Kanjingo – A Mobile App for Post-Editing

Sharon O'Brien
CNGL/SALIS
Dublin City University
Ireland
sharon.obrien@dcu.ie

Joss Moorkens
CNGL/SALIS
Dublin City University
Ireland
joss.moorkens@dcu.ie

Joris Vreeke
CNGL
Dublin City University
Ireland
joris.vreeke@dcu.ie

Abstract

This paper describes the Kanjingo post-editing application for smartphones. The application was developed using an agile methodology at the Centre for Global Intelligent Content (CNGL) at DCU and a first stage of user testing was conducted using content from Translators Without Borders.[1] Initial feedback on this app was quite positive. Users identified some particular challenges, e.g. input and sensitivity limitations, insufficient Help, lack of automatic punctuation and capitalization. Development and further testing are ongoing and may include interactive MT, speech as input and focus on Asian languages as target languages in the future.

1 Introduction

Kanjingo is a mobile app for translating a source text and post-editing machine translated target text on a mobile interface. It was developed in the CNGL (Centre for Global Intelligent Content) at Dublin City University.[2] This paper describes the first round of user testing where the objective was to obtain feedback and improve the application.

2 User Testing

2.1 Motivation

The objective of the first stage of user testing focused on Kanjingo's suitability for post-editing machine translated output in a mobile scenario. The motivation for doing so is based on the increasing evidence that volunteers are willing to translate or post-edit for causes they wish to support (Munro, 2010; Petras, 2011)

Our use case scenario for this first round of testing is volunteers for an organization such as "Translators Without Borders" (TWB). The volunteers wish to contribute to the translation effort of this organization, but possibly only have time to translate or post-edit a few segments of text per day on their way to and from work. Our assumption is that volunteers may not wish to sit at a desk to do this work and might like to post-edit a few segments of text while waiting at a bus stop, for example.

The Kanjingo App is not intended to replace a desktop CAT environment. However, since MT suggestions sometimes need to be deleted outright due to poor quality and retranslated by a human, we decided to also test the App's potential to support the human translation task in addition to the post-editing task.

Figure 1. The Kanjingo post-editing screen

[1] We are very grateful to Translators Without Borders for their collaboration in this project.

[2] This research is supported by the Science Foundation Ireland (Grant 12/CE/I2267) as part of CNGL (www.cngl.ie) at Dublin City University.

2.2 UI Description

When the App is first accessed, the user selects a language pair, e.g. English-French. Source segments are listed in the initial screen presented to users. The user selects a source segment at which point a machine translated segment is presented on the screen in a vertical tiled format (see Figure 1).

If at first users do not know how to interact with the UI, they can click on a Help link which presents them with a screen shot explaining the basic features of the UI (see Figure 2).

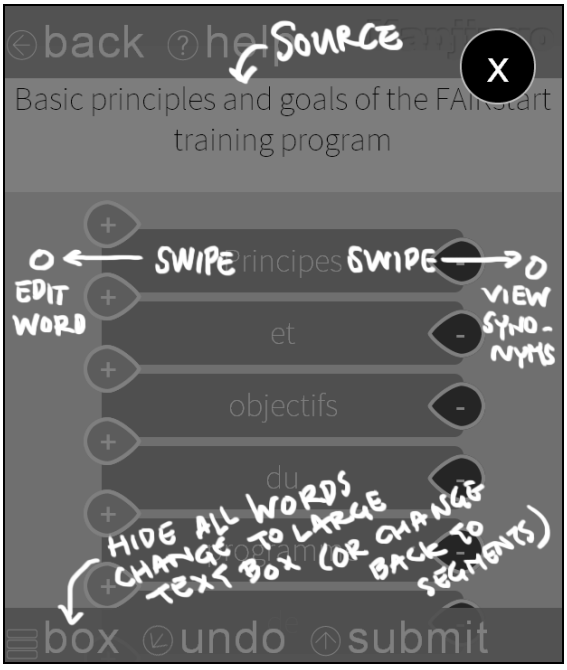

Figure 2. Basic help

As can be seen in Figure 1, each word tile has a "+" symbol on the left which, when tapped, inserts a new empty tile above that word, enabling the post-editor to insert a new word. The '-' sign on the right side of the word tile deletes that tile completely. Tiles can be reordered by dragging the tile up or down on the screen. Users can scroll down through the MT segment by dragging elsewhere on the screen, as with the regular scrolling feature on a smartphone. To edit a word, the user taps on the word, which appears in an edit box. A second tap in the box enables the appearance of the smartphone keyboard.

