

INTERACT Deliverable 8.3: Panel on Crisis MT

The INTERACT project has received funding from the European Union's Horizon 2020 Research and Innovation programme under the Marie Skłodowska-Curie grant agreement No 734211.

Building offline, compact, ready to use MT systems for crisis translation

Antoni Oliver (Universitat Oberta de Catalunya, Barcelona)
María del Mar Sánchez Ramos (Universidad de Alcalá, FITISPos-UAH, Madrid, Spain)
Celia Rico (Universidad Europea, Madrid, Spain)



Crisis Machine Translation Workshop, MTSummit 2019, Dublin, 20th August

- Project funded by Universidad of Alcalá (Madrid, Spain), 2018-2019
- Research on multilingual crisis communication and the possible benefits (and challenges) of using technology (MT) in the third social sector
- Main objectives:
 1. Describing Spanish policies for action in multilingual communication in crisis situations
 2. Identifying translation needs in crisis situations within Spanish organizations offering international cooperative services
 3. Describing translators' perceptions towards MT in the migratory context
 4. Building a customized MT system to be used in multilingual crisis communication settings

Challenges in building MT in crisis settings

Where is MT in the aid chain?

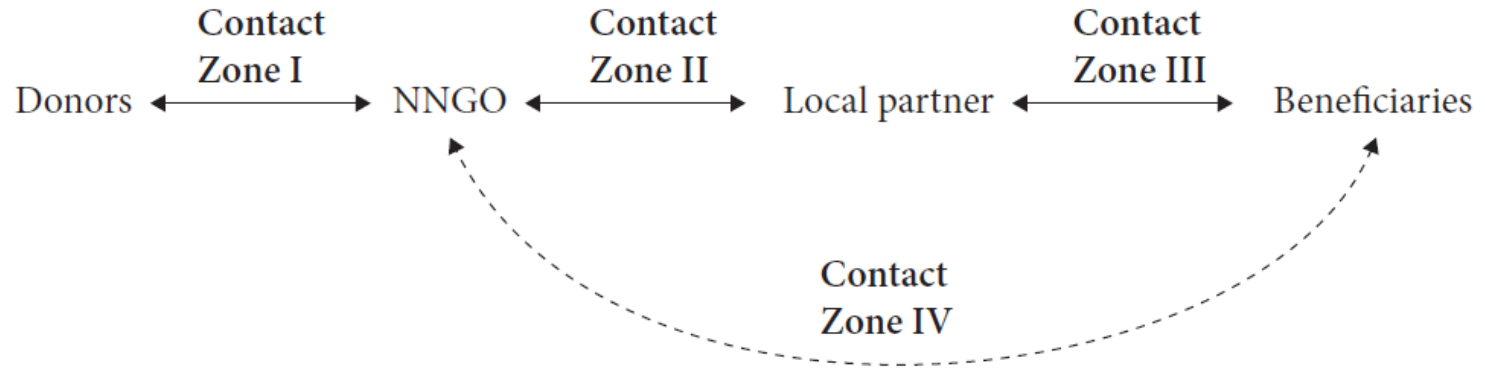


Figure 1. The generic aid chain and its contact zones (Delgado Luchner 2018)

Challenges in building MT in crisis settings

- Unpredictable nature of emergencies
- Emergencies often occur in areas where languages are low resourced
- Cost, time and infrastructure restrictions

How is MT approached then when hardware requirements are relatively high, especially for training neural systems?

Hardware requirements much lower for translation

MT Toolkits

Moses and Marian:

- Mature enough and used in real commercial applications
- Permissive free license
- Can be compiled in compact applications with almost no dependencies
- Easy distribution

Corpus - MultiUN

	segments	L1 tokens	L2 tokens
eng-spa	9,2 M	254 M	300 M
fra-spa	11,4 M	333 M	335 M
rus-spa	10,6 M	228 M	308 M
zho-spa	9,8 M	260 M	288 M

Trained MT-Systems

- **Moses (SMT)** (language model of order 5; alignment using mgiza with grow-diag-final-and.)
 - eng-spa - spa-eng
 - fra-spa - spa-fra
 - rus-spa - spa-rus
 - zho-spa
- **Marian (NMT)** (RNN-based encoder-decoder model with attention mechanism (s2s), layer normalization, tied embeddings, deep encoders of depth 4, residual connectors and LSTM cells; subword-nmt algorithm)
 - eng-spa - spa-eng
 - fra-spa - spa-fra
 - rus-spa - spa-rus
 - zho-spa - spa-zho

