Post-editing a chapter of a specialized textbook into 7 languages: importance of terminological proximity with English for productivity

Ritesh Shah[1,2*], Christian Boitet[1†], Pushpak Bhattacharyya[2‡],
Mithun Padmakumar[1], Leonardo Zilio[1], Ruslan Kalitvianski[1],
Mohammad Nasiruddin[1], Mutsuko Tomokiyo[1], and Sandra Castellanos Páez[1]

[1]GETALP-LIG, Université Grenoble-Alpes, Grenoble, France
[2]CFILT, Indian Institute of Technology, Mumbai, India

Abstract

Access to textbooks in one's own language, in parallel with the original version in the instructional language, is known to be quite helpful for foreign students studying abroad. Cooperative post-editing (PE) of specialized textbook pretranslations by the foreign students themselves is a good way to produce the "target" versions, if the students find it rewarding, and not too time-consuming, that is, no longer than about 15-20 minutes per standard page (of 1400 characters or 250 words). In the experiment reported here, PE has been performed on a whole chapter of 4420 words (in English), or about 18 standard pages, into 7 languages (Portuguese, Japanese, Russian, Spanish, Bengali, Hindi, Marathi), native tongues of the participants. Average PE time has been measured, and when possible correlated with primary PE time (the time spent in editing a MT pre-translation in the PE text area). When terms are cognates of English terms (as in French, Spanish, Portuguese, and even Russian or Japanese), because neologisms are directly borrowed from English, or built using similar roots (often Latin or Greek) and similar word formation mechanisms (composition, affixation of special prefixes and suffixes), target terms can be "guessed" and PE time is in the order of 15 minutes per page, even if the target language is considered "distant" from English. On the other hand, PE times increase by a factor of 3 to 5 when the target language is terminologically distant from English. We found that is the case of Hindi, Bengali and Marathi, and probably of all Indic languages. Previous experiments seem to have missed that important point, because they were performed on too short texts (often, only a few paragraphs), and on "easy" pairs like English-French. A consequence is that, for terminologically distant language pairs, one should begin by separately collecting, or if necessary coining, the terms in the target languages.

1 Introduction

We are interested by using existing Machine Translation (MT) systems in the situations where their output does not (and often cannot, because, in its general form, MT is an "AI-complete" problem) provide "good enough" results. Post-editing MT results has been a professional activity in Japan since about 1985 (Nagao et al., 1985), and professional translators have begun to adopt that approach in other countries. Speaking of "translation accelerators", see for example the systems deployed at WIPO[1] and UN[2] (Pouliquen et al., 2013; Pouliquen and Mazenc, 2011a; Pouliquen and Mazenc, 2011b).

There are other situations in which MT outputs could be brought to a quality "good enough" for goals requiring a high level of precision and reliability. One of them is making pedagogical material accessible to foreign students in their own language. Cooperative post-editing (PE) of free MT pre-translations by the foreign students themselves is a good way to produce the "target" versions, if the students find it rewarding, and not too time-consuming., that is, if PE takes them no longer than about 15-20 minutes per standard page (of 1400 characters or 250 words).

*Ritesh.Shah@imag.fr
†Christian.Boitet@imag.fr
‡pb@cse.iitb.ac.in

[1]World Intellectual Property Organization
[2]The United Nations

Having validated that idea since 2-3 years on the French-Chinese (fr-zh) language pair, with the participation of 6 Chinese students in an experiment concerning mainly informatics and mathematics, we wanted to investigate the feasibility of using that approach for other language pairs and other scientific domains. Note that, in that particular experiment, students had studied informatics at home before going abroad, knew most of the Chinese terminology, and had access to online French-English and English-Chinese lexicons. A natural question arises: is this approach usable for all situations $<$ domain, instructional language, native tongue $>$? By 'native tongue', we understand not only the language spoken at home, but the language or languages of education, like Hindi and English in India.

In the fall of 2013, we had the opportunity to start an experiment involving a group of researchers, mostly PhD students, of 8 different native languages[3]: Portuguese (pt), Japanese (ja), Russian (ru), Spanish (es), Bengali (bn), Hindi (hi), Marathi (mr), and Malayalam (ml).

