a professional translation program. For these reasons, I chose post-editing alone to be the focus of a pedagogical hands-on experiment.

3. THE ENGLISH-TO-FRENCH PE EXPERIMENT

The goals of this experiment were to compare performance between "from-scratch" translation and revised MT output (for convenience referred to here as PE), and to determine student reaction to this exercise. Power Translator Pro (PTP) was the selected MT program, after a comparison with Systran and Softissimo. The first stage consisted of an error count of a passage translated by students without any MT intervention. The next year's class was asked to revise MT output of the same passage, and another error count was done, both of the output and the PE results, using the same error classification. Finally, students filled out a questionnaire on the usefulness of the exercise.

Before giving the results of the error count, I will explain the categories underlying it:

Agreement: incorrect morphological concord, or lack thereof, between subject and verb, subject and reflexive pronoun, or noun and adjective

Anaphora: lack of object pronoun, incorrect person and/or number of possessive or relative pronoun, personal pronoun in place of demonstrative pronoun *ce*.

Article: use of definite article where none is required

Generic: use of partitive instead of definite article, or vice-versa

Literal: translation is word-for-word

Mistranslation: carelessness (e.g. *Zimbabwe* for *Zambesi*, plural for singular, unnecessary word, invented word)

Omission: missing content word (noun, verb, adjective, adverb)

Preposition: incorrect or missing preposition

Punctuation: mainly lack of a comma

Spelling: a typo rather than a grammatical error

Structure: syntactic error, e.g. using an intransitive verb transitively, incorrect pronoun object placement, gerund in place of infinitive or vice-versa

Tense: incorrect verb tense or mood

Word choice: error in polysemy or homonymy, e.g. *au larynx fort* for *loud-larynxed*, *tandis que* >while= (contrastive sense) for *pendant que* >while= (duration)

Word order: e.g. failure to invert subject and communication verb in interpolated clause, placing an adverb after a heavy object

Table 1 gives the results of the raw MT output and the “from-scratch” translation done by six students in the year preceding the PE exercise¹. In Tables 1 and 2, * indicates that $p < .05$ according to chi-square test with 1 degree of freedom, comparing Power Translator Pro and student results. ** indicates that $p < .01$.

Table 1: Error counts of MT and student translation

<u>Power Translator Pro</u>			<u>Student Translations</u> (N = average of 6 students)			
	<u>N</u>	<u>%</u>		<u>N</u>	<u>%</u>	
agreement	5	5.43	agreement	5.67	12.56	
anaphora	1	1.09	anaphora	1.83	4.05	
article	0		article	.17	.38	
generic	4	4.35	generic	.33	.73	
literal	19	20.65	literal	1.83	4.05	**
mistranslation	2	2.17	mistranslation	2.83	6.27	
omission	0		omission	.83	1.84	
preposition	14	15.22	preposition	6.83	15.13	
punctuation	0		punctuation	.5	1.11	
spelling	1	1.09	spelling	3.83	8.48	
structure	17	18.48	structure	7.83	17.34	
tense	3	3.26	tense	1.83	4.05	
word choice	25	27.17	word choice	10.67	23.63	*
word order	1	1.09	word order	.17	.38	
Total errors: 92			Average total errors: 45.15			
Total words: 388			Average total words: 418.67			

See the appendix for the raw MT output and 1 student translation “from scratch” of the article. *Word Choice* is by far the most common error category for both PTP and students, but PTP had a significantly higher incidence, with p just above .01. Polysemy, always a challenge human translators, is handled far worse by PTP. Other categories posing problems for both MT and students are *structure* and *preposition* (with nearly identical incidences for each group), surprising in that the target text was in the students’ native language. Notable

¹ This error taxonomy is largely word-based, i.e. it can’t show the real scope of a syntactic error, as for instance with causative verbs, whose syntactic properties go beyond individual words.

differences include *literal*, accounting for one-fifth of PTP errors but only 4 percent of human errors ($p < .0002$), and *spelling*, almost 4 times more frequent among students, although the difference was not significant. These differences are not surprising: given PTP’s propensity for word-for-word output that often would never occur to a human translator, *literal* is far more often a problem category for non-humans. The converse holds with spelling: just as a word processor avoids most careless spelling errors via its internal dictionary checker, this is the least of PTP’s weaknesses.

