Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Text Entry for All: Towards Speech-based Multimodal Interaction for Inclusion, Accessibility and the Preservation of the World's Linguistic Heritage

Julián Zapata[1], *Lara Hanna*[2]

[1]Department of Languages, Literatures and Cultures, Toronto Metropolitan University, Canada
[2]Department of Anthropology, University of Toronto, Canada
julian.zapata@torontomu.ca, lara.hanna@mail.utoronto.ca

Abstract

This paper describes an emerging project that aims at better understanding how speakers of different languages, with diverse needs, produce texts today, and rethinking how they should do so in the future. Text entry (TE), the process of inputting text into a computer, is a crucial part of our interaction with technology. However, research has shown that the conventional TE method—typing on a keyboard—is not the most efficient or appropriate way of inputting text in various contexts, for diverse users. Now, how can we design more effective, accessible and inclusive TE methods that consider the uniqueness of different languages and diverse users and use cases? This fascinating question is likely to motivate speech technology, writing-process and human-computer interaction researchers alike, particularly in the generative AI age. But we are also living in the age of speech and multimodal interactions, which offer unprecedented opportunities to reinvent how we learn, write and communicate.

Index Terms: text entry, speech input, multimodal input, writing, natural languages, inclusion, accessibility.

1. Introduction

Chances are the reader is reading this paper on a conventional keyboard-based personal computer or laptop. If they are a mobile device user, chances are they will manipulate and interact with this text (e.g., add notes, forward it on social media while adding a personal thought, copy/paste a provoking passage, etc.) using an app that emulates a physical (or mechanical) keyboard, but on a touchscreen instead.

Text entry (TE) is the linchpin of modern global communication. From typing on physical keyboards to swiping on touchscreen devices, people of diverse linguistic backgrounds rely on text input methods to express thoughts and ideas in writing and interact with textual information. That being said, a growing body of research has indicated that (1) humans have typed only for a few decades, yet written in multiple other ways for millennia, (2) not everyone is able to type, and (3) certain languages pose challenges for traditional keyboard-based input due to character complexity or particular phonetic structures. This paper offers speech recognition (SR) and multimodal interaction (MI) perspectives for an emerging research project that aims at better understanding how speakers of different languages (beyond English), in different use scenarios, produce texts today, and rethinking how they should more efficiently do so in the future. The fields of TE, SR and MI have long shared interests and worked towards common goals, namely the demonstration of the inappropriateness of existing keyboard-and-mouse interfaces and, by the same token, the development of devices and applications that accommodate the practical aspects of usability while people are mobile and in field contexts. Indeed, one of the impetus for developing speech-based multimodal interfaces has been supporting people's ability to communicate fluently as they think and work using multiple modalities (speech, touch, pen, gestures, etc.), multiple representation systems (symbols, numbers, language, etc.), and multiple linguistic codes (e.g., Western, Asian or Indigenous languages, etc.) [1, pp. 421–423]. Now, while English remains a global lingua franca in many digital spheres, speakers and users of other languages contribute significantly to the digital discourse. As a result, understanding and rethinking TE in different languages are pivotal considerations in promoting inclusive and efficient global communication. Most TE research has focused on English-language-based tasks, thus precluding research using non-English-speaking populations [2]. Furthermore, technological advances in language education, in particular, have emphasized the importance of accommodating individuals with disabilities who may find mechanical keyboards challenging to use [3], [4].

The need for stronger, more comprehensive multidisciplinary research in TE has never been greater, particularly with the proliferation of mobile computing devices and the increased representation of the world languages in digital communications. As Dunlop et al. [2] put it, "effortless [TE] has become a badly supported requirement for a very wide diversity of population, including those with disabilities and niche languages, or dialect speakers. However, even when considering support for large user populations, the community is still unearthing strong evidence that better [TE] method design is still needed, for example in non-English-speaking countries." Consequently, we believe the fields of SR and MI can significantly advance TE research and development. This is because SR and MI inherently prioritize the end users of computer systems and emphasize interacting with machines through natural human behavior. By leveraging diverse TE input methods to enhance accessibility and efficiency for all, including those with disabilities and users of any of the world's written languages, and by integrating SR and e-pen input, gesture control, and other forms of natural human behavior, multimodal interfaces can potentially offer more intuitive and adaptable TE solutions.

