The International Journal for Translation & Interpreting Research
trans-int.org

Testing interaction with a mobile MT post-editing app

Olga Torres-Hostench
Universitat Autònoma de Barcelona, Spain
olga.torres.hostench@uab.cat

Joss Moorkens
Dublin City University, Ireland
joss.moorkens@dcu.ie

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

Joris Vreeke
Dublin City University, Ireland
joris.vreeke@adaptcentre.ie

DOI: 10.12807/ti.109202.2017.a09

Abstract: Kanjingo is a post-editing application for iOS devices developed at the ADAPT Centre (formerly CNGL) at Dublin City University (DCU). The first stage of user testing was conducted in 2014 (reported in O'Brien, Moorkens & Vreeke, 2014), and improvements were made based on the initial feedback. This abstract describes further exploratory testing based on the second iteration of the Kanjingo application. The new tests were designed with several aims: (1) testing Kanjingo for post-editing using the phone's keyboard (2) testing Kanjingo for post-editing with voice input; (3) testing Kanjingo for revision of post-edited texts; (4) testing Kanjingo general usability; and (5) testing Kanjingo interface design. This paper presents the results of the various tests, issues identified, and ideas for improvements. For example, the use of Kanjingo for post-editing with voice input, one of the most innovative forms of interaction with MT in the test, worked much better than participants expected, and this mode of input was preferred for translating from scratch when MT quality was very poor, whereas post-editing short words or phrases was found to be faster with the iPhone keyboard. In addition, we present some reflections on the strengths and weaknesses of the testing methods employed.

Keywords: machine translation, machine translation post-editing, mobile devices, Kanjingo, mobile app, translation

1. Introduction

Working with machine translation (MT) using a mobile MT post-editing app implies different levels of interaction. On the one hand, there is the interaction with the hardware, that is, a "mobile device". On the other hand, there is the

interaction with the software, in this case, the Kanjingo app. Both kinds of interaction are addressed in this paper. A priori, post-editing from a smartphone seems unreasonable. The space limitations of a mobile device with limited screen space and a small keyboard might be expected to make the post-editing experience deteriorate instead of improving it, taking into account the fact that post-editing in itself has often been considered frustrating (Guzmán, 2007; Specia, 2011; Koponen, 2012).

In 2014 a prototype of Kanjingo, a web-based post-editing application for smartphones, was developed by the ADAPT Centre (formerly CNGL) at Dublin City University (DCU). Usability for post-editing was tested and explained in a research paper (in O'Brien et al., 2014; Moorkens, O'Brien & Vreeke, 2016). A native iOS app with improved functionality was developed based on the initial feedback and a second series of tests were carried out in 2016, which will be described in this paper. However, before presenting the test results, a brief explanation of the app may be helpful in order to explain its functionality.

Figure 1. Kanjingo interface.

When Kanjingo is first accessed, the user selects a language pair. Then source segments are presented to the user. When a source segment is selected this opens the Editor view, in which the source segment is displayed at the top of the screen, with the corresponding machine translated segment below in a vertical tiled format, where each tile contains a word. The user may add empty tiles with a "+" symbol to insert new words, and delete tiles with the "-" symbol to delete existing words. Tiles can be dragged up and down to rearrange the sentence structure. When the user taps once on a word, an edit box appears, within which the user may edit the word, and tapping a second time enables the appearance of the smartphone keyboard. Once the user finishes post-editing, the Send button must be pressed to submit the sentence.

Besides the post-editing feature, there is also a translation feature consisting of an empty tile below a source segment where the user may translate the whole source segment using the phone keyboard. The Help menu may be accessed from the Editor view.

From the above description it can be deduced that Kanjingo is fairly easy to use, but is it usable for post-editing? According to Kukulska-Hulme & Traxler (2005, p. 45), "usability in its most basic definition is whether something can be used for its intended purpose". Kanjingo usability was first tested in 2014 (O'Brien et al., 2014) and some usability issues arose that would be addressed in the later iOS app development as described in (Moorkens et al., 2016). Some problems were related specifically to the HTML application design. For instance, users could lose unsubmitted work if they left the UI to check the Internet or dictionaries/glossaries. Insufficient help was mentioned. Users also complained that they had to manually add capitals at the beginning of a segment and to append a full stop at the end because the app did not provide automatic punctuation and capitalisation.