Table 2 gives a comparison of raw MT output errors with averaged error counts of post-edited versions. Again, the student numbers are the average of six students’ scores.

Table 2: MT error counts (repeated) and post-editing error counts

	PTP		Student Mean		
	N	%	N	%	
agreement	5	5.43	3	10	
anaphora	1	1.09	0		
article	0		.67	2	
generic	4	4.35	.33	1	
literal	19	20.65	.67	2	**
mistranslation	2	2.17	1	3.33	
omission	0		0		
preposition	14	15.22	6.66	22	
punctuation	0		.33	1	
spelling	1	1.09	1	3.33	
structure	17	18.48	7.66	25	
tense	3	3.26	1.33	4	
word choice	25	27.17	8	26	**
word order	1	1.09	0		
TOTAL	92		30.65		

Again, only *literal* and *word choice* showed significant differences between machine and student output. *Word choice* is again the most frequent category among the post-editing results, but still significantly less than the PTP score ($p = .003$). The other high-frequency categories noted for the “from scratch” translations remain so here, although with slightly different ranking. The most notable difference, between the average total errors of the “from scratch” group (45.15) and the PE result (30.65), is not significant.

3.1. Student comments

Students were asked to make a detailed analysis of the MT output and give a general assessment of the exercise both as a practical way to improve their translations and as a means of gaining insight into the nature and difficulties of translation. The analysis turned out to be a fine-tuned contrast between the target and source texts, e.g. the difference between English directional particle verbs like *reflect off* and their non-directional French equivalents. What is more interesting for our purposes is the second task, that of evaluating MT and considering what insights into translation the exercise provided. Here is a sampling of comments:

“This exercise may help me understand the logic or functioning of computers but it doesn’t help me improve my translations. Its mistakes are mechanical errors (e.g. word order, over-use of the article) which we have already learned to avoid”.

“The exercise demonstrates basic problems underlying any translation, e.g. computer does largely a word-by-word replacement, ignoring context, general ideas (background knowledge, pragmatic conditions), and the culture of the source language”.

“The notion that “sum is equal to more than its parts” is lost: e.g. the machine is incapable of choosing among several possibilities the sense that fits best with the overall message”.

“The translator serves as a filter vis-à-vis the source text; the computer does a pale imitation of such a filter”.

“The exercise shows importance of being aware of concepts, rather than just searching for word or phrase equivalents. It made me aware of how much translation is an art rather than something mechanical”.

“It’s harder to find errors in a badly written text”.

“This exercise was extremely useful. My grammar isn’t very strong, so this exercise helped me check grammar points I’d forgotten”.

“The computer can’t detect fixed expressions; it translates them word-for-word, which results in nonsense”.

Finally, a number of students focussed on specific limitations of MT, namely, that it can’t handle part of speech distinctions, word order shifts, pronoun reference, changes in register, context-dependent senses, polysemy, particle verbs, proper syntactic grouping, tone (e.g. humorous vs. serious), and agreement of discontinuous elements. Rather than merely downplaying the abilities of MT, the comments generally implied that such lacunae made the student aware of how important such factors are in any translation task.

3.2. Comparison with the French-to-English exercise

Table 3 gives the results of the French-to-English and English-to-French exercises, the latter taken from Kliffer (2001):

Table 3

French-to-English	English-to-French
Translations from scratch: PTP (3 texts): 22.9% errors (total errors/total words)	Translations from scratch: PTP (1 text): 23.8% errors
Students (3 texts) average: 10.8% errors	Students (1 text): average: 14.2% errors
Post-editing: (Montagnier text only) average: 3.5% errors	Post-editing: average: 7.2% errors

The PTP output displays about the same error incidence in each direction. The students translating into English, their native language in nearly all cases, show a lower error rate (10.8% vs 14.2%). The post-editing shows the same native-language effect: French-to-English had half the error incidence of English-to-French.² Krings (2001) speculates that if the target text is L2, MT might be more useful because it partially makes up for the translator's competence gap between L1 and L2. Our results cast doubt on this: with L1 the target language, the post-editing results had an error incidence of about a third of the "from scratch" group (10.8% vs 3.5%), while with an L2 target text, the incidence was reduced by only a half (14.2% vs 7.2%). Krings, though, was dealing with professional translators whose L2 competence was undoubtedly higher than the students'. Clearly, more experimentation is needed for us to generalize with any confidence.

