

INMIGRA3: building a case for NGOs and NMT

Celia Rico*
Universidad Europea de
Madrid

María del Mar Sánchez Ramos**
Universidad de Alcalá

Antoni Oliver***
Universitat Oberta
de Catalunya

*celia.rico@universidadeuropea.es

**mar.sanchezr@uah.es

***aoliverg@uoc.edu

Abstract

INMIGRA3 is a three-year project that builds on the work of two previous initiatives: INMIGRA2-CM¹ and CRISIS-MT². Together, they address the specific needs of NGOs in multilingual settings with a particular interest in migratory contexts.

Work on INMIGRA3 concentrates in the analysis of how to use NMT for the purposes of translating NGOs documentation.

1 Translation needs of non-governmental organisations

The third sector is experiencing an increasing relevance as the number of people in vulnerable circumstances grows. Natural catastrophes, wars, political and religious persecutions, or economic crisis are some of the conditions that leave people unprotected. These are situations that pose a challenge as complex linguistic situations arise (Federici and O'Brien, 2020). And migration flows are not an exception.

Previous research has revealed a series of gaps still to be filled if we are to understand the true nature of multilingual needs in not-for-profit settings. For instance, which are their working conditions as related to multilingual needs? How technology can be best put to use? (see, for

instance, the work of INTERACT project³, *Language on the Move*⁴ or *Translators without Borders*⁵).

In the case of NGOs working in the migratory context, one the main issues is that usually the budget allocated to translation resources is scarce or non-existent. This might explain why catering for the multilingual needs of migrant population is not considered among their core activities—at least in the minds of official donors (Footitt, Crack and Tesseur, 2018). Consequently, translation is mostly conducted as volunteer work and using *ad hoc* materials and tools. In the case of MT, volunteers mostly use free online engines (Rico, 2020). This involves a high risk when dealing with confidential and personal data such as donors' information, reports to official bodies regarding field actions or personal documentation from most vulnerable people.

2 Building a case for the use of neural machine translation

INMIGRA3 aims at building a case for the use of NMT for the specific translation needs of NGOs. This involves an experimental setting along the following lines:

- The participation of two NGOs working with migrant people and refugees in Spain: *Cáritas Española*⁶ and the Spanish Committee for Refugee Help (CEAR)⁷.

¹ The work of INMIGRA2-CM was presented at EAMT 2017: https://ufal.mff.cuni.cz/eamt2017/user-project-product-papers/Conference_Booklet_EAMT2017.pdf

² CRISIS-MT is a project funded by Universidad de Alcalá (CCG2018/HUM-043). Among other objectives, it aims at designing an MT system that can be easily put to use in multilingual crisis communication.

³ INTERACT website:

<https://sites.google.com/view/crisistranslation/home?authuser=0>

⁴ Language on the move website:

<https://www.languageonthemove.com/>

⁵ TWB website: <https://translatorswithoutborders.org/>

⁶ *Cáritas Española* is the Spanish chapter of *Caritas Internationalis*, a not-for-profit organization associated to the Roman Catholic Church.

⁷ CEAR's website: <https://www.cear.es/>

- Definition of a use-case scenario according to the specific needs of the two NGOs participating in the project. This involves different text typologies and topics: donors' funding information, economic reports to official bodies, documentation for asylum petition, administrative forms and instructions for housing benefits, workshop materials in food safety and health issues, tax waiver forms, and informative texts on access to education. The source language of all texts is Spanish and target languages are English, French, Russian, Arabic and Chinese⁸.
- A compact NMT engine developed at the Universitat Oberta de Catalunya that can be used offline in a regular consumer computer. As demonstrated in Oliver et al (2019) it is possible to develop neural MT systems that can be integrated in a very compact set of applications that work in computers with limited hardware resources and without Internet connection. In our implementation, we are using the Marian toolkit with an s2s architecture. The subword-nmt algorithm (Sennrich, 2016) has been applied to both sides of the corpus to minimize the effect of out-of-vocabulary words. As a training corpus, we use MultiUN corpus.
- The evaluation of the MT output in terms of text usability (Suojanen et al, 2015).
- The publication of a best practice report with specific recommendations for NGOs on how to implement NMT. This will include specifications on how these systems can be adapted to the needs of a given situation (language pairs, subjects, etc.), and how to implement them in a regular consumer computer to provide a useful solution to communication needs in migratory contexts.

References

Federici, F.M. and O'Brien, S. 2020, *Translation in Cascading Crisis*. London: Routledge.

Footitt, HA, Crack, A & Tesseur, W. 2018, *Respecting Communities in International Development: Languages and Cultural Understanding*.

<https://www.intrac.org/resources/respecting-communities-international-development-languages-cultural-understanding/>

Oliver, A., M. Sánchez Ramos and C. Rico. Building offline, compact, ready to use MT systems for crisis translation. *Workshop on Crisis Machine Translation ("Crisis MT") at MT Summit 2019*

<https://sites.google.com/view/crisistranslation/crisis-mt-workshop>

Rico, C. 2020. Mapping translation technology and the multilingual needs of NGOs along the aid chain. In Federici, F.M. and O'Brien, S. 2020, *Translation in Cascading Crisis*. London: Routledge.

Sennrich, R. Barry Haddow and Alexandra Birch (2016): Neural Machine Translation of Rare Words with Subword Units *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*. Berlin, Germany.

Suojanen, T., Koskinen, K. and Tuominen, T. 2015. *User-Centered Translation*. London and New York: Routledge.

Acknowledgements

INMIGRA3 is a project jointly funded by the Autonomous Community of Madrid (Spain) and the European Social Fund under grant H2019/HUM-5772 (start date: 1 Jan 2020; end date: 31 Dec 2023).

⁸ We are aware that these languages do not cover the full spectrum of languages spoken by migrant people in Spain. Less resourced languages are not the specific subject of this project as the participating NGOs issue their content in the language combinations mentioned.