However, other problems were related specifically to the limitations of a handheld device. For instance, users highlighted sensitivity issues due to the limited display and said that more space should be provided on either side of the segment display for scrolling and requested a feature for grouping words for combined drag and drop functionality. They also mentioned input challenges such as speed of typing, which is so much slower than for a desktop scenario.

Taking into account the physical limitations of mobile devices, the authors considered it worthwhile investigating how they could be overcome in order to improve interaction with MT post-editing (MTPE). One way of doing this would be, for example, using voice input instead of the keyboard. Although voice recognition research is quite developed (Dymetman et al., 1994; Vidal et al., 2006; Dragsted et. al., 2011; Toselli et al., 2011; O'Brien, 2012), it has not been applied often in MT or PE research. One example is the SEECAT Project and the research carried out by Mesa-Lao (2014) at the University of Copenhagen. It would make sense then to explore the interaction of three areas: voice input, MT post-editing, and mobile. Thus, the usability testing carried out in 2016, and reported on here, was focused on two main aims. First, analysing the limitations of mobile devices for MT post-editing via three different tasks: (1) testing Kanjingo for post-editing using the phone's keyboard (2) testing Kanjingo for post-editing with voice input; (3) testing Kanjingo for revision of post-edited texts. Second, improving the app itself by (4) testing Kanjingo usability; and (5) testing the Kanjingo interface.

2. Method

The Kanjingo 2016 testing followed a qualitative methodology adopted from Krug (2006), a renowned web usability consultant. While for traditional usability testing "you need quantitative testing, with a large sample size, a clearly defined and rigorously followed test protocol and lots of data gathering and analysis", his approach "is a qualitative method whose purpose is to improve what you're building by identifying and fixing usability problems (…) The result is actionable insights, not proof" (Krug, 2006, p. 119). This method consists of testing an app with a few participants (five in our case) while selected observers (again, five in our case) watch the tests in real time from a room nearby. At the end of the testing day, observers and the testing

team meet to discuss preliminary results. Taking into account that the app is still under development and has not been launched to market yet, it was reasonable to take this qualitative approach in line with Nielsen, who stated that "the best results come from testing no more than five users and running as many small tests as you can afford" (2000). The tasks to be carried out in the testing were the following:

Task 1: Using Kanjingo to revise previously post-edited texts
As one of the drawbacks for post-editing using a smartphone is input limitation, as detected in the first Kanjingo user testing, a test was set up to use Kanjingo in such a way that typing was reduced to a minimum. The task consisted of reviewing segments that had been translated from English to Spanish in a real post-editing assignment carried out at Universitat Autònoma de Barcelona (UAB)[i]. MTPE carried out by participants was to be screen recorded, with gestures and audio comments also recorded with a webcam.

Task 2: Using Kanjingo to post-edit machine translation with voice input
This task consisted of post-editing segments from the Global Voices website translated from English into Spanish by the MTradumàtica MT engine (Martín-Mor, 2017).[ii] Participants were asked to post-edit the segments using voice input. They could decide whether they would prefer to delete the whole proposal and translate it from scratch using voice input or reuse as much of the MT as possible and just post-edit some Kanjingo tiles. PE, gestures, and audio comments were again to be recorded.

Task 3: Using Kanjingo to post-edit machine translation with smartphone keyboard
Selected segments from the Global Voice website (different from the ones used in Task 2) were translated from English into Spanish using the MTradumàtica MT engine. Participants were asked to post-edit the segments using the smartphone keyboard. Despite the fact that this is the most common use for Kanjingo, it was decided to do this in Task 3 as it is the task that requires most knowledge of the app and at that point participants would be quite familiar with it.

Five participants were selected for testing. Some days before the test, the participants filled in an online pre-test questionnaire on their personal and professional profile and their opinion of MT and post-editing. They were also asked about the professional tasks that they perform with their smartphone. From this pre-test questionnaire we extracted the following information: two were product managers in language companies, one was a university translation lecturer, one was a researcher in Translation Studies and one was a freelance translator; 2 men and 3 women; between 29 and 49 years old; 3 with previous experience in MTPE and 2 without previous experience in MTPE. The participants had in common that all of them were daily users of smartphones although none of them translated nor post-edited using the phone. Four out of five consider MT very useful and the fifth considers MT useful sometimes. Users with a positive attitude towards MT were chosen because they were expected to focus on the app usability and design rather than the quality of the MT. As the app is not an MT engine itself, it has no control over the quality of the MT output.