10.21437/Interspeech.2025-1036

2. An Overview of Text Entry Methods in Languages Beyond English

The global adoption of the mechanical keyboard significantly contributed to the spread and popularity of personal computers across new regions, from Europe to the Middle East to Asia. Today, even the operating systems (OSs) of some (hybrid) laptop computers –for which a rather thin physical keyboard or "type cover" is nonetheless optional and sold separately– offer a staggering 96 possible keyboard layouts during installation, ranging from "Canadian Multilingual Standard" to "Hungarian 101-key" to "Inuktitut - Latin" to "Wolof," to name a few.

The introduction of the QWERTY keyboard layout, initially designed for the English language, paved the way for new typing methods in regions like the Middle East, for example, where languages such as Arabic and Persian are widely spoken. However, it also posed challenges for typing in such languages, which use different scripts and are written from right to left [5, p. 43]. As a response to the typographical barriers hindering the use of Arabic script in the 1990s, for example, Arabic-speaking typists popularized a hybrid form of typing that utilizes both Latin letters and numbers in place of Arabic characters when needed (e.g., the number "3" to substitute the letter "ع"). This hybrid form of communication, known as "Arabizi," or "Franco-Arabic," was especially prevalent in informal online communication, where it became easier for users to type Arabic sounds with the existing Latin-based QWERTY keyboard rather than switching to an Arabic layout. Arabic versions of the QWERTY keyboard were later developed, with Arabic characters replacing the Latin ones on the keys or appearing alongside Latin keys, thus allowing for bilingual typing. However, Arabizi also remains widespread throughout the Middle East today for several reasons. In a study at the University of Petra, Yaghan [6] found that Arabic-speaking youth preferred typing using Arabizi as they believed that it could express emotions and tone more efficiently due to its ability to capitalize letters that the Arabic keyboard and script lack. However, the use of QWERTY-based Arabic keyboards and the rise of mixed typing practices (e.g., Arabizi) continue to spark debates about linguistic purity and the preservation of the Arabic script. Some purists argue that Arabizi and similar adaptations might erode the traditional use of Arabic script, while others see them as pragmatic solutions for modern communication [7]. While Arabizi did initially provide a solution to the allegorical irony which was the comically small appearance of Arabic next to English script in the early aughts of the Internet, and still has its benefits, considering other methods of Arabic TE for the future would be fruitful.

The introduction of mechanical keyboards in Arabic-speaking regions eventually spurred the development of software and operating systems that supported TE in that language. This included localizing OSs, developing word processing tools, and integrating Arabic input in mainstream applications like MS Office and various web browsers. Several open-source projects developed software for Arabic on Drupal, Yamli, Google, and other platforms in the years that followed, thereby enabling Arabic-language content to grow dramatically [5, p. 43]. Innovations like predictive text, SR, and advanced input methods for Arabic also improved the efficiency and accessibility of Arabic typing on QWERTY keyboards, addressing the initial challenges of using a Latin-based layout for Arabic. Despite these developments, the Arabic QWERTY keyboard still has its limitations, especially regarding technical problems.

Arabic, like Persian and other languages, is written in a cursive style where letters within a word are connected, and each letter has up to four forms: isolated, initial, medial, and final. This means that software must dynamically adjust the shape of each letter depending on its position within a word, making it complex to implement and maintain. Also, the shape of a letter can change depending on its context and the letters around it [8]. This requires robust software that can correctly render each form based on its position in a word. Additionally, Arabic typing can be slower due to the need for managing character variants and the use of additional keystrokes for diacritical marks. Also, the frequent need to switch between Arabic and Latin characters can cause ergonomic discomfort and affect typing speed. Research has been conducted in attempts to simplify typing in Arabic by implementing the best or the most efficient single-finger Arabic keyboard layout that allows users of mobile devices to write text and carry out written conversations for a long time with comfort, ease, and speed [9]. Such hindrances further push the need to consider new TE methods such as multimodal speech- and e-pen-based input; see Oviatt [10]. SR can bypass the difficulties posed by Arabic's cursive script and multiple letter forms by allowing users to input text without manually navigating character variants or diacritics. Moreover, e-pen-based input could offer an additional layer of flexibility in a multimodal interface with speech. These methods can adapt to both casual users and those with accessibility needs, providing smoother transitions between spoken and written forms of Arabic. Our study will include user evaluations comparing typing with SR and e-pen and other touch-based inputs, focusing on speed, accuracy, and user satisfaction in real-world scenarios. This approach aims to demonstrate how speech-based multimodal input can offer more intuitive and efficient TE methods, especially for languages with complex writing systems.