3.3. Differences between our experiments and professional PE

A major difference between the two in-class experiments and PE in a professional setting lay in the texts. Students worked on general-interest rather than the specialized texts which are typical of professional translation. This is because the purpose was not to evaluate PE but to impress upon students the limitations of MT: with specialized language, the MT

² Admittedly, the two exercises do not constitute a rigorously controlled experiment. The French-to-English translation from scratch was based on 3 texts, as opposed to only 1 for English-to-French. Moreover, a rigorous experiment comparing L1 vs L2 target text status would require 2 groups, one with L1 as native language, the other with L2, translating the same text, a condition not attainable in this setting, where nearly all students are Anglophones. The other comparison, where all students with the same L1 translate 2 texts, one into L1 and the other into L2, would not be reliable because the inevitable differences between the two texts would constitute an intervening variable. Even if we dealt with the same source text and its translation, and gave the two versions to two different groups, the inevitable differences in translation challenges between L1 and L2 might be partly responsible for any observed differences.

program would have given better results, so that these limitations would have been less obvious. A second difference lay in the level of competence: a professional translator is more qualified than a student; hence the former should be able to edit much faster, especially after training in MT post-editing. Thirdly, aside from the usual time constraints of an assignment, the students faced no productivity demands or efficiency considerations, factors which a professional must always live with. Finally, the students had access to fewer tools, i.e. no translation memory or specialized lexicons and only about half of them used an on-line dictionary.

4. CONCLUSION: PROS AND CONS OF TEACHING PE AT A PRE-PROFESSIONAL LEVEL

After reviewing the literature on professional PE training and the results of my in-class PE experiments, I have concluded that having language students perform PE is a double-edged sword.

On the negative side, we have the disdainful reaction of some good students whom the exercise exasperated because of the low-level, repetitive MT errors. Their attitude recalls the ALPAC report, which dismissed the viability of MT and seriously curtailed American research on it for over a decade. Secondly, the experiment did not allow time for real training, comparable to the minimum of one month which far more experienced professionals need to be comfortable with PE. The result is that students don't really end up with an accurate view of what PE truly entails.

On the other hand, the experience turned out to be a confidence-building exercise for average and even poor students, who until then had always done translations from scratch. In a similar vein, Niño (2004b) reports that PE brought about increases in learners' motivation, participation and, consequently, higher confidence in expressing themselves.

In particular, post-editing drove impressed upon our students the importance of a holistic approach to interpreting the source text and translating by phrase rather than word. The activity also provided them with a taste of what to expect if they undertake a career in translation. From an intellectual standpoint, it gave students insight into the huge challenges which have confronted MT, especially the questions of how to deal with syntactic and lexical ambiguity, non-literal language, and inferencing. In a broader sense, these are also challenges for linguistic theory, and confronting them drives home to the student that they are part of the design-features of human language, i.e. properties that all human translators must constantly bear in mind.

