Developing a Customisable Subtitling Tool Based on Academic Research and User Needs

Željko Radić [0009-0003-9451-5557]
University of Surrey, Guildford, Surrey GU2 7XH, UK
z.radic@surrey.ac.uk

Abstract. SpeakSubz is a tool for translating non-live subtitles primarily relying on Google’s speech recognition technology: Google Voice Typing. This specialised software also includes functionalities that make it a viable tool for academic research on various topics related to subtitling and voicing/respeaking. It was used as a tool within a larger doctoral study called ‘Interlingual Subtitle Voicing: A New Technique for the Creation of Interlingual Subtitles, A Case Study in Croatian’. This study aimed to measure speech recognition accuracy for Croatian subtitles, the duration of various elements of a subtitling workflow, and, ultimately, to assess participants’ reception of the technique and the tool and elicit their feedback for its future development. The tool is being developed based on academic scholarship, industry insights and, most importantly, subtitlers’ needs. Within the framework of action research and the notion of subtitler experience and as highlighted in this paper, the tool is developed in cycles akin to software updates. In each cycle, a major new variable is introduced and tested with other minor changes related to either functionality or the interface. This paper will also highlight some of the most relevant functionalities that distinguish the current version of SpeakSubz from similar tools. The goal is to create a customisable tool for human-informed translation of subtitles that can be used for training and research in academia as well as professional environments.

Keywords: SpeakSubz, ISV, Subtitling, SUBX, Action Research, Software Development, Research Cycles

1 Introduction

Interlingual Subtitle Voicing (ISV) is a technique devised for translating non-live subtitles by voicing, traditionally known as respeaking. SpeakSubz, a specialised tool, was created to assess the technique’s viability. This tool enabled quality measurements of target language output, emphasising speech recognition (SR) accuracy, as well as durations of various subtitling workflow elements such as typing, voicing and editing. Eventually, the ISV study resulted in subtitlers’ evaluation of both the technique and the accompanying software, which was crucial for further development.
This strand of research started with an MA thesis completed at the University of Roehampton in 2018 [1], which resulted in the desktop version of the tool (the first ISV cycle) from an improvised methodological setting: uploading subtitles to the Google Voice Typing (GVT) interface and translating them by voice. The desktop version of SpeakSubz was further developed as part of a doctoral thesis currently being finalised at the University of Surrey. After internal testing within the Centre for Translation Studies (CTS) at Surrey and the pilot study with students, an online version of SpeakSubz was created (the second ISV cycle), further improved after the main experiment.

2 Development Cycles

Per action research [1]-[3], the ISV doctoral study was exploratory, participatory and conducted in cycles. Each cycle was informed by relevant academic research, industry experiences at that time and feedback coming from participants of pilot and main studies. This enabled the shaping of the tool according to future users’ needs and preferences, in line with the concept of agile software development [5]; it also offered them multiple functionalities that they could use as needed, based on individual preferences. The tool was developed by a Croatian software developer, A. Vrešić, and integrated with the ISV website1 created by a Croatian web designer, A. Prskalo.

The ISV website, primarily the result of the Covid-19 pandemic, is an online hub for experiments containing all the necessary information and materials for ISV training and testing. Such interdisciplinary cooperation enabled a streamlined training and testing environment for the new technique. The following sections will chronologically describe the cycles of the software development process in brief.

2.1 Desktop SpeakSubz

As mentioned above, the desktop version of SpeakSubz resulted from an MA study that compared the speeds of typing subtitles interlingually versus voicing them. The experiment was conducted in the GVT application [6] in Google Docs. Participants had to copy subtitles into the GVT interface and translate them by voicing. At the same time, they had to track and note down the times they spent on each exercise, including possible breaks. In addition, they had to also record their process with an external tool: Screencast-o-Matic (now ScreenPal2).

---

1 https://isvresearch.eu/
2 https://screenpal.com/
Besides collecting empirical data, participants were also invited to give feedback about the technique, which, among other factors, shaped the next iteration of desktop SpeakSubz (Figure 1). This version was also used for internal testing within CTS at Surrey and piloting with translation students from the Faculty of Humanities and Social Sciences at the University of J.J. Strossmayer in Osijek, Croatia. Although some of the piloting was initially supposed to occur face-to-face, because of the Covid-19 pandemic, the methodological process had to be transferred online. This resulted in the ISV website that contains all the information about the ISV technique and the software in the form of PowerPoint and video presentations, as well as short video tutorials on various functionalities of SpeakSubz. This piloting process resulted in an online SpeakSubz version integrated with the ISV website.

### 2.2 Online SpeakSubz

The online iteration of SpeakSubz was a more advanced version because it included a plethora of functionalities suggested by participants of the previous experiment cycles. In Figure 2, for example, in comparison with the desktop version of SpeakSubz subtitle boxes are much bigger, there is a big green button to activate or stop a speech recognition session, visual reading speed markers (thin green bars that turn red once the reading speed is exceeded) and comment boxes. The interface is similar to some existing subtitling platforms, especially cloud-based ones, to evoke a sense of familiarity.