The app was installed on an iPhone 6, which has a screen size of 4.7 inches (12cm that contain 1334x750 pixels, displayed in 16:9 format). The

tests were carried out using three spaces in DCU in July 2016. The first space was a testing room, equipped with the iPhone with the Kanjingo app, a webcam fixed to a 'sled' to record the participants' hand movements and audio, phone screen mirroring software (Mirroring 360[iii]) and a computer with Camtasia screen recording software to record the webcam and the phone mirroring activity.

During testing, the participants came to the Testing Room, signed consent forms, and were handed task scenarios. The testing facilitator[iv] welcomed each participant and showed them the basic features of Kanjingo. Each participant was given a couple of sentences to post-edit as a warm-up exercise before beginning the test tasks. Once the participant understood how the app worked and could use it on his/her own, the first task was presented: a revision of post-edited sentences. Then the second task (MTPE using voice input) and finally the third task (MTPE using keyboard). Participants were asked to comment on their experiences during MTPE (using the Think Aloud method), and comments were recorded via the webcam. After the test, participants answered the following general questions on the app and its use:

1) What are your initial thoughts on the Kanjingo app?
2) How could it better support post-editing?
3) How does post-editing with the app compare to desktop post-editing?
4) What does the app do well?
5) What does the app do badly?
6) Did you find the app speedy or responsive?
7) How could the look of the interface be improved?
8) In what circumstances could you envisage yourself using the app?
9) In what circumstances would you use Kanjingo for reviewing?
10) Would you recommend Kanjingo to your colleagues and/or customers for post-editing tasks?
11) Which features would you like to improve in the application if you were to use it on a regular basis?

The second testing space was the Observation Room, from where the test was followed in real time by the Kanjingo architect and a group of selected observers with knowledge of MT who could provide valuable insight on the test via their own reflections.

Five observers participated in the test. Observer 1 has a background in Translation Studies, does research in MT and has experience in MTPE. Observer 2 has a background in Computer Science and does research on Machine Learning methods, Natural Language Processing and Computational Linguistics. Observer 3 has a background in languages and Specialised Translation and does research in Translation Studies. Observer 4 is a professional translator with experience in MTPE. Observer 5 is the app architect, a software engineer and web developer. The Observation Room was equipped with a 50" screen which mirrored the computer screen in the Testing Room, with accompanying audio via Skype. Observers were provided with instructions and specific questions on the three main usability issues identified in the app and the three main ideas for further developments, with some free space for comments. The third space was the Debriefing Room where the testing team and some observers met to discuss their notes immediately after the test. The purpose of the debriefing meeting was to obtain preliminary hot results on the testing day instead of waiting for the full report of results.

After the testing day the participants' recordings were transcribed and coded manually according to different tags, under the following main branches: usability (positive features, features to be improved, ideas for future developments); interface (positive features, features to be improved, ideas for future developments); preferences of use; use cases for the app; and emotions involved in the app experience. The main branches would divide into many different tags such as voice input, keyboard, productivity, learnability, mobility, etc. Then the number of occurrences for each tag was annotated together with the participant who made the comment. Occurrences were ordered from the most frequently mentioned to least frequent, so in the results report it was easy to identify the most relevant issues first.

3. Results

3.1 On mobile physical limitations: small keyboard and small screen

Due to space limitations, participants said that they preferred using Kanjingo for short sentences rather than long sentences, and would prefer to use Kanjingo for short assignments consisting of few words. If sentences were too long, it was not possible to see the whole text without scrolling down. Enabling the app for landscape mode was suggested. Participants also mentioned that they often forgot to look at the target sentences at the bottom of the screen and requested that the target text could be kept at the top of the screen, beside or below the source text. Regarding the small screen, one participant said that they preferred to post-edit using a traditional desktop computer, preferably using two large computer screens, rather than on the small phone screen.

There were comments about the small keyboard. One participant mentioned that the mode for inputting punctuation with the keyboard was annoying and made it difficult to check information online (terminology, etc.). Another mentioned that s/he would connect an external keyboard to the phone to improve the typing experience and would add a swiping feature in order to ease the typing process (this is possible by downloading a swipe keyboard from the iPhone app store).