5. REFERENCES

- Allen, J. (2004). "Translation Speed vs. Content Management", in *Perspectives on Machine Translation*. 62 Supplement:8-10.
- ALPAC (1966). *Languages and Machines: Computers in Translation and Linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C. (Publication 1416.)
- Belam, J. (2003). «Buying up to falling down: a deductive approach to teaching post-editing», Paper presented at MT Summit IX workshop on Teaching Translation Technologies and Tools, New Orleans. Available at www.dlsi.ua.es/t4/docum/belam.pdf.
- Bouillon, P. (1993). «Introduction et bref historique», in P. Bouillon and A. Clas (eds.), *La Traductique*. Montréal: Presses de l'Université de Montréal. Pp. 15-20.
- Carbonell, J. G. and M. Tomita (1987). "Knowledge-Based Machine Translation, and the CMU Approach", in S. Nirenburg (ed.), *Machine Translation, Theoretical and Methodological Issues*. Cambridge University Press.
- Coughlin, Josette (1988). "Artificial Intelligence and Machine Translation, Present Developments and Future Prospects", in *Babel* 34:1. 3-9, pp. 1-9.
- Hutchins, W. J. and H. L. Somers. (1992). *An Introduction to Machine Translation*. London: Academic Press.
- Kliffer, M.. (2001). «Exploitation pédagogique de la traduction automatique», in *Distances*. 5.2:83-102.
- Krings, H. P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Kent, Ohio and London: Kent State University Press. (Translated from 1994 German thesis and edited by G. S. Koby).
- McElhaney, T. and M. Vasconcellos. (1988). "The Translator and the Postediting Experience", in M. Vasconcellos (ed.), *Technology as Translation Strategy*. Vol. 2. Binghamton, N. Y.: Center for Research in Translation, SUNY at Binghamton. Pp. 140-148.
- Melby, A. and C. T. Warner. (1995). *The Possibility of Language: A Discussion of the Nature of Language, with Implications for Human and Machine Translation*. Amsterdam: Benjamins.
- Mossop, B. (2001). *Revising and Editing for Translators*. Manchester: St. Jerome Publishing.
- Nirenburg, S. (1996). "The Inflexible Fickleness of Fashion", in *IEEE Expert Intelligent Systems and Their Applications*, 4: 15-16.
- Niño, A. (2004a). "Recycling MT: A Course on Foreign Language Writing Via MT Post-editing", Paper presented at Computational Linguistics UK, University of Birmingham (6th and 7th January 2004).
- Niño, A. (2004b). "MT Post-editing as a Challenge to Promote Grammar Accuracy and to Enhance Foreign Language Written Production", Paper presented at CamLing, University of Cambridge (March 19, 2004).
- O'Brien, S.. (2002). "Teaching Post-editing: a Proposal for Course Content", in *Sixth European Association for Machine Translation Workshop* (pp. 99-106) Manchester, England.: UMIST
- Povlsen, C. and A. Bech. (2001). "Ape: Reducing the Monkey Business in Post-editing by Automating the Task Intelligently", in *Machine Translation Summit VIII* (pp. 283-286) Santiago de Compostela, Spain.
- Somers, H. (2003). "Why Translation is Difficult for Computers", in H. Somers (ed.), *Computers and Translation: A Translator's Guide*. Amsterdam: Benjamins. Pp. 319-340.
- Vasconcellos, M. (1988a). "Factors in the Evaluation of MT: Formal vs. Functional Approaches", in M. Vasconcellos (ed.), *Technology as Translation Strategy*. Vol. 2. Binghamton, N. Y.: Center for Research in Translation, SUNY at Binghamton. Pp. 203-213.
- Vasconcellos, M. (1988b). "SPANAM and ENGSPAN: m-Machine translation at the Pan American Health Organization". in *Studies in Natural Language Processing*. J. Slocum (Ed.) Cambridge, England: Cambridge University Press. Pp. 187-236.

APPENDIX: The Assignment

Dans le cadre d'une recherche sur l'exploitation pédagogique de la traduction automatique, nous vous demandons de produire une version corrigée de l'extrait suivant, produit par un logiciel de traduction. Le texte de départ précède sa traduction. Après avoir fait la correction, pourriez-vous répondre aux questions se trouvant à la page 3.

Original text:

Zimbabwe's bad-tempered bullies

Sure they may mangle and stomp you to death. But rest assured: Being vegetarians, hippos have no taste for human beings.

Are that big-mouthed, loud-larynxed hippo's tonsils pink? Or is it just the setting sun reflecting off the water into its face?

We are not going to get close enough to find out. Sure, this is one of those creatures that seemed so lovable as portrayed in Walt Disney's *Fantasia*, decked out in tutus and gavotting sweetly to the Waltz of the Flowers.