Once the user has post-edited the segment to his/her satisfaction, the segment can be submitted by clicking on the Submit button, located at the bottom of the screen.

As mentioned above, we also tested the App for translation functionality. When users selected a segment in the translation mode, an empty text box was presented into which they had to type their translation using the phone's keyboard (see Fig. 3).

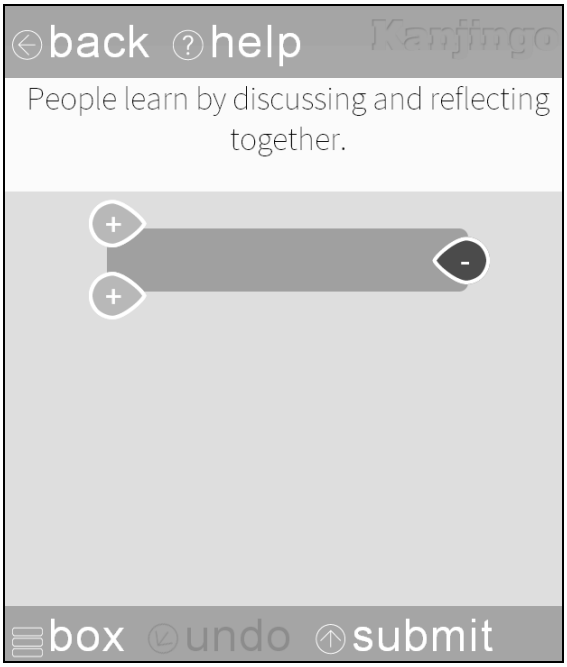

Figure 3. The Kanjingo translation screen

2.3 Setup

Translators Without Borders provided us with sample content in source languages English and French, which were machine translated using Microsoft Bing Translator into French, Spanish, and English.

We recruited five users (2=male) with different backgrounds with the objective of including users of different profiles who were likely to fit the profiles of those who might volunteer to post-edit MT output. Their profiles are listed below with the language pair and direction they worked with listed in brackets.

- Professional translator who also has experience of post-editing in desktop scenarios (Fr-En);
- Research engineer who works with machine translation (En-Fr);
- PhD candidate who is currently researching audio-visual translation/fansubbing (En-Sp)
- Lecturer in language studies who has a Master's qualification in translation and interpreting (En-Sp)
- Master's student of translation, with undergraduate studies of translation (En-Sp).

Although this is a small group of users, this was an adequate number for initial user feedback. Nielsen has written that several iterations of usability testing at this scale maximizes the cost-benefit: "The best results come from testing no more than 5 users and running as many small tests as you can afford" (2000). The variation in profiles is also in keeping with best practice in UX testing.

Participants were requested to use concurrent Think Aloud Protocol (TAP), that is, to speak their thoughts about the task or the user interface during the task (Ericsson and Simon, 1980; 1999). Nielsen wrote that "thinking aloud may be the single most valuable usability engineering method" with some caveats, in that it may bias user behavior, and decrease productivity (1994, p195). This study, however, focuses on usability rather than productivity, so TAP was considered worthwhile, although in practice it transpired that "some test users have great difficulties in keeping up a steady stream of utterances as they use a system" (ibid., p196). Whatever TAP content was produced was transcribed and analyzed for comments that allowed us to identify the strong and weak points of the UI in both the post-editing and translation modes. Following the user interface test, participants were asked ten questions as part of a structured debriefing interview to help elucidate their evaluation of the App.

3 Results

The four users most familiar with smartphones were quickly able to edit the machine translated segments and had fairly positive attitudes towards the App in general, e.g. User 4 saw it as "ideal for short messages or perhaps emails with two or three sentences." Several participants said that they found the App intuitive, with user 5 commenting that "I think it's quite friendly, usable as well - easy-to-use." Most participants were pleased with the drag-and-drop functionality. On the other hand, the user with least experience of smartphones struggled to use the App and disliked it more than the others. This user did not appear to understand the drag-and-drop functionality, despite having "accidentally reorganized the sentence without wanting to", and found the App largely frustrating. She was one of several users who hit or touched buttons by mistake.

Accidental manipulation of the UI was one of several problems or frustrations encountered repeatedly during the tests. In summary, these were:

- The lack of automatic punctuation and capitalization
- Problems with sensitivity
- Loss of unsubmitted work if the user leaves the UI to check the Internet or dictionaries/glossaries
- Insufficient Help
- Input functionality challenges

Each of these issues is discussed in more detail below.

