MaTIAS: Machine Translation to Inform Asylum Seekers

Lieve Macken, Ella van Hest, Arda Tezcan, Michaël Lumingu, Katrijn Maryns and July De Wilde
Department of Translation, Interpreting and Communication
Ghent University
Belgium
{firstname.lastname}@ugent.be

Abstract

This project aims to develop a multilingual notification system for asylum reception centres in Belgium using machine translation. The system will allow staff to communicate practical messages to residents in their own language. Ethnographically inspired fieldwork is being conducted in reception centres to understand current communication practices and ensure that the technology meets user needs. The quality and suitability of machine translation will be evaluated for three MT systems supporting all target languages. Automatic and manual evaluation methods will be used to assess translation quality, and terms of use, privacy and data protection conditions will be analysed.

1 Project overview

Machine translation plays a key role in contexts of migration (Valdez et al., 2023; Vieira et al., 2021). Known problems with the use of MT in migration settings are related to translation quality, lack of domain-specific vocabulary and privacy concerns (Liebling et al., 2020).

This project aims to develop a prototype of a multilingual notification system tailored to asylum reception facilities. The project is being conducted in collaboration with Fedasil, the federal agency responsible for the provision of asylum reception in Belgium. The project is funded by AMIF, the EU Asylum Migration and Integration Fund, and started in July 2023 and will end in December 2025.

The notification system will allow reception centre staff to convey practical messages and instructions to centre residents in the latter's own languages. In doing so, the project aims to support the reception facilities in two main needs: rapid communication of practical information with residents (e.g. "the teacher is sick so there will be no Dutch classes today") and language support for minority languages.

The prototype will consist of a web platform that Fedasil employees can use to translate English, French or Dutch text messages into a set of at least 14 languages[1], including low-resourced languages such as Pashto, Somali and Tigrinya. The translations are either retrieved from a context-specific translation memory or generated by a machine translation (MT) system. These translations are then automatically sent using an existing messaging system (e.g. WhatsApp, Signal, Telegram).

2 Project phases

MaTIAS is an interdisciplinary project that combines methods from linguistic ethnography and translation technology research. The project is divided into four phases: (1) practice-oriented research comprising ethnographic fieldwork and the evaluation and selection of suitable MT systems; (2) content and product development; (3) training; and (4) user evaluation and satisfaction research. The activities and preliminary results of the first phase of the project are detailed below.

3 Ethnographic fieldwork

The ethnographic fieldwork at the reception centres aims to (1) understand the current communication practices and to make an inventory of

[1] All languages available at https://www.fedasilinfo.be/

the most common messages (announcements, instructions) in the reception facilities. This inventory, which consists of frequent phrases and their translations, forms one input source of the context-specific translation memory; (2) gain an insight into the necessary preconditions to optimally use the developed technology (e.g. required expertise among users, attitudes and trust among users, receptivity towards this technology, IT infrastructure, user security, etc.); and (3) gain an understanding of residents' communication preferences to determine which messaging system is best suited to link to the web platform.

So far, the fieldwork has yielded important insights on users' expectations and potential usage obstacles. Firstly, the participants have high expectations on translation accuracy. The staff believe the notification system will facilitate their communication practices significantly. Residents, on the other hand, expressed their support for the system but would prefer technology that allows two-way messaging. Secondly, low literacy in some residents has been identified as a major obstacle for the reception of messages in written form. Therefore, we will explore possibilities of using text-to-speech settings to enable the read-aloud function on different types of smartphones. Finally, the fieldwork data highlights the benefits of incorporating the source language text into the message, rather than simply presenting the translated version. This practice not only serves as an invaluable resource for residents seeking more detailed information from staff, but also provides an excellent opportunity for residents to familiarise themselves with the local language.

4 MT evaluation

We will test the quality and suitability of three different MT systems that support all target languages: two commercial engines (ModernMT and Google Translate), which both allow for some degree of customisation and one open-source model (Meta AI's *No Language Left Behind*-model). For each of these systems, we will also evaluate the impact of the source language on translation quality, in particular English versus Dutch or French for translation into the minority languages.

We will use both automatic and manual evaluation methods to assess MT quality. As a first step, MT quality will be assessed using reference-based automatic evaluation metrics such as BLEU, TER, chrF, BLEURT, BERTscore and COMET. Translation memories linked to www.fedasilinfo.be were obtained from Fedasil for all languages except Dutch, French, German and Spanish, from which we extracted reference translations. In addition, to create a test set, we identified 43 web pages from www.fedasilinfo.be that were most relevant to our project[2] and extracted and sentence aligned all textual information for the missing languages.

From this set, we collected all sentences for which translations are available in all 14 languages. The resulting test set of 577 sentences will be used for all automatic evaluations. Based on the results of the automatic evaluations, a decision will be taken as to which MT systems will be manually evaluated by language experts.

In addition to assessing the MT quality, a thorough analysis of the terms of use for the three different MT systems will be conducted to gain a comprehensive understanding of the privacy and data protection conditions. Furthermore, factors such as deployment features and the ease of integration into the web platform will be taken into consideration. Based on the results obtained from these evaluations, a final decision on which engines to integrate will be made in consultation with Fedasil.

Acknowledgements: This project is co-financed by the European Commission under the Asylum, Migration and Integration Fund (AMIF 093-133).

References

Liebling, Daniel J, Michal Lahav, Abigail Evans, Aaron Donsbach, Jess Holbrook, Boris Smus, and Lindsey Boran. 2020. Unmet needs and opportunities for mobile translation AI. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13.

Valdez, Susana, Ana Guerberof Arenas, and Kars Ligtenberg. 2023. Migrant communities living in the Netherlands and their use of MT in healthcare settings. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 325–334, Tampere, Finland, June. European Association for Machine Translation.

Vieira, Lucas Nunes, Minako O'Hagan, and Carol O'Sullivan. 2021. Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, 24(11):1515–1532.

[2] e.g. https://www.fedasilinfo.be/en/your-stay-reception-place