However, you do not want to cruise up and check those tonsils by applying a tongue suppressor and coaxing «Say aaahhh, please.» No. When a hippo snorts, «Aaahhh!» as this one is doing without any prompting, it shivers the timbers of any foolishly really-up-close boat, and threatens passengers' eardrums, at the least.

These are bad-tempered water dudes, says the skipper of our 7.5 metre craft that seems so frail in the presence of a Zambezi River hippopotamus.

«This is close enough», says Larry Munkombwe at the wheel of the flat-bottomed power boat that is idling about 30 metres from the protruding grey head and shoulders of a resident monster. «They are very dangerous. Hippos are responsible for several thousand human deaths a year in Zimbabwe.»

«They can't swim, you know.» I didn't know.

But they run on the bottom, and they run very fast. This one has popped up for air. But they can stay under for quite a time, and they like to tip over boats. They account for a lot of drownings. He could overturn us, so we'll keep our distance.»

Hippos are vegetarians with no taste for human beings, Mukombwe says. «But they can mangle and stomp you to death, knock you around badly if they catch somebody on land. They look ponderous, but are deceptively fast for a short distance.»

Beside us four hooded eyes appear like floating Ping-Pong balls, behind two snouts. A pair of crocodiles are watching us watching the hippo. They would love the hippo to play bump-the-boat and force a turnover. Crocs aren't squeamish or fussy about people flesh...

Power Translator Proûs Output:

Les acariâtres tyrans de Zimbabwe

Assurément ils peuvent déchirer et marcher d'un pas lourd vous à mort. Mais le reste a assuré: Les étant végétariens, les hippopotames n'ont aucun goût pour les êtres humains.

Est ce grand a marmonné, larynxed bruyant la rose des amygdales d'hippopotame? Ou est-ce que c'est seulement le soleil du cadre qui reflète fermé l'eau dans son visage? Nous n'allons pas chercher fermez assez pour trouver. Assurément, c'est une de ces créatures qui ont paru si sympathiques comme peint dans la Fantaisie de Walt Disney, ornée dans les tutus et les gavotting à la Valse des Fleurs doucement.

Cependant, vous ne voulez pas croiser au-dessus et vérifier ces amygdales en appliquer un antiparasite de la langue et cajolant « Dites aaahhh, s'il vous plaît ». Non. Quand un hippopotame renifle, Aaahhh «! comme celui-ci fait sans en inciter, il frissonne les bois de construction d'en bateau vraiment en haut - fin sottement, et menace les tympanes de passagers, au plus petit.

Ce sont d'acariâtres mecs de l'eau, dit le capitaine de notre habileté de 7.5 mètres qui paraît si fragile dans la présence d'un Zambezi Rivière hippopotame.

«C'est fin assez », dit Larry Munkombwe à la roue du bateau du pouvoir à fond plat qui met au chômage approximativement 30 mètres de la sortant tête grise et épaules d'un monstre résidant. « Ils sont très dangereux. Les hippopotames sont responsable pour plusieurs mille morts humaines une année au Zimbabwe ».

«They can't swim, you know.» Je ne savais pas.

Mais ils courent sur le fond, et ils courent très vite. Celui-ci a crevé au-dessus pour l'air. Mais ils peuvent rester sous pour un vrai temps, et ils aiment renverser des bateaux. Ils expliquent beaucoup de drownings. Il pourrait nous renverser, donc nous garderons notre distance ».

Les hippopotames sont des végétariens sans goût pour les êtres humains, says' Mukombwe. «Mais ils peuvent déchirer et marcher d'un pas lourd vous à mort, vous frappe autour mal si ils s'accrochent quelqu'un sur terre. Ils semblent massif, mais est fausement rapide pour une courte distance ».

À côté de nous quatre yeux à capuche paraissent aimez des balles du Ping-pong flottantes, derrière deux museaux. Une paire de crocodiles nous regarde regarder l'hippopotame. Ils aimeraient l'hippopotame jouer la bosse le bateau et forcer un chausson. Crocs ne sont pas délicats ou méticuleux au sujet de gens écharnez...