Automatic Evaluation

ENG-SPA	BLEU	NIST	WER
Marian	0.4547	9.262	0.4427
Moses	0.4906	9.483	0.4672
Google Translate	0.4183	8.672	5.5318

SPA-ENG	BLEU	NIST	WER
Marian	0.4871	9.65	0.4301
Moses	0.4647	9.074	0.4999
Google Translate	0.4362	8.929	0.5165

Automatic Evaluation

FRA-SPA	BLEU	NIST	WER
Marian	0.4132	8.607	0.492
Moses	0.4544	8.927	0.512
Google Translate	0.3626	7.927	0.5882

SPA-FRA	BLEU	NIST	WER
Marian	0.4204	8.784	0.4964
Moses	0.4629	8.942	0.5032
Google Translate	0.3565	7.841	0.5922

Automatic Evaluation

RUS-SPA	BLEU	NIST	WER
Marian	0.3746	7.943	0.5413
Moses	0.4409	8.665	0.5685
Google Translate	0.3346	7.386	0.6613

SPA-RUS	BLEU	NIST	WER
Marian	0.3254	7.189	0.597
Moses	0.3747	7.522	0.6293
Google Translate	0.2711	6.359	0.705

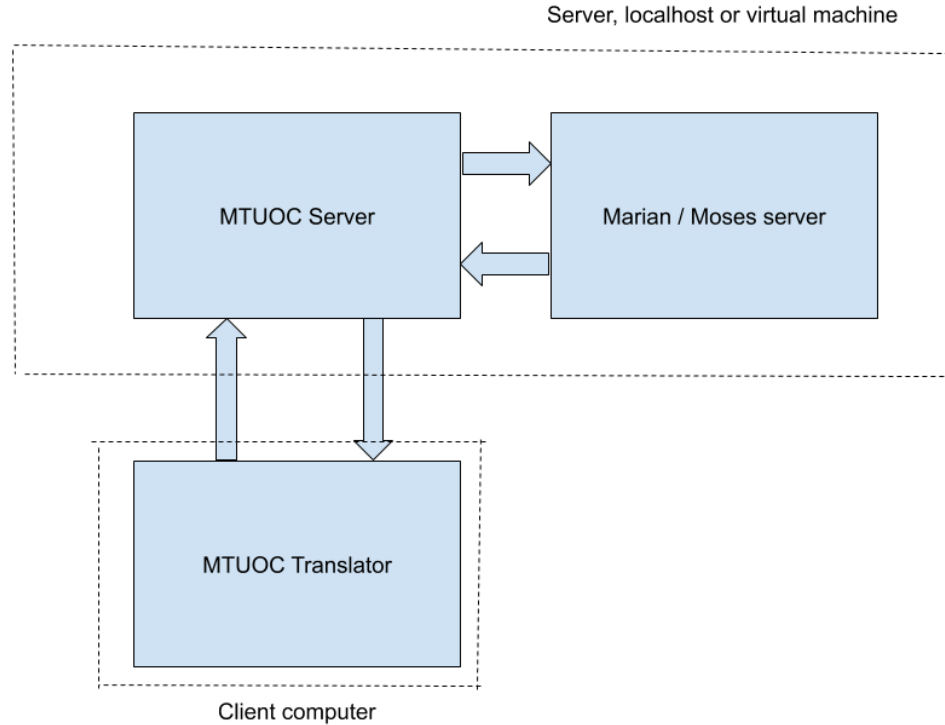
Automatic Evaluation

ZHO-SPA	BLEU	NIST	WER
Marian	0.2304	4.895	0.698
Moses			
Google Translate	0.2396	6.151	0.79