While the idea of post-editing from anywhere is appealing for participants – several proposed use of Kanjingo on-the-go while travelling or commuting – ergonomics seems to be the biggest disadvantage for using Kanjingo for professional purposes. For this reason, participants mentioned that Kanjingo is suitable for not-for-profit projects such as 5min4good (2013), for fun, for crowdsourcing, and even for translation training. They also suggested using Kanjingo to post-edit something urgent or when a client is in a rush. However, they said that professionals who are post-editing would ideally work with two big screens to switch between the translation and documentation resources.

3.2 On overcoming physical limitations: voice input

Regarding voice input, it was mentioned that voice input "works ok", "is very good", and "is a plus", although it was also "cumbersome" for another user. The opinions on voice input were not unanimous. They seemed surprised by the high quality of automatic speech recognition (ASR). One participant was surprised that voice input utilised different Spanish accents properly. ASR also recognized orders such as *comma* and *full stop* and the whole dictation

could be added in one Kanjingo tile. The fact that lower case was written automatically in the middle of a sentence was seen as a positive feature.

In Task 2 participants were invited to post-edit five segments using voice input or keyboard, and they tried different combinations, which we will describe here in detail. From the 25 segments to post-edit (5 segments per participant), ten were post-edited using only the keyboard, as participants mentioned that changes were minor and it was easier to use the keyboard than voice input. Eight segments were post-edited using only voice input in some tiles. Two segments were deleted completely and translated with voice input from scratch (this option was chosen only to see how it worked, and participants said that they would use it only if MT output was very bad). For two other segments that were deleted and translated with voice input from scratch, some minor edits were made with the keyboard to amend the voice input. Finally, for three segments, voice input was used in some tiles followed by keyboard edits to amend the voice input.

Four out of five participants preferred typing instead of using voice input because they were more used to typing, but one of the participants, who was not so used to typing, said that voice input was easier to use than the keyboard. Participants said that voice input was good for re-translating long sentences with bad MT output, i.e., sentences with such low MT quality that they were not easy to post-edit, but it was not worthwhile using voice input to post-edit just one word. One participant recommended not using voice input in noisy places for obvious reasons.

As a future improvement it was suggested that there could be MT quality scores and thresholds next to the sentences so that bad quality sentences could be re-translated directly with voice input. Or even better, MT below a certain quality threshold might not even be shown to the user, as is done in certain CAT applications which integrate MT and TMs.

3.3 On reducing keyboard use: use Kanjingo for review

The testing team wanted to know whether Kanjingo was suitable for performing traditional linguistic review, which, theoretically, would imply less editing if the MT target text had good quality in general. While one participant said it could be convenient although not for long sentences, another participant said that tiles might not be the optimal solution for review and maybe it would be better to have the whole text in a window (instead of sentences or tiles) and edit directly only what you need to change if changes are minimal and only a few sentences would need changes. The rest of the participants felt satisfied with the tiles.

Participants suggested using Kanjingo with short texts, standalone sentences, and informal types of translation. It could be used for review with the caveat that it may be problematic to review a long text with many internal references and target segments cannot be changed once they have been submitted. In its current state Kanjingo only allows editing sentence by sentence and it is not possible to review the whole text. In summary, review could be a suitable task because the keyboard is not used as much. However, due to the segment orientation of the app, internal references and text coherence would be challenging to deal with.

3.4 On Kanjingo usability

Krug (2006, p. 155) defines usability as the extent to which "a person of average ability and experience can figure out how to use the thing [i.e. "it's

learnable"] to accomplish something [effective] without it being more trouble than it's worth [efficient]."

Taking into account the above-mentioned criteria and the participants' feedback, we concluded in this round of testing that Kanjingo's general usability is good. Participants said the app was intuitive, easy to use, user-friendly and very handy.

As for the Kanjingo-specific features, the participants mentioned the tiles (adding, moving and deleting tiles) as a positive feature. Also mentioned positively was the fact that capitalisation changes for the first tile and the "suggested proposals" feature. As for responsiveness, it was mentioned that the app's productivity was ok because "it does not block", "it goes very well" and the "app is fast". As for learnability, participants agreed that the app was easy to learn, with a low learning curve and it is self-explanatory.