The Malayalam native speaker, PhD student of biological engineering, selected the text to be translated in his speciality, and organized the experiment, which lasted several months as we wanted it to be performed in realistic conditions, that is, with post-editors contributing occasionally. He could not do PE into Malayalam as there was no available en-ml MT system at the time, but he was initially expected to help with terminology in Hindi (his second language of education, English being his main one since college).

The text he selected was chapter 21 of the BEMbook (Malmivuo and Plonsey, 1995), a textbook in English on bioelectromagnetism, of about 700 pages, that is available freely online. This chapter contains 4420 words (in English) or about 18 standard pages.

The rest of this paper is organized as follows. We begin with more details on our motivations and discuss some related work on post-editing (PE). We then present the experimental setup and the use of the iMAG/SECTra tool along with the PE statistics. The next section gives quantitative and qualitative results, and analyzes why PE time, typically between 15 and 30 mn/page, was considerably longer for the 3 Indian languages (bn, hi, mr).

In short, the reason reported by the participants (post-editors) is that they ignored many existing technical terms *(latent terminology)*, or that, even worse, but a frequent case, there were no equivalent terms in their language *(absent terminology)*, in which case it was necessary to coin them after a long and fruitless search.

A deeper reason is probably that these languages are "terminologically distant" from English, that is, neologisms are not "cognates" of English neologisms: most of the time, they are not directly borrowed from English, and they are not built using similar roots (often Latin or Greek) and similar word formation mechanisms (composition, affixation of special prefixes and suffixes) as in English. Rather, they use roots coming from the Sanskrit or the Arabo-persian lexical store.

2 Motivations and objectives

2.1 Need for enabling access to pedagogical documents by foreign students

The needs for enabling access to pedagogical documents by foreign students in their native tongues has been long recognized. In Europe, the Bologna EU research project (2010-2013) attacked the task with the goal to make all syllabi of the European universities involved in the Erasmus project available in English and Chinese (Pietrzak et al., 2013; Depraetere et al., 2011; Depraetere and Van de Walle, 2012; Van de Walle et al., 2012b; Van de Walle et al., 2012a).

That was meant as a first step toward making course material also available, in those and more languages. The envisaged method was to build high-quality (HQ) specialized Statistical Machine Translation (SMT) systems to translate the corresponding web (*HTML*) pages and *PDF* files. Post-editing was seen only as a complement, to be added later if output quality did not meet expectations.

The choice of these 2 target languages was due to the large proportion of Chinese students in Europe, English being a natural "first step" as almost all European students studying within Erasmus must have quite good TOEIC[4], TOEFL[5] or IELTS[6] grades.

Like the proponents of the Bologna project, we think the real need is that foreign students get ac-

[3] Language code ISO639-2: https://www.loc.gov/standards/iso639-2/php/code_list.php

[4] Test of English for International Communication

[5] Test of English as a Foreign Language

[6] International English Language Testing System

cess not only to courses and tutorials, but also to notes from fellow students, in their respective languages, and possibly in parallel with the original (Kalitvianski et al., 2012).

Simultaneously, they should also be invited to participate and improve the translations as time progresses. A subgoal is also that they learn better the language of their host university.

2.2 Obstacles

The Bologna project started well, with clearly defined goals, and produced the beginning of a web service. However, very few syllabi were accessible, and the quality was not comparable with that of GT (Google Translate), Systran or Reverso. As no collaborative PE framework was included in the design, it could not be added afterwards and also could not produce a long-lasting service. This is no exception: less than 18 months after the end of the project, the Bologna web service was discontinued.

With our approach, immediate access in all languages for which MT is available is provided. But how to guarantee that a "good enough" quality level will be attained, and when?

A first answer is that students interested in accessing the content of these web pages and documents are usually only interested in some portion of them, and will "contribute" some corrections to improve segments they have not well understood. An important idea is that they will not do it only for themselves, to find "their" version when coming back later to the web page or document, but also for their fellows in the "virtual community" of the internauts sharing their language and wanting to access the same content.

A second answer is that "good enough" can be defined by some "self-scoring" mechanism. In SECtra/iMAG, each segment has a *reliability level*[7] and a *quality score* between 0 and 20[8].While the reliability level is fixed by the tool, the quality score can be modified by the post-editor (initially, it is that defined in his profile) or by any reader.

We can then say that the quality of the PE of a segment is deemed to be "good enough" if its quality score is higher or equal to 12/20.