Asian languages also provide other key examples of the importance of better understanding the particularities of a language and the needs of its users to rethink TE methods. An application such as Katakana EdgeWrite, for example, has transformed e-pen-based TE for Japanese users by providing an innovative method for inputting Katakana characters. This software combines stroke recognition with context-aware algorithms, allowing users to simply draw patterns with their e-pen to form the characters. With its predictive-text capabilities and a low learning curve, the app has significantly improved the efficiency and accuracy of TE on touchscreen devices for Japanese language users. By simplifying the process of producing Katakana characters, this technology enhances the user experience and showcases the potential for innovative solutions in language-specific TE methods. Moreover, Katakana EdgeWrite stems from its original Korean version with hopes of facilitating TE on a variety of other East Asian

languages [11]. Moreover, models such as Distance and Statistical Language Model have facilitated further research to improve smartphone-based TE, with the latter focusing on the pre-existing keyboard on most smartphones, as evidenced in the study by Janpinijrut et al. [12] for the Thai language.

Beyond Middle Eastern and Asian languages, research on European languages has highlighted the need to move past archaic forms of mechanical TE. Böcker et al. [13]'s study sheds light on how Europe's nearly 230 indigenous languages take a backseat to languages with Latin alphabets when it comes to standard mechanical keyboards, and suggested methods to improve TE using a 12-key keypad in mobile devices for a range of European languages. Dagienė et al. [14], who analyzed the existing keyboard layouts on mobile devices for various languages including Lithuanian, demonstrated the influence of keyboard usability on the number of typing errors, and proposed keyboard design guidelines for non-English languages.

These research projects are a testimony of the continuous interest in TE and highlight the potential of speech-based multimodal input to become the ultimate solution for future generations wishing to bypass the barriers imposed by mechanical keyboards, and that, in any of the world's written languages. In the words of Dunlop et al. [2] "work into supporting local dialects, niche or disappearing languages through better typography, semiotics and intelligent entry support, must form a core target as we aim to culturally preserve, rather than help towards elimination, of this rich heritage", a statement that is greatly supported by researchers in the MI field such as Oviatt, for instance, who believes that one of the goals "of developing richer communication interfaces is [the] preservation of the world's diverse heritage languages, many of which are poorly supported by western-European keyboards" [1, p. 423].

3. Text Entry for Inclusion and Accessibility

The need for speech-based and multimodal approaches to TE for future generations cannot be understated. Kristensson [15] expands on the realities of smartphone users that crave convenience in their TE methods, and suggests how an amalgamation of different forms of TE can ameliorate this process: "Users need to be able to enter text when they're on the go, encumbered, lack access to a full-size keyboard, and in other challenging situations [by] integrat[ing] speech, typing, and gesturing. Such an approach can provide higher performance without forcing users to invest considerable time in learning new [TE] techniques." These findings imply an apt method to face different challenges of TE, regardless of the use situation. Among these challenges lies the need to facilitate TE for those with disabilities. Issues with accessibility and inclusion demonstrate how traditional TE methods may be challenging for individuals with disabilities or who have difficulty using traditional input devices. For example, individuals with visual impairments may have difficulty typing on a physical keyboard or touchscreen, while those with motor impairments may have difficulty using a mouse or a keyboard. For those visually and motor impaired, research involving different methods of TE have been published in the past decade, with the aim of facilitating additional research [16]. Efficiently so, recent work has emerged presenting new methods relying on single-switch TE for those that are motor-impaired. For instance, Ticker [17] is an audio-based single-switch TE method catering to non-speaking individuals with motor disabilities. It has proved to be a successful model to hone down on improved accessible TE methods. Lastly, it is worth noting that disabilities are not always permanent: they may also be temporary or situational [18, pp. 41–42]. TE should be able to cater for those scenarios, and users should be able to switch, combine or alternate between input modes, and adapt quickly. For example, in the case of a temporary disability, such as when someone drives a car, and thus, manual and gaze input are less available than speech, we realize how crucial modality choice can be [1, p. 410].