3.1 Punctuation and capitalization

Four of five participants voiced frustration at having to manually add capitals at the beginning of a segment and having to append a full stop at the end. When the capitalised word from the MT output was moved, this exacerbated the problem. This can be seen in Figure 4. In the next version of the App, any word moved to the top of the tiled list will be automatically capitalized and the full-stop will be attached only at the end of the segment, even if the last word is moved up.

Figure 4. Incorrect capitalization.

3.2 Sensitivity

One of the main challenges in designing a smartphone App for text editing purposes is the limited space available to display the text. This problem is exacerbated when two segments have

to be displayed within the UI. The challenge increases if the segments are "long". The shortest segment in our content contained 3 words while the longest contained 20 words. The limited display led to issues regarding over- and under-sensitivity in the App. One user in particular had trouble hitting the plus and minus signs (see Figure 1 above). Some users accidentally tapped the 'undo' button when intending to 'submit' the segment that had been processed. Users also dragged word tiles accidentally when they simply meant to scroll up or down. Users mentioned that they wanted to group words and drag and drop them together, but this facility was unavailable. Our development team is investigating providing more space on either side of the segment display for scrolling and means of grouping words for combined drag and drop functionality.

3.3 Retention of unsubmitted work

Once a segment is edited, the users could submit it using the "Submit" button. One comment in relation to this was that they were unsure what had happened to the submitted segment because there was no confirmation message. A confirmation message (or other form of visual feedback) will therefore be added to the next version.

It may occur that a user is half-way through a segment and needs to abandon it for a period of time (the bus has arrived or an interesting message has popped up on Twitter!). The half-edited segment was then lost if the user toggled to another app. The development team is investigating ways of using the phone's local storage to save edits in progress. However, we also need to consider what impact this has on progress if the user decides never to come back to the segments and they cannot be picked up by an alternative user.

3.4 Insufficient Help

We wished to investigate how intuitive the App was with only limited Help available to users. The level of Help available is depicted in Figure 2 above. It became obvious that the Help function in the App was inadequate. The user who was least familiar with smartphone apps, tried to avoid clicking 'Help' but eventually relented. Other users commented that they would like to see walk-though instructions before using the App for the first time.

3.5 Input Problems

Due to the limited area available, input was challenging. Particularly for the human translation scenario, input was a frustrating bottleneck between the users and the App. One participant commented that the speed for typing was so much slower than for a desktop scenario. The keyboard sometimes got in the way of the text box for translation. Another user complained because no synonyms or auto-completions were offered. For the translation scenario, one possible solution might be speech as input, though of course this is limited by the environment in which the translation takes place (e.g. if it is noisy, speech recognition may be compromised). We will also look at connecting the App to resources that would allow for synonym suggestion and auto-completion.

We were aware that the content we selected for this initial stage of testing was particularly challenging because (i) it was continuous text for which context was important and (2) some segments were relatively long. Limiting the length of segments would help solve the input problems, but this would also mean ruling out the use of the App for content that is typical to an organization like Translators Without Borders. Moreover, shorter segments bring their own challenges in respect of machine translation output quality and Tweets, or other forms of user-generated content, can also be difficult to decipher. Just limiting to short segments or Tweets is, therefore, not very desirable.

4 Conclusions

In embarking on this small-scale user testing of the mobile post-editing and translation App, Kanjingo, we expected a rather negative response from users given the severe space limitations of the mobile text-editing environment. However, although they were critical of certain aspects, the users were fairly positive about the App and gave some highly useful feedback. This feedback has been taken on board by the development team who are now in the process of developing a new version, for which we intend to do larger-scale user testing.

Future development work could potentially focus on interactive machine translation, speech input and Asian languages as target languages.

References

Ericsson, Anders K. and Herbert A. Simon. 1980. Verbal reports as data. *Psychological Review* 87, 215-251.

Ericsson, Anders K. and Herbert A. Simon. 1999. *Protocol Analysis: Verbal reports as data*. Cambridge, MA: MIT Press, 3rd edition.

Munro, Robert. 2010. Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge. *AMTA 2010: The Ninth Conference of the Association for Machine Translation in the Americas*, October 31 - November 4, 2010, Denver, Colorado.

Nielsen, Jakob. 1994. *Usability Engineering.* Academic Press, Cambridge, MA.

Nielsen, Jakob. 2000. Why You Only Need to Test with 5 Users. Nielsen Norman Group. [Accessed from http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/]

Petras, Rebecca. 2011. Localizing with community translation. *Multilingual 22(7)*, 40-41.