2.3 Hypotheses and motivations behind the experiment

Our first hypothesis is that the objective is reachable only if PE is put at the center of the approach, and if foreign students that can benefit from it, do it themselves in a voluntary fashion.

Our second hypothesis is that we need to consider many more target languages than English and Chinese, including languages that are terminologically less equipped, for instance Arabic and South and South-East languages. Having the knowledge of an experiment (Wang and Boitet, 2013) in fr-zh for 1 year (in which the authors did not take part), we wanted to investigate the possiblity to do it for many languages, and to isolate the positive and negative factors, according to the target languages and to the profiles of the contributors.

3 Experiment

3.1 Setting

We considered the following constraints:

- to translate into as many target languages as possible, some of them being "distant" from English,
- to have as participants mostly PhD students, and all native speakers of the target languages,
- to tackle a text of significant length, in a representative document,
- to use available and free online MT systems.

3.1.1 Participants

The first thing was to assemble a team of volunteers, native speakers of several target languages. The candidate languages (native tongues of members of our lab) were Arabic, Bengali, Bulgarian, Chinese, French, German, Gujarati, Hindi, Japanese, Malayalam, Marathi, Portuguese, Russian, Somali, Spanish, Ukrainian, and Vietnamese. However, participants were available only for 8 languages (Portuguese, Japanese, Russian, Spanish, Bengali, Hindi, Marathi, Malayalam). Considering this set of target languages, and direct MT

[7]* for dictionary-based translation, ** for MT output, *** for PE by a blingual contributor, *** for PE by a professional translatior, and ***** for PE by a translator "certified" by the producer of the content.

[8]10: pass, 12: good enough, 14: good, 16: very good, 18: exceptional, 20: perfect. 8-9: not satisfied with something in the PE. 6-7: sure to have produced a bad translation! 4-5: the PE corresponds to a text differing from that of the source segment. That happens when a senetnce has been erroneoulsy split into 2 segments and the order of words is different in the 2 languages. 2: the source segment was already in the target language.

systems availability, English was the only choice for the source language. Unfortunately, at the time there was no MT system into Malayalam, so we had to skip the en-ml pair.

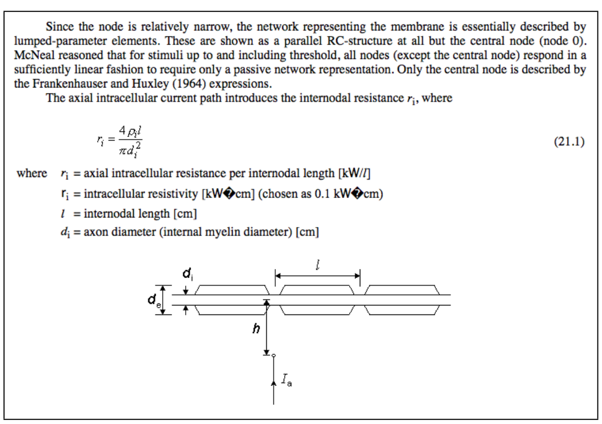
3.1.2 Choice of text

We had set up in 2011 an iMAG for the BEM-book(Malmivuo and Plonsey, 1995), for the benefit of a few French students in 2nd year of biology. Although they had bought the hard copy and knew English quite well (B2 level), they said it was quite helpful to access some difficult passages in French. In this case, MT outputs from GT were used, and the students post-edited 15% to 20% of the segments (sentences or titles) they accessed, but did not touch Chapter 21.

As said above, we selected chapter 21 because it was the most interesting for the PhD student in biology (native speaker of Malayalam).

This chapter has 374 segments, 4420 words and an average segment length of 22.22 words. It includes 10 diagrams, 13 equations, 11 section titles and sub-titles, 15 tables, 660 hapaxes (single-occurence words, and 6 scientific definitions spanning 62 segments.

Figure 1 below gives a screenshot of a typical passage from Chapter 21.



Since the node is relatively narrow, the network representing the membrane is essentially described by lumped-parameter elements. These are shown as a parallel RC-structure at all but the central node (node 0). McNeal reasoned that for stimuli up to and including threshold, all nodes (except the central node) respond in a sufficiently linear fashion to require only a passive network representation. Only the central node is described by the Frankenhauser and Huxley (1964) expressions.