4. Speech Recognition as a Text-entry Method in Teaching and Learning

Researchers have long been interested in SR as an attractive text-input mode for an array of applications. In educational contexts, in particular, SR technology has been shown to offer transformative benefits, revolutionizing the way in which students access and engage with textual content, enhancing their learning experiences. Several studies have found that SR may hold strong potential as a tool for students of all ages with reading and writing difficulties [3], [4], [19]. However, some discrepancies exist on how effective SR is regarding reading comprehension for those with reading and language impairments. Regardless, language instructors in university sectors may find comfort in the fact that SR has helped students improve language fluency. According to Chen [20]'s study on using SR to teach English, "the [speech] app tasks that provide repetitive training of spoken English, particularly on pronunciation, fluency, and vocabulary acquisition, are more beneficial to learners than traditional teaching methods." While not all languages have the same educational trajectory as English, educators may find that rethinking TE can open up new possibilities for innovation and creativity in academic environments. These can include speech-based multimodal interfaces that cater not just to the learning needs of students, but also the working needs of teachers and staff, improving "support for human cognition and performance […] [since o]ne major goal of such interfaces is to reduce cognitive load and improve communicative and ideational fluency" [1, pp. 422–423].

5. A Vision for Future Explorations

In our effort to understand and rethink TE beyond the English language and for an array of users and use cases, we are at the early stages of planning a multi-stage mixed-method study involving speakers of French, Spanish and Arabic.

Speech-based multimodal interaction in different languages and for various specific tasks has been studied in recent years [21], [22], [23], [24]. Zapata [21], for instance, conducted an experiment with 14 French-speaking professional translators.

Participants performed transcription and translation tasks using conventional TE methods (i.e., keyboard and mouse) and speech-and-touch interfaces. Some of the tasks involved reading out loud a text in French and transcribing the same text. It was observed that reading times were comparable among participants, and that typing took from nearly three to over six times longer. It was argued that a person's typing skills (speed and accuracy) can have a significant impact on their performance when carrying out linguistic tasks such as transcribing or translating, and that French can be seen as not fit for typing, given the wide gap there is between the number of keystrokes needed to transcribe a text and the number of phonemes involved in dictating it. (As an example, to type "*Monsieur Vigneault*," no fewer than 20 keystrokes are needed, whereas the utterance contains 9 phonemes: /mə.sjø vi.ɲo/). This suggests that SR technology has the potential to level the playing field, offering a more inclusive approach to TE by minimizing the disparities caused by varying typing skills and physical abilities, and the particularities of certain languages, thereby enabling a broader range of users to perform linguistic tasks effectively.

Our pre-pilot experiments to this day have also led us to similar conclusions. We recently conducted an informal experiment, collecting transcription and reading-out-loud data from a handful of speakers in French, Spanish and Arabic from our entourage. For instance, in Spanish, we noted that our 200-word text requires a minimum of 1,262 keystrokes (i.e., for anyone with 100% typing accuracy, which is rare: the average is 95%). Depending on the person's typing skills, transcription task times can be between 2.5 and 6 times higher than reading out loud. We also categorized different types of keystrokes and correlated this to the technical effort needed to type a text. In French, for instance, the 200-word text used in the experiment requires 1,334 keystrokes distributed as follows: 1082 correspond to letters (18 of which correspond to 9 special letters (e.g. *â*) or capital letters, which require two keystrokes each), 25 correspond to one-keystroke punctuation marks, 24 to two-keystroke punctuation marks, 200 to hitting the space bar, and 3 to hitting the Enter key to move to a new paragraph. The phoneme count was 775 for the same text.

For our main study, the objectives are to:

- Investigate the current state of TE in major world languages. Along with a deeper literature review, we will conduct a series of controlled experiments to collect quantitative and qualitative TE data from French-, Spanish- and Arabic-speaking members of our university community, including professors, students and staff (5–10 participants per language). We will also collect English-language data for comparison purposes.
- Analyze characteristics of languages that affect TE, and impact the physical or cognitive effort put into typing or dictating with SR: e.g., some languages have words with letters that are written but not pronounced (as noted above); some have syntactic or grammatical structures that require anticipation in dictation, etc.
- Provide recommendations for future research in multimodal TE, SR and dictation training.

Regarding the first bullet point above, we expect to collect data both in-lab and in-the-wild. Lab experiments analyze user behavior during TE, focusing on typing speed and error rates through specific tasks. A mixed-method data-collection approach will allow for active observation of users with video and screen recordings, input logging using Translog II [25] and InputLog [26], and semi-structured interviews. Now, while effective for detailed observation, lab-based experiments do miss the complexities of real-world scenarios. Thus, as Simou and Komninos [27] argue, there is also a need for in-the-wild experiments, where users perform TE in their everyday environments with their own devices, capturing a broader range of influencing factors.