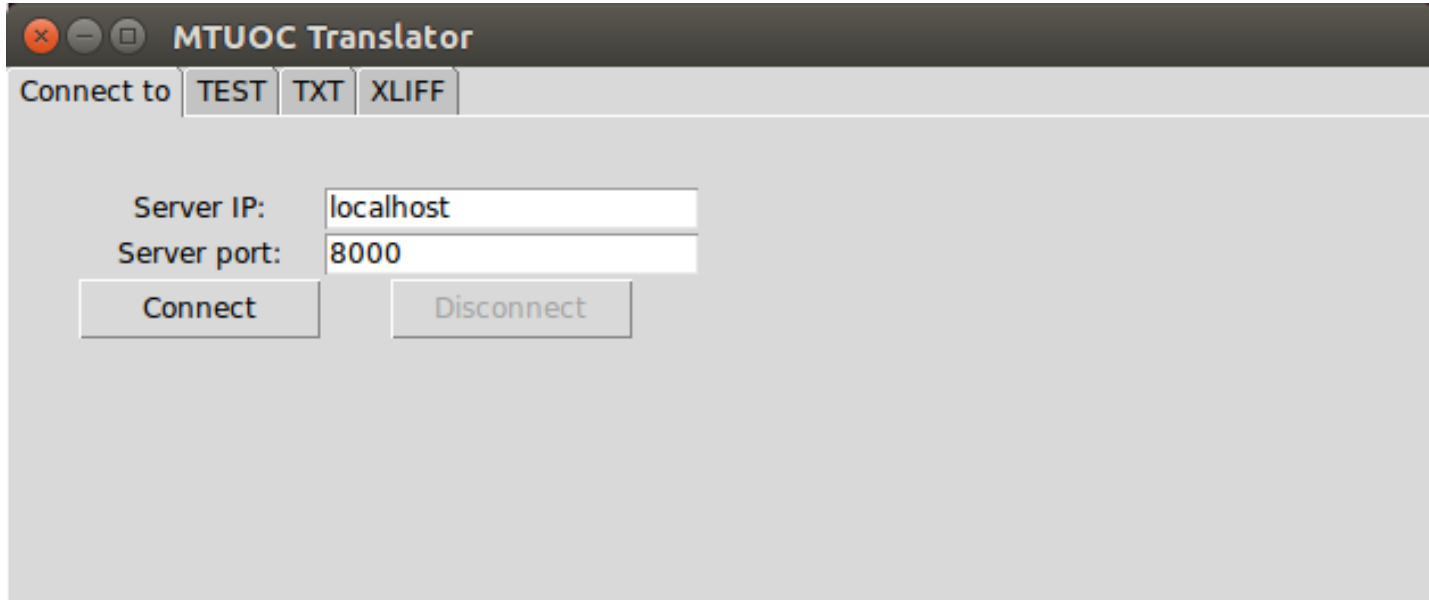
SPA-ZHO	BLEU	NIST	WER
Marian	0.279	7.167	0.6837
Moses			
Google Translate	0.1643	6.024	0.8738

Distribution

- <http://lpg.uoc.edu/MTUOC/>
- 2 components:
 - Server
 - Moses/Marian
 - MTUOC
 - Client



Client - MTUOC Translator



The image shows a screenshot of a software window titled "MTUOC Translator". The window has a dark gray title bar with standard window control buttons (close, minimize, maximize). Below the title bar, there is a tabbed interface with three tabs: "Connect to", "TEST", "TXT", and "XLIFF". The "Connect to" tab is currently selected. Inside this tab, there are two text input fields. The first is labeled "Server IP:" and contains the text "localhost". The second is labeled "Server port:" and contains the text "8000". Below these fields are two buttons: "Connect" and "Disconnect". The "Disconnect" button is currently disabled, indicated by its gray color.

MTUOC Translator

Connect to TEST TXT XLIFF

Server IP: localhost

Server port: 8000

Connect Disconnect

Pros and cons

Pros:

- System can be used in consumer computers
- Linux, Windows or Mac
 - Server - Linux or virtual machine
 - Client - written in Python and Windows exe distributed
- No internet connection required

Cons:

- Systems trained using large parallel corpora. Not available for all languages
- Translation segment by segment is slow

Contribution and further work

Effective communication as an strategic need (Seeger, 2006). Our specific contributions to this strategic need are:

- We offer a replicable model for the development of offline MT.
- We distribute a very compact set of applications in order to work in computers with limited hardware resources and without Internet connection.

As future work:

- How to train systems for low resourced languages
- Unsupervised Machine Translation (as Artetxe's Monoses system)

Thank you very much for your attention!

Antoni Oliver

aoliverg@uoc.edu

Maria del Mar Sánchez

mar.sanchez@uah.es

Celia Rico

celia.rico@universidadeuropea.es