The axial intracellular current path introduces the internodal resistance r_i, where

$$r_i = \frac{4\rho_i l}{\pi d_i^2} \qquad (21.1)$$

where r_i = axial intracellular resistance per internodal length [kW/l]
r_i = intracellular resistivity [kW�cm] (chosen as 0.1 kW�cm)
l = internodal length [cm]
d_i = axon diameter (internal myelin diameter) [cm]

Figure 1: Passage of BEMbook Chapter 21, with 5 sentences. The 5th contains in-line mathematical symbols, not to be translated but kept as is, a numbered stand-alone mathematical equation. The last part is a diagram.

Segmentation is often a problem. For example, in Figure 2 we show 3 successive segments produced by Google's segmenter for the 5th sentence of that screenshot. Actually, the linguistic "segment", corresponding to a sentence, is much longer: the beginning of the sentence, then an in-line mathematical relation that can function as a verbal phrase or as a proper noun, and the end of the sentence, made of a "where" dominating 3 bullet items, each of them giving the definition of some symbol.

The axial intracellular current path introduces the internodal resistance r_i, where

---

$$r_i = \frac{4\rho_i l}{\pi d_i^2}$$

---

where ra = axial intracellular resistance per internodal length [kW/l]; ri = intracellular resistivity [kWcm] (chosen as 0.1 kWcm); 1 = internodal length [cm]; di = axone diameter(internal myelin diameter) [cm]

Figure 2. Segments produced by Google

3.1.3 Post-editing times

The time we want to measure is Tpe_{tot}, the *total PE time*, expressed in mn/p (minutes per page). That time can be split in two parts, Tpe_1 and Tpe_2, such that $Tpe_{tot} = Tpe_1 + Tpe_2$:

1. Tpe_1, the *primary PE time*, is the time taken to perform editing operations in SECTra PE text area, or in the iMAG palette text area.

2. Tpe_2, the *secondary PE time*, is the time taken to perform other activities, essentially looking for lexical equivalents.

For a segment seg, $Tpe_1(seg)$ is measured by SECTra when PE of seg is done in SECTra (advanced mode). It is not yet measured by the iMAG. An ongoing research shows however that $Tpe_1(seg)$ can be estimated from the "mixed PE distance" $\Delta_\alpha(pt, pe)$[9] between the pre-translation pt and its post-edition, pe.

[9] For two strings A and B, the distance in characters, $\Delta_c(A, B)$, is the usual Levenshtein distance, with all weights set to 1 (for inserting, deletion and exchange). If the words of A and B are $a_1 \ldots a_m$ and $b_1 \ldots b_n$, the distance in words, $\Delta_w(A, B)$, is again the edit distance, but this time considering the set of words in A and B to the a new "alphabet", and the weights of exchanges, insertions and deletions are the character distances: $EXC(u, v) = \Delta_c(u, v)$, $INS(u) = DEL(u) = \Delta_c(\varepsilon, u)$. Finally, $\Delta_\alpha(pt, pe) = \alpha\Delta_c(pt, pe) + (1 - \alpha)\Delta_w(pt, pe)$ with $\alpha \in [0, 1]$.

In this experiment, the post-editors have been asked to time their total times globally, that is, to record every time they did some PE, typically for less than 1 hour, the list of segments post-edited, and the total elapsed time. Using a proportional rule, we could then estimate $Tpe_{tot}(seg)$ for each segment seg.

3.2 Environment

We used the iMAG/SECTra tool as it was, without any special addition. PE is possible in both direct mode (on the web page) and advanced mode (on SECTra table-like PE interface). Figure 3 shows the direct interface (a palette on the web page, much like the Google Translate palette) and Figure 4 the advanced interface. Participants were asked to measure their overall time, and to use the interface to "self-score". As said above, PE time (only Tpe_1) is measured for each segment by the system, but only in the dedicated interface.

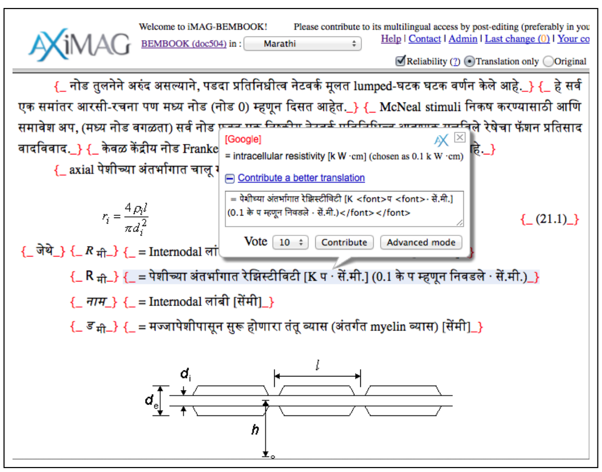


Figure 3. Post-editing on the web page. Reliability brackets appear around each segment: red for MT result, green for PE by a connected participant, orange for an anonymous contributor.