In studying the usability and quality-in-use [28] of TE devices and applications, we will adapt to the working models proposed by Hornbæk [29] and Zapata [21], which include objective measures such as expert assessment of task outcomes, error rates, task completion times, input rates, deviation from an optimal solution, etc.; as well as subjective measures such as users' perception of outcome, subjectively experienced time, perception of task difficulty, users' sense of success, etc. It will also be important to consider and analyze the correlations between the different aspects of usability (i.e., effectiveness, efficiency and user satisfaction), as well as objective and subjective measures. One of the reasons to consider both types of measures is that they may lead to different conclusions regarding usability. For instance, studies have proposed as a usability measure the ratio between objective time and subjectively experienced duration. Psychologists have long recognized and quantified the difference between these two measures; therefore, using both may lead to a better picture of usability and of the user experience, since "differences between interfaces in objective time may not be found for subjectively experienced duration, and vice versa" [29, p. 92].

6. Conclusions

In the rapidly evolving landscape of TE methods, innovation is crucial to addressing the manifold challenges posed by linguistic diversity, regional scripts, and digital accessibility. Tailored solutions, like speech-based MI, may offer a versatile path forward by integrating speech, typing, gesturing, etc. This approach may provide users with adaptable options across various communication contexts, improving efficiency and convenience. Additionally, addressing the needs of individuals with disabilities is paramount, with innovative approaches showing promise in enhancing accessibility. The adoption of SR technology in education also presents opportunities to support students with reading and writing difficulties, potentially benefiting language fluency and overall learning outcomes.

We believe that the fields of SR and MI can contribute tremendously to the advancement of TE research efforts and have a positive impact on the world's linguistic heritage, and on humankind, since novel (speech-based) multimodal interfaces "literally reflect and acknowledge the existence of human users, empower them in new ways, and create for them a 'voice.' [...] They represent [...] a sociopolitical statement about our collective desire to humanize the technology we create." [1, p. 423].

7. Acknowledgements

We would like to gratefully acknowledge the financial support of the Office of the Dean of Arts, Toronto Metropolitan University, which allowed us to conduct and present this exploratory research.

8. References

[1] S. Oviatt, "Multimodal Interfaces," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, 3rd ed., J. A. Jacko, Ed., New York: Lawrence Erlbaum Associates, 2012, pp. 415–429.

[2] M. D. Dunlop, W. Stuerzlinger, A. S. Arif, A. Komninos, and O. Alharbi, "TEXT2030 - Shaping Text Entry Research in 2030," in *Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA: ACM, Sep. 2022, pp. 1–4. doi: 10.1145/3528575.3551429.

[3] J. L. Keelor, N. A. Creaghead, N. H. Silbert, A. D. Breit, and T. Horowitz-Kraus, "Impact of text-to-speech features on the reading comprehension of children with reading and language difficulties," *Ann Dyslexia*, vol. 73, no. 3, pp. 469–486, 2023, doi: 10.1007/s11881-023-00281-9.

[4] S. Kraft, "Revisions in written composition: Introducing speech-to-text to children with reading and writing difficulties," *Front Educ (Lausanne)*, vol. 8, Mar. 2023, doi: 10.3389/feduc.2023.1133930.

[5] L. S. Sakr, *Arabic Glitch: Technoculture, Data Bodies, and Archives.* Statford, CA: Stanford University Press, 2023.

[6] M. A. Yaghan, ""'Arabizi": A Contemporary Style of Arabic Slang.,'" *Design Issues*, vol. 24, no. 2, pp. 39–52, 2008, Accessed: Jul. 07, 2024. [Online]. Available: http://www.jstor.org/stable/25224166

[7] M. Warschauer, G. R. El Said, and A. G. Zohry, "Language Choice Online: Globalization and Identity in Egypt," *Journal of Computer-Mediated Communication*, vol. 7, no. 4, pp. 0–0, Jun. 2006, doi: 10.1111/j.1083-6101.2002.tb00157.x.

[8] B. Esfahbod, M. Hajizadeh, N. Tounsi, R. Ishida, S. Afshar, and T. Nemeth, "Arabic & Persian Layout Requirements," W3C Group Draft Note.

[9] N. Alswaidan, M. I. Hosny, and A. B. Najjar, "A Genetic Algorithm Approach for Optimizing a Single-Finger Arabic Keyboard Layout," 2015, pp. 261–277. doi: 10.1007/978-3-319-14654-6_16.

[10] S. Oviatt, "Interfaces for Thinkers : Computer Input Capabilities that Support Inferential Reasoning," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, pp. 221–228.