Student sample of post-editing with subsequent categorization of errors:

Le mauvais caractère des brutes **structure**: zimbabwéennes

Bien sûr, ils peuvent vous mutiler et vous broyer avec **anaphora**:ses pas lourds jusqu'à la mort. Mais soyez'**wordchoice**:assuré: en tant que végétariens, les hippopotames **wordchoice**: n'ont pas de goût pour les êtres humains.

Sont-elles

agreement: rose, les amygdales d'un hippopotame à la grosse gueule démesurée et au larynx **wordchoice**: fort? Ou peut-être n'est-ce tout simplement que l'effet du soleil couchant qui se reflète **prepverb**:à travers l=eau sur son **wordchoice**:visage?

Nous ne nous **wordchoice**:rapprochons pas assez **literal**:proche pour découvrir la vérité. Évidemment, voici une de ces créatures qui semblaient si adorables d'après l'impression donnée dans "Fantasia" de Walt Disney, animaux décorés **prepverb**:en tutus et **structure**:gambadant **literal**:à la "Valse des Fleurs" une gavotte d'un pas léger.

Cependant, on ne veut pas s'approcher d'eux et leur examiner ces amygdales en appliquant **wordchoice**:une spatule pour déprimer la langue, **wordchoice**:tandis qu'on dit d'un ton **mistrans**:câlin: * *Tirez la langue, s'il vous plaît.* Non. Quand un hippopotame s'ébroue, * *Aaahhh!* ainsi que celui-ci le fait de son propre chef, cela **structure**:tremble **preverb**:les émotions de toute embarcation qui est assez bête pour **wordchoice**:se rapprocher si **literal**:proche et cela met **wordchoice**:au moins en danger les tympans des passagers. Ce sont des phénomènes aquatiques **prepnoun**:de mauvais caractère, dit le capitaine de notre **wordchoice**:navire de 7,5 mètres qui paraît tellement fragile en présence d'un hippopotame du Fleuve Zambèze.

* *On s'est assez approché,* +dit Larry Munkombwe au gouvernail du hors-bord à fond plat qui tourne au ralenti à une distance d'à peu près 30 mètres **mistrans**:loin de la tête grise **spelling**:emergeante et des épaules d'un monstre **spelling**:résidant. * *Ils sont très dangereux. On peut leur attribuer des milliers de morts humaines par an au Zimbabwe.* +

* *Vous savez qu'ils ne savent pas nager.* Je ne le savais pas.

* *Mais ils courent au fond et ils courent très rapidement. Celui-ci est venu à la surface pour **wordchoice**:prendre l'air. Ils sont capables de rester sous l'eau pendant assez longtemps et ils aiment **structure**:basculer les bateaux. Ils sont responsables d'un grand nombre de noyades. Il peut nous faire chavirer, donc on va garder **anaphora**:ses distances.*

* *Les hippopotames sont des végétariens **wordchoice**:sans goût pour les êtres humains,* explique Mukombe. * *Mais ils peuvent tuer quelqu'un en le mutilant et en le broyant avec **anaphora**:ses **wordchoice**:pas lourds, et aussi en le bousculant **wordchoice**:grièvement s'ils surprennent quelqu'un **preposition**:à terre. Ils ont l'air balourd, mais **prepnoun**:pendant **agreement**:un faible distance, ils sont beaucoup plus rapides qu'on ne l'imagine.*

À côté de nous, quatre yeux mi-clos apparaissent en arrière de deux **mistrans**:grogniers, comme des balles de ping-pong flottantes. Une paire de crocodiles nous observe regarder l'hippopotame. Ils aimeraient bien que l'hippopotame **wordchoice**:frappe notre bateau afin de nous faire chavirer. Les crocodiles ne sont pas **literal**:tatillons ni dégoûtés par la chair humaine.

Survey on the post-editing experience:

1. Encerchez la lettre de votre choix:
Cet exercice a été
 - a) extrêmement utile
 - b) assez utile
 - c) un peu utile
 - d) inutile
 - e) une grosse perte de temps
2. Dressez une liste des points forts et des points faibles de l'exercice, du point de vue de votre apprentissage.
3. Quelles recommandations feriez-vous pour améliorer l'exercice de postédition?

Nous vous remercions vivement de votre collaboration.