3.3 Unfolding

The experiment spanned about 9 months (1/10/2013-15/6/2014), because participants worked incrementally for short periods of time, which is what is expected in "real life".

3.3.1 Organization

A strong organization was not needed, but something of that kind emerged:

- There was a global "animation" by Partic_ml[10], who could not PE himself due to the lack of MT into Malayalam.
- Joint work was done on Hindi and Marathi. It was coordinated and done 80% by Partic_hi_mr with the cooperation of Partic_ml.
- Partic_bn and other participants worked alone.

Figure 4. Table-like SECTra "advanced" PE interface. In this interface, the post-editor can choose from which MT result to start, in case more than one is available, and give himself or herself a score differing from the default score associated to his/her profile. The "reliability level" (between 3 stars and 5 stars) cannot be modified, but the "quality score" can. If a segment has been post-edited, the time taken in the PE window (Tpe_1) appears (in blue).

3.3.2 Particular points

Partic_sp had deadlines for exams and reports and could only post-edit about 60% into Spanish.

The terminological help initially expected from Partic_ml to find or create terms in Hindi did not materialize, because Partic_ml was educated in Malayalam and English, and learned Hindi only upto something like B2 level, with no introduction to technical terms, that are no more similar between Dravidian and Indo-Aryan languages than between English and Indo-Aryan languages.

By contrast, new terms can usually be coined quite easily from English (or French) terms that have a Latin or Greek etymology, into Romance languages as well as into Slavic languages. For Japanese (and French, for that matter), new terms

[10] Participant for Malayalam language.

are often initially directly borrowed from English, to be later (not always) superseded by terms coined using the lexical store and word formation methods of the language.

Some automatic term-coining from English into Bulgarian was already reported in (Nikolova and Nenova, 1982), and in this case it worked also into Russian.

4 Evaluation and discussion

4.1 Quantitative results

Language	Segments	Words	Words/segment (average)	Standard pages (250 words)	Time/std_page (seconds)	Time/std_page (minutes)
English (source)	*374*	*4420*	*11.82*	*17.7*	-	-
Portuguese	374	6232	16.66	24.9	926.2	15.5
Japanese	374	5000	13.37	20.0	937.2	15.6
Russian	374	4800	12.83	19.2	1037.4	17.2
Spanish	164	2583	15.75	10.3	1940.0	32.2
Bengali	365	4472	12.25	17.9	4666.5	**77.7**
Hindi	374	4547	12.16	18.2	6549.7	**109.2**
Marathi	374	4672	12.49	18.7	7087.6	**118.1**

Table 1. Time (Tpe_{tot}) taken for post-editing Chapter 21 of the BEMbook into 7 languages

From Table 1, we observe that PE was complete or nearly complete for all languages, save Spanish.

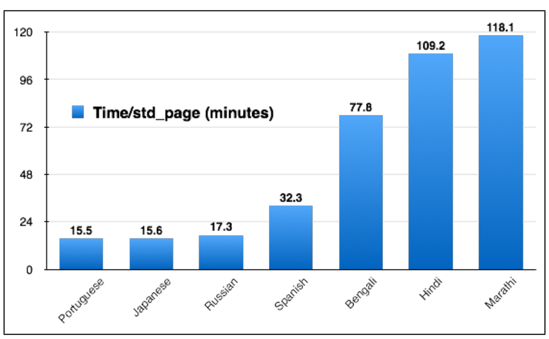


Figure 5: Plot of total post-editing times (in minutes per standard page)

4.2 Qualitative results

The final outputs obtained are all of the *expected quality level*. That means, as said above, that the quality score of each segment is higher or equal to 12/20. When the posteditors set that score, they do it by answering the question: "how good is this translation now for my usage, and presumably for the usage of future readers in my language?"

That means that, in this case, there is not much more one can do: producers are consumers of their own productions, hence they are in the bst position to judge them.