Online SpeakSubz was created in dark mode, per software development industry standards. Certain academic research also explores various aspects of the dark mode. Some studies show that dark mode reduces eyestrain in lower ambient illumination [7] and increases visual acuity, which is important for an efficient and pleasant subtitler experience (SUBX). In addition, one study shows that dark mode
could potentially save energy [8]. This is especially relevant for SpeakSubz because it is aimed to be used on mobile devices such as tablets or hybrid laptops and away from energy outlets, thus becoming an ergonomic tool.

SpeakSubz is an example of how a tool can improve the existing underlying technology and tailor it to specific needs. For instance, the desktop version of SpeakSubz introduced voicing punctuation, which even GVT does not offer for Croatian. The online version went even further and enabled entering SR misrecognitions into the SpeakSubz dictionary for better accuracy. The online version of SpeakSubz also includes automatic spell-checking to increase output quality further. This version of SpeakSubz also shows a number of characters per row, which is critical for readability in non-live subtitling.

Regarding research methodology, the online version of SpeakSubz also has an integrated screen and voice recording functionality and pre-loaded tasks that had to be loaded and exported manually in previous versions. This automatisation of various processes resulted in participants’ positive experiences with the ISV technique and SpeakSubz software, resulting in positive reviews – scoring above 4 out of 5 (4.06 and 4.13, respectively) – despite imperfections in the underlying technology (GVT). This data stems from post-testing questionnaires (15 participants) and optional interviews (6 participants) conducted after the main ISV experiment, providing us with a wealth of information crucial for future developments.

To sum up, SpeakSubz is a tool that aims to combine different functionalities needed to create non-live subtitles in multiple hybrid workflows. The interface’s similarities to existing online solutions help users adjust to the tool. However, most current subtitling software does not include SR as an input method, even for major languages, leaving much space for exploration. Additionally, SpeakSubz enables entering misrecognitions, which the underlying technology, GVT, does not offer in its interface. This makes
further improvements possible in SR accuracy, especially for lower-resourced languages.

Regarding teaching and research, SpeakSubz’s integrated recording functionality allows users to re-watch their performance, reflect on it and improve aspects that need to be practised. Professional subtitling software generally used in audiovisual translation (AVT) research usually does not offer such functionality. Researchers, in addition, can utilise these recordings to observe the translation process and participants’ behaviours. In short, SpeakSubz offers functionalities that facilitate methodological processes that most professional desktop or cloud-based software like WinCaps, Spot, EZTitles and Oona do not provide, primarily to protect copyrighted source materials.

Based on the above and to my knowledge SpeakSubz would be the only subtitling tool from academia tailored to academic teaching, training and testing that offers technologies and functionalities – such as SR input and integrated recording – that professional tools do not provide. Furthermore, SpeakSubz is highly adaptable and can include and adapt new and emerging technologies in each new cycle to further test them to gauge whether they increase efficiency. Such an improved tool could then be presented to industry stakeholders, both subtitlers and clients.

3 Future Development

Based on the data from the main ISV experiment, the following goals have been established for the subsequent research and development cycle (the third ISV cycle): introducing a new major technology into the workflow (machine translation), testing the technique and the software on mobile and touch-screen devices and adding other functionalities suggested by participants in the ISV main experiment from the second cycle. For example, the current version of SpeakSubz includes virtual buttons for adding punctuation by touch as an alternative to voicing or typing them; this feature is yet to be tested and evaluated within the ISV workflow.

These elements will be included in the next stage of testing hybrid workflows. However, this time, we will provide the participants with various functionalities and measure and test what they prefer the most and why so that the tool can be customised even more in future. These will be explored under the concept of subtitler experience (SUBX) introduced in the ISV study, stemming from translator experience – (TX) [9–11], the term that Zapata derived from the general user experience (UX). In short, SpeakSubz will be explored not only as a training and testing, as well as a professional tool but also as an ergonomic one that enables subtitlers to work in more comfortable environments, away from the traditional keyboard-desk setting.
4 Conclusion

This demo paper describes the development of specialised software based on academic research for academia, primarily for creating and quantitatively testing hybrid subtitling workflows and new technologies. The software also enables gathering qualitative data about the specialised subtitling tool and the new technique being developed around it. Suppose empirical research shows this tool can increase the efficiency of subtitlers while keeping in mind their physical and mental well-being. In that case, it might also be developed commercially and offered to major clients in the AVT industry.

This interdisciplinary and participatory research not only gathers data from participants but also teaches them new skills and aims to inform them about the benefits of ergonomics, which can be done anonymously online. Such an environment is conducive to collecting objective data and ample qualitative feedback, with one main goal: to research what subtitlers need and how this novel form of human-machine interaction can produce the best results.
References

6. Google Docs Editors Help: Type with your voice.