Among the usability issues to be improved were two issues mentioned by all participants, suggesting that they should be prioritised in the next build. The first suggestion was related to the Accept button to approve post-editing. Currently it is placed in a location where is it easy to forget to click on it. Some participants asked if it would be possible to accept a sentence by pressing anywhere on the screen, not on the Accept button, but this request might have a drawback because the user might touch the screen accidentally before finishing the sentence. Second, in its current state of development it is not possible to edit in any location within the tile. It is only possible to delete from the end of the tile. So, if a tile has many words, it is not possible to go directly to the first word and change it. Besides these two main improvements, there were some other suggestions. Participants would prefer to keep the post-edited sentences in the app and not send them right away, and to be able to see the whole post-edited text at the end. They would like to be able to review the post-edited sentences when required.

They found editing to be the least intuitive feature of the app and suggested that it needs improvement. They mentioned that a double click is needed to open a segment and it would be better to open it with just one click (this iteration of the app was developed prior to the introduction of 3d Touch in iOS9, permitting different screen responses based on finger pressure). Another participant mentioned that too many clicks are needed (to open a sentence, to delete letters in a tile when there are several words, etc.). It was also mentioned that the 'suggested proposals' feature adds an extra space at the end and that MT output should be recoverable after clicking on the Clean button.

Participants suggested ideas for future developments such as adding spellcheck, copy & paste features, a bin, numbering segments, a search tool for sentences and a 'Language resources' button to go directly to online dictionaries or databases directly from the app. It was also suggested to have a Pro version that would include a link to an MT engine so that anyone could control the whole workflow and post-edit their own texts. There could be a lighter version for post-editors who just receive texts to post-edit.

3.5 On Kanjingo design

Kanjingo design was responded to very positively by participants with words such as "clean", "big enough" "great resolution" and "nice". They also mentioned that there is no need for the "Show sentences" screen which appears after sending a segment and that it should be possible to go directly to the next sentence. Some ideas for future designs would be to add project

folders and a brief explanation on each project, and to add ticks for confirmed sentences.

3.6 Unexpected results

The testing team would like to add a comment on the subjective experience of participants. During the tests, the participants adopted a critical attitude towards Kanjingo because this is what they were asked to do. They did a good job of taking into account all suggestions above. However, it is worth mentioning the personal attitudes and emotions involved while "playing" with the tool. Expressions from the participants were "This is great fun", "This is fun", "It is fun at the end", "It's addictive this software", "It's a bit like a game", "Great. Very exciting!". The positive emotions and expressions by the participants suggest the possibility to explore further the gamification of the tool in order to make of Kanjingo an even more pleasant and playful app for non-for-profit post-editing. It is worth pointing out, at the same time, that previous exploration of this topic resulted in negative responses from the users, so it looks like the improvements made between the first and second testing phases improved the user experience.

3.7 Results from the observers

The observers took notes during the tests from the Observation Room. As they were not as focused on the app but on the user experience, they would take notes on what they considered the most important things from the user perspective. Usability issues highlighted by observers from the participants' comments made during the test were related, for instance, to visualization problems, ways to improve the speed of the MTPE process, voice recording and editing features which need improvement.

On the other hand, observers could reflect and connect what they were looking at with their own expertise, they added valuable comments for the testing team that were not mentioned by the participants. In the following we present a summary of the ideas provided by observers that were not included in participants' comments. Taking into account that the size of the screen seems to be a drawback, an observer suggested designing the app for mobile devices with bigger screens such as tablets (actually, this development is in progress now). Some ideas were suggested to improve existing features such as adding a double confirmation before sending the sentence, a button to check saved segments or some punctuation improvements. Other ideas were new features for the app such a progress button, an automatic conversion of measures, an intelligent system for tile segmentation, a MT quality estimation score. An observer suggested that Kanjingo would be best used for light post-editing (instead of full post-editing), and ideal text types would be short documents, business emails, brochures, short web pages (such as the "who we are" section) and restaurant menus.

3.8 Results from the debriefing meeting

The main aim of the debriefing meeting held by the testing team and some observers after the test was to determine the top three usability problems, the top three interests for the industry, and the top three further developments so that action could be taken immediately after the test (while detailed results were being processed). The conclusions of the meeting were:

Top three usability problems (detected by all participants): (1) The Accept button seems to be a problem (many users forget to click on it); (2) if you have several words in a single tile (because you post-edit one tile and

write several words in it or via voice input), you have to delete everything to change something at the beginning. It is not possible to place the cursor in the text; (3) not being able to see submitted segments (participants would prefer to save their own work after post-editing).