4.3 Impact of lack of terminology or terminological competence

As can be seen on Figure 5, the PE time per standard page is considerably higher for the Indic languages than for the others. Participants for the three Indic languages indicated that this increase was essentially due to lack of knowledge of specific terms, or to their absence in the language, meaning neologisms had to be created.

On the other hand, when terms are cognates of English terms (as in French, Spanish, Portuguese, and even Russian or Japanese), because neologisms are directly borrowed from English, or built using similar roots (often Latin or Greek) and similar word formation mechanisms (composition, affixation of special prefixes and suffixes), target terms can be "guessed" and PE time is in the order of 15 minutes per page, even if the target language is considered "distant" from English.

In fact, posteditors in the other languages (non Indic, here) had no problem to "deduce" terms in their languages from the English terms, by applying some simple spelling and transliteration changes.

4.3.1 Difficulty in finding an existing term in the 3 Indic languages considered

There were nearly 660 hapaxes in the post-edited chapter, about 300 of them being technical terms. Almost all these terms were understood by the Indian post-editors, but they did not know their equivalent in their languages. They of course began by looking for them in terminological sources, such as [http://cstt.nic.in] for instance, for English-Hindi pair. A successful search typically took 3 to 5 minutes, but an unsuccessful search (in several sources) took 4 times as much.

As a consequence, PE times increased by a factor of 3 to 5 when the target language was Hindi, Bengali and Marathi. The same thing would most probably happen with all other Indic languages.

4.3.2 How did we coin a non-existing term in Hindi or Marathi?

In case a term was not found, it had to be assembled (in case of a compound term in English), or transliterated. We give examples below. Assembling a term took typically 20 to 30 minutes, because there are usually several plausible choices, that had to be discussed between participants and sometimes with colleagues over the web.

Some examples of ”assembled” terms are shown in Figure 6.

hyperpolarization अतिध्रुवण
stimulation उद् दीपन
electrode characteristics इलैक्ट्रोड अभिलक्षण (found together)
electrode current इलैक्ट्रोड धारा (found together)
electrode dissolution इलैक्ट्रोड विलयन (found separately)
capacitance धारिता
nodal membrane निस्पंद/निस्पंदीय/नोड/नोडल झिल्ली (found separately with variations in dictionary itself)
threshold level देहली स्तर

Figure 6: Examples of ”assembled” technical terms in Hindi

4.3.3 Why this factor is predominant but has not been considered in most previous studies?

Going from 15 mn/page to 1h15 mn/page makes of course a huge difference, and it is mainly attributable to the ”terminological problem”. However, previous studies such as (Green et al., 2013) did not mention it.

There may be several reasons for that. First, these experiments have usually been done on a few paragraphs only. Second, the material was selected in domains that were quite familiar to the participants, and not too specialized, so that post-editors always knew the lexical equivalents in their languages.

In fact, the *terminological problem* appears only (with that severity) when

- the 2 languages are terminologically distant (as defined above),
- the terminology in the target language is largely latent (hence, ignored by the post-editors), or worse absent.

5 Conclusion and perspectives

In this paper, we have presented an experiment in post-editing part of the online version of a textbook in English on bioelectromagnetism, the BEMbook, by a group 6 PhD students and a senior researcher, of 7 different native tongues (Portuguese, Japanese, Russian, Spanish, Bengali, Hindi, Marathi), using the iMAG/SECtra tool.

The goal was to investigate the feasibility of using that approach to make pedagogical material available to foreign students or to students studying abroad in their native tongues, with a PE time that was hoped to remain aound 15 mn/page.

The result is that it is quite possible, but that the post-editing time, typically between 15 and 30 mn/page, can considerably increase if many technical terms are ignored by the post-editors, or even worse when they are absent, which was the case for the 3 Indic languagues (bn, hi, mr). More precisely, about 1 hour per page can be added in case the lack (or ignorance) of terminology is significant.

A consequence is that, for terminologically distant language pairs, one should begin by collecting, or if necessary coining, the terms in the target languages, before asking foreign students to use our web service to access pedagogical material in their native languages, improving the quality of the text as a byproduct of their learning activity, spending a reasonable time in cooperative PE.

That is actually what is done in all professional or semi-professional translation and localization projects, for example in the localization projects integrated into large open source projects such as Mozilla.