[11] K. Go and Y. Kinoshita, "Katakana EdgeWrite: An EdgeWrite Version for Japanese Text Entry," *IEICE Trans Inf Syst*, vol. E97.D, no. 8, pp. 2053–2054, 2014, doi: 10.1587/transinf.E97.D.2053.

[12] S. Janpinijrut, C. Nattee, and P. Kayasith, "An Approach for Improving Thai Text Entry on Touch Screen Mobile Phones Based on Distance and Statistical Language Model," in *Knowledge, Information, and Creativity Support Systems*, T. Theeramunkong, S. Kunifuji, V. Sornlertlamvanich, and C. Nattee, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 44–55.

[13] M. Böcker, K. I. Larsson, and B. von Niman, "Increasing the Usability of Text Entry in Mobile Devices for European Languages and Languages Used in Europe," in *Usability and Internationalization. Global and Local User Interfaces*, N. Aykin, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 13–21.

[14] V. Dagienė, G. Grigas, and T. Jevsikova, "Text Input on Mobile Devices from Cultural and Educational Aspects," *Informatica*, vol. 25, no. 3, pp. 415–424, Jan. 2014, doi: 10.15388/Informatica.2014.21.

[15] P. O. Kristensson, "Next-Generation Text Entry," *Computer (Long Beach Calif)*, vol. 48, no. 7, pp. 84–87, Jul. 2015, doi: 10.1109/MC.2015.185.

[16] O. Polacek, A. J. Sporka, and P. Slavik, "Text input for motor-impaired people," *Univers Access Inf Soc*, vol. 16, no. 1, pp. 51–72, Mar. 2017, doi: 10.1007/s10209-015-0433-0.

[17] E.-M. Nel, P. O. Kristensson, and D. J. C. MacKay, "Ticker: An Adaptive Single-Switch Text Entry Method for Visually Impaired Users," *IEEE Trans Pattern Anal Mach Intell*, vol. 41, no. 11, pp. 2756–2769, Nov. 2019, doi: 10.1109/TPAMI.2018.2865897.

[18] A. Shum *et al.*, "Inclusive Microsoft Design," 2016. Accessed: Jul. 10, 2024. [Online].

[19] L. P. Bruno, A. M. Lewis, E. R. Kaldenberg, P. A. Bahr, and J. Immerfall, "Direct Instruction of Text-to-Speech Software for Students with Intellectual Disability," *Educ Train Autism Dev Disabil*, vol. 55, no. 4, pp. 424–437, 2020.

[20] K. T. C. Chen, "Speech-to-text recognition in University English as a Foreign Language Learning," *Educ Inf Technol (Dordr)*, vol. 27, no. 7, pp. 9857–9875, Aug. 2022, doi: 10.1007/s10639-022-11016-5.

[21] J. Zapata, "Translators in the Loop: Observing and Analyzing the Translator Experience with Multimodal Interfaces for Interactive Translation Dictation Environment Design," PhD thesis, University of Ottawa, 2016. Accessed: Mar. 19, 2023. [Online].

[22] N. Herbig, "Multi-modal post-editing of machine translation," PhD Thesis, Saarländische Universitäts-und Landesbibliothek, Saarbrücken, 2022.

[23] Ž. Radić, S. Braun, and E. Davatti, "Introducing Speech Recognition in Non-live Subtitling to Enhance the Subtitler Experience," in *Proceedings of the International Conference HiT-IT 2023*, 2023, pp. 167–176.

[24] K. Giancola, "Orality in Post-Editing: Methodological Insights," [Poster presentation]. Mardi des chercheurs, Mons, Belgium. 2024.

[25] M. Carl, "Translog - II: a Program for Recording User Activity Data for Empirical Reading and Writing Research," in *Proceedings of the Eight International Conference on Language Resources and Evaluation*, 2012, pp. 4108–4112.

[26] M. Leijten and L. Van Waes, "Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes," *Writ Commun*, vol. 30, no. 3, pp. 358–392, 2013, Accessed: Aug. 21, 2014. [Online].

[27] I. Simou and A. Komninos, "Modelling human behaviour during text entry," in *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction*, New York, NY, USA: ACM, Sep. 2023, pp. 1–3. doi: 10.1145/3565066.3609739.

[28] N. Bevan, "Measuring Usability as Quality of Use," *Software Quality Journal*, vol. 4, pp. 115–150, 1995.

[29] K. Hornbæk, "Current practice in measuring usability: Challenges to usability studies and research," *Int J Hum Comput Stud*, vol. 64, no. 2, pp. 79–102, Feb. 2006, doi: 10.1016/j.ijhcs.2005.06.002.