Top three interests for the industry: (1) Crowdsourcing; (2) there is potential for the gamification of post-editing ("it is addictive", "it is fun"); (3) there is potential for voluntary contributions in the user's spare time (participants do not really envisage that the app is appropriate for professional use).

Top three further developments: (1) Copy and paste words (to reuse proper names, etc.); (2) use swipe feature for the keyboard; (3) add MT confidence or quality score.

4. Discussion and conclusions

An MTPE app has to live with the constraints of mobile devices. As Krug (2009, p. 147) states, "serious usability problems are the result of a poor decision about a trade-off (…) Most of the challenges in creating good mobile usability boil down to making good trade-offs". In the use of Kanjingo, users lose the benefits of big screens in return for gaining ubiquity and they lose the benefits of big keyboards for gaining availability anytime. The trade-off criteria may explain why one participant was very convinced that it is good to be able to post-edit on mobile, while another participant said that the use case for Kanjingo for professional post-editing was not convincing. Taking into account that Kanjingo was not designed for professional post-editing, it seems that participants were willing to accept the trade-off for post-editing for not-for-profit projects, whereas when productivity is at stake, they would prefer the best ergonomics possible. As development has begun for creating multimodal editor for larger touchscreen devices, it remains to be seen whether an increase in screen real estate will change user perceptions.

The testing methodology also required some trade-off. Testing was done with only five participants but it added five observers, one of whom was the developer of the app, who had never seen users testing the app before: in total, ten people providing valuable feedback. From our point of view, this methodology has proved useful for several reasons: 1) qualitative assessment is enriched with different perspectives; 2) more members of the research centre are actively involved in the testing, providing valuable insight; 3) useful preliminary results are obtained and discussed on the testing day, while information is fresh.

Four main conclusions can be drawn from this evaluation:

1. MT post-editing with a mobile app is feasible and realistic (when limitations are taken into account);
2. Improvements are needed for Kanjingo, mainly in physical ergonomics and editing features;
3. The small keyboard and small screen-size mean that the app is not very suitable for professional post-editing tasks (as opposed to not-for-profit post-editing);
4. The potential for voice input as a solution to keyboard limitations is a topic to be further explored.

Productivity and quality were not evaluated in this study, as the focus was on the interaction with the app rather than post-editing results. Users had total freedom to interact with the app as long as they adhered to the test script and post-edited the same text, and they did not focus on providing a fast or defined-quality output, so post-edited data would not necessarily be comparable. Proper productivity and quality tests should be designed to compare the app with a desktop environment. Nevertheless, from the screen recording of this test we may suggest two starting hypotheses for future productivity and quality tests:

1. Productivity would be faster in a desktop environment. From our screen recordings, we have noticed that participants use just one finger to type, which results in limited productivity. Participants also admitted they were more used to typing using a desktop PC and keyboard, although they felt that their productivity with the app would increase with practice.
2. Quality would be similar in both environments. In our screen recordings, we have observed that post-editing segment by segment in a mobile app makes the user concentrate their focus at the segment level. Moreover, the fact that only the current segment appears on screen (as compared to desktop applications with many elements on screen at once), using a relatively largely font size, makes the identification of mistakes an easy process. In our test, mistakes were quickly identified by all participants. Finally, the app's colourful and aesthetically clear design (as opposed to dull text editors) might help to maintain the users' focus.

As described above, these are only starting hypotheses derived from the observation of the screen recordings and these need to be tested. After this evaluation we conclude that more testing would be beneficial, with more users and with different professional and personal profiles. Also, more tests are needed with different tasks and with different levels of MT quality. For instance, tests on MTPE productivity, tests with longer texts or more sentences (to test intensity of use), testing real assignments, testing in different environments (not only a test controlled environment), testing the app as a tool for MTPE training (to spot MT mistakes or to get familiar with specific MTPE guidelines), or even for foreign language learning. An assessment of how use of the app impacts on the final quality of post-edited text would also be important for defining potential use cases.

Acknowledgements

This work was supported by the ProjecTA project "Translation projects with Statistical Machine Translation and Postediting", grant number *FFI2013-46041-R [MINECO / FEDER, UE]*, and by the ADAPT Centre for Digital Content Technology, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and co-funded under the European Regional Development Fund.