In the context of ”contributive PE”, something more will be needed, so that the initially collected or coined bilingual terminology does not become obsolete with time, but ”lives” with the PE activity using it.

What we plan, then, is to set up a contributive web service containing a multilingual lexical database with a simple ”sharing-oriented” structure, like PIVAX (H.T. Nguyen 2008), and an instance of SepT[11], a kind of lexical network structure for collaboratively proposing terms as a kind of parallel ”preterminological” activity.

Acknowledgements

We are grateful to UJF (Université Joseph Fourier), IRD (Institut de Recherche pour le Développement) and LIG (Laboratoire Informatique de Grenoble) for supporting this experiment.

References

Mohammad Daoud. 2010. *Usage of non-conventional resources and contributive methods to bridge the terminological gap between languages by developing multilingual ”preterminologies”*. Ph.D. thesis, University of Grenoble, France.

Heidi Depraetere and Joeri Van de Walle. 2012. Bologna Translation Service: An enabler for international profiling and student mobility. In *”Proceedings of the 6th International Technology, Edu-*

[11] System for Eliciting pre-Terminology (Daoud, 2010).

cation and Development Conference", pages 5907–5912, Valencia, March. IATED.

Heidi Depraetere, Joachim Van den Bogaert, and Joeri Van de Walle. 2011. Bologna Translation Service: Online translation of course syllabi and study programmes in English. In *"Proceedings of the 15th Conference of the European Association for Machine Translation"*, pages 29–34, Leuven.

Spence Green, Jeffrey Heer, and Christopher D Manning. 2013. The efficacy of human post-editing for language translation. In *"Proceedings of the SIGCHI conference on Human factors in Computing systems"*, pages 439–448. ACM.

Ruslan Kalitvianski, Christian Boitet, and Valérie Bellynck. 2012. Collaborative computer-assisted translation applied to pedagogical documents and literary works. In *"Proceedings of the 24th International Conference on Computational Linguistics"*, pages 255–260, Mumbai. Citeseer.

Jaakko Malmivuo and Robert Plonsey. 1995. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press.

Makoto Nagao, Jun-ichi Tsujii, and Jun-ichi Nakamura. 1985. The Japanese Government Project for Machine Translation. *Computational Linguistics*, 11(2-3):91–110.

Bonka Nikolova and Irina Nenova. 1982. An Automated System for Term Services. In J Horeckp, editor, *"Proceedings of the 9th Conference on Computational Linguistics"*, pages 265–270, Prague. North-Holland.

J Pietrzak, A Jauregi, Joeri Van de Walle, and A Eriksson. 2013. Improving access to educational courses via automatic machine translation - New developments in post-editing. In *"Proceedings of the INTED 2013 Conference"*, pages 5521–5529. IATED.

Bruno Pouliquen and Christophe Mazenc. 2011a. Automatic translation tools at WIPO. *Aslib, Translating and the Computer*, 33:17–18.

Bruno Pouliquen and Christophe Mazenc. 2011b. COPPA, CLIR and TAPTA: Three tools to assist in overcoming the patent language barrier at WIPO. In *"Proceedings of the 13th Machine Translation Summit"*, pages 24–30.

Bruno Pouliquen, Cecilia Elizalde, Marcin Junczys-Dowmunt, Christophe Mazenc, and José García-Verdugo. 2013. Large-scale multiple language translation accelerator at the United Nations. In *"Proceedings of the 14th Machine Translation Summit"*, pages 345–352.

Joeri Van de Walle, Heidi Depraetere, and J Pietrzak. 2012a. Bologna Translation Service: High-quality automated translation of study programmes into English. In *"Proceedings of the EDULEARN 2012 Conference"*, pages 5831–5835. IATED.

Joeri Van de Walle, Heidi Depraetere, and J Pietrzak. 2012b. Bologna Translation Service: Making study programmes accessible throughout Europe by means of high-quality automated translation. In *"Proceedings of ICERI 2012 Conference"*, pages 3910–3918. IATED.

Lingxiao Wang and Christian Boitet. 2013. Online production of HQ parallel corpora and permanent task-based evaluation of multiple MT systems: both can be obtained through iMAGs with no added cost. In *"Proceedings of the 2nd Workshop on Post-Editing Technologies and Practice at MT Summit 2013"*, pages 103–110, Nice, September.