References

5min4good (Producer). (2013, November 4) Welcome to 5mins4good [Video file]. Retrieved from https://www.youtube.com/watch?v= qBEoufZ_qJE

Dragsted, B., Mees, I. M., & Gorm Hansen, I. (2011). Speaking your translation: Students' first encounter with speech recognition technology. *Translation & Interpreting 3*(1), 10-43.

Dymetman, M., Brousseau, J., Foster, G., Isabelle, P., Normandin, Y., & Plamondon, P. (1994, September). Towards an automatic dictation system for translators: The TransTalk Project. *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*. Retrieved from http://arxiv.org/abs/cmp-lg/9409012

Guzmán, R. (2007). Manual MT post-editing: "If it's not broken, don't fix it!" *Translation Journal 11*(4). Retrieved from http://translationjournal.net/journal/42mt.htm

Koponen, M. (2012, June). Comparing human perceptions of post-editing effort with post-editing operations. *Proceedings of the Seventh Workshop on Statistical Machine Translation* (pp. 181-190). Stroudsburg, PA: Association for Computational Linguistics.

Krug, S. (2006). *Don't make me think!: A common sense approach to web usability*. San Francisco: New Riders.

Krug, S. (2009). *Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems.* San Francisco: New Riders.

Kukulska-Hulme, A., & Traxler, J. (2005). *Mobile learning: A handbook for educators and trainers*. London, UK: Routledge.

Martín-Mor, A. (2017). MTradumàtica: Statistical machine translation customisation for translators. *Skase Journal of Translation and Interpretation*, *11*(1), 25-40.

Mesa-Lao, B. (2014). Post-editing through speech recognition: a feasibility study with post-editor trainees. *Proceedings of 2014 CRITT - WCRE Conference*. Frederiksberg, Denmark.

Moorkens, J., O'Brien, S. & Vreeke, J. (2016). Developing and testing Kanjingo: A mobile app for post-editing. *Tradumàtica*, *14*, 58-66.

Nielsen, J. (2000, March 19). *Why you only need to test with 5 users*. Retrieved from https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users

O'Brien, S. (2012). Translation as human-computer interaction. *Translation Spaces*, *1*(1), 101-122.

O'Brien, S., Moorkens, J., & Vreeke, J. (2014). Kanjingo: A mobile app for post-editing. *Proceedings of Third Workshop on Post-Editing Technology and Practice* (pp. 137-141). Dubrovnik, Croatia.

Specia, L. (2011, May). Exploiting objective annotations for measuring translation post-editing effort. *Proceedings of the 15th Conference of the European Association for Machine Translation,* 73-80.

Toselli, A. H., Vidal, E., & Casacuberta, F. (2011). *Multimodal interactive pattern recognition and applications*. Berlin: Springer Science & Business Media.

Torres-Hostench, O. (2016). *Pros and cons of real life assignments in a localization course*. TAUS blog. Retrieved from https://events.taus.net/blog/pros-and-cons-of-real-life-assignments-in-a-localization-course

Vidal, E., Casacuberta, F., Rodriguez, L., Civera, J., & Hinarejos, C. D. M. (2006). Computer-assisted translation using speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(3), 941-951.

[i] Kanjingo testing task #1 was related to a real post-editing assignment. Fourth-year students from a Localization course in the degree programme of Translation and Interpreting at Universitat Autònoma de Barcelona participated in a service-learning project consisting of the localization from English into Spanish of the TAUS website <taus.net> (Torres-Hostench, 2016). Localization students translated and post-edited

up to 20,000 words of the website as part of a real assignment. Smartling was the localization tool used in the project. It includes project management features plus translation memory and machine translation features integrated into the editing area.

[ii] Global Voices corpus, found on the Opus open parallel corpora repository, was used to train the machine translation engine MTradumàtica developed by Prompsit as part of the ProjecTA project [Ref. *FFI2013-46041-R*]. Segments from the Global Voices website not included in the corpus were machine translated with MTradumàtica before and after the training with the specific Global Voices corpus. Segments machine translated after the training were chosen for the Kanjingo testing task 2. The reason why MTradumàtica was used in this study is because the Kanjingo team also collaborates on the ProjecTA project, and MTradumàtica is in testing phase.

[iii] http://www.mirroring360.com/

[iv] The testing facilitator belongs to the Universitat Autònoma de Barcelona. The Kanjingo testing in 2016 was part of a summer research stay at the ADAPT Centre in collaboration with the ProjecTA project.