TraMOOC

Translation for Massive Open Online Courses

Dr. Sheila Castilho
Dr. Joss Moorkens
Dr. Federico Gaspari
Prof. Andy Way

Table of contents

- Intro to TraMOOC
- DCU's Work Package
- Comparative Evaluation of Neural MT and Phrase-Based SMT
- Crowdsourcing Evaluation
- Specific Constraints
- Specific Solutions
- What is still needed

Translation for Massive Open Online Courses

- **2015-2018**; ICT-17-2014: Cracking the language barrier

- **Reliable Machine Translation (MT)** for **Massive Open Online Courses (MOOCs)**

- The main expected outcome is a **high-quality machine translation service** for educational text data on a MOOC platform

- Open educational platform for MT and a replicable process for creating such a service

- Make existing monolingual educational material available to speakers of other languages
  - multi-genre and heterogeneous textual course material
  - Subtitles – video lectures
  - assignments
  - tutorial text
  - social web text posted on MOOC blogs and fora (questions/answers/comments)

- Reusing existing linguistic infrastructure and MT resources extending existing models

- Test on a MOOC platform and on the VideoLectures.Net digital video lecture library

The targeted audience

- Users who want access to open online education that is not constrained by language barriers.

- MOOC providers, who wish to offer high-quality, integrated multilingual educational services.

- Machine Translation developers, who need a platform for promoting, testing and comparing their solutions.

- Language Technology Engineers, who want access to accurate and wide-coverage linguistic infrastructure, even for less widely spoken languages.

The Consortium

- 10 partners from 6 European countries
  - Humboldt University (Coordinator)
  - Dublin City University
  - University of Edinburgh
  - Ionian University
  - Radboud University
  - Tilburg University
  - Deluxe Media Europe LTD
  - Knowledge 4 All Foundation LTD
  - EASN Technology Innovation Services
  - (Iversity) Coursera

Activities

- 9 Work Packages
  - o WP1 - Management and Coordination
  - o WP2 - Architecture and Requirements Analysis
  - o WP3 - Data Collection and Infrastructure Exploration/ Adaptation/ Bootstrapping
  - o WP4 - Machine Translation
  - o WP5 - Explicit Translation Evaluation
  - o WP6 - Implicit Translation Evaluation
  - o WP7 - System Integration/Expandability/Updateability
  - o WP8 - System Viability/Exploitation/Commercialization
  - o WP9 - Dissemination and Diffusion

Crowdsourcing

- PBSMT
  - o Moses, MGIZA is used to train word alignments, and KenLM is used for language model training and scoring (Huck and Birch 2015)
- NMT
  - o attentional encoder-decoder networks trained with Nematus (Sennrich et al. 2016)
- Training data:
  - o WMT training data
  - o OPUS
  - o TED from WIT3
  - o QCRI Educational Domain Corpus (QED)
  - o a corpus of Coursera MOOCs
  - o TraMOOC's own collection of educational data

- Domain adaptation:
    - Models initially trained on all available data, then continually trained on in-domain data, which effectively adapts the system to the domain NMT (check Rico's answer)

- Tools Used:
    - Nematus: https://github.com/rsennrich/nematus
    - Amun: https://github.com/amunmt/amunmt (for deploying the models)

- Human and automatic translation evaluation of prototype 1 vs prototype 2 (PBSMT vs NMT)

- Crowdsourcing evaluation prototype 2

- Crowdsourcing evaluation prototype 2 vs prototype 3

NMT vs. PB-SMT

- 4 datasets (250 segments) from real EN MOOC data translated into German, Greek, Portuguese, and Russian

- PB-SMT/NMT mixed, random task order

- 2-4 professional translators

NMT vs. PB-SMT

- Comparative ranking of 100 randomised translations

- Post-editing using PET (Aziz, Castilho, Specia 2012)
  - Temporal effort – time spent post-editing (Krings 2001)
  - Technical effort – edit count
- Rating of fluency and adequacy (1-4 Likert scale)
- Error annotation
  - Inflectional morphology, Word order, Omission, Mistranslation, Addition

NMT/SMT Ranking

EN-EL Evaluations	PB-SMT preference	NMT preference
400	174	226
	43.5%	56.5%

EN-DE Evaluations	PB-SMT preference	NMT preference
300	61	239
	20.3%	79.7%

EN-RU Evaluations	PB-SMT preference	NMT preference
300	110	190
	36.7%	63.3%

EN-PT Evaluations	PB-SMT preference	NMT preference
300	115	185
	38.3%	61.7%

NMT/SMT Fluency

- For all 4 language pairs:

FLUENCY
1. No fluency
2. Little fluency
3. Near native
4. Native

	EN-DE		EN-EL		EN-PT		EN-RU	
% scores assigned 3-4 fluency value (SMT, NMT)	54.2	67.6	65	75	73.8	79.5	60.2	75.1
% scores assigned 1-2 fluency value (SMT, NMT)	45.8	32.4	35	25	26.2	20.5	39.8	24.9

NMT/SMT Adequacy

- For all 4 language pairs:

ADEQUACY
1. None of it
2. Little of it
3. Most of it
4. All of it

	EN-DE		EN-EL		EN-PT		EN-RU	
% scores assigned 3-4 adequacy value (SMT, NMT)	73.5	66.4	89	89	94.7	97.1	72.8	77.5
% scores assigned 1-2 adequacy value (SMT, NMT)	26.5	33.6	11	11	5.3	2.9	27.2	22.5

NMT/SMT PE Temporal Effort

Words per second (all PEs)	SMT	NMT
German	0.21	0.22
Greek	0.22	0.24
Portuguese	0.29	0.30
Russian	0.14	0.14

Previous work by Moorkens & O'Brien (2015) found an average speed of 0.39 WPS for EN-DE professional PE.

SMT, NMT	German		Greek		Portuguese		Russian	
Post-edited sentences (changed)	940	813	928	863	874	844	930	848
Unchanged smt, nmt	60	187	72	137	126	156	70	152

NMT/SMT Error Markup

- Fewer overall errors for all language pairs
- Marked improvement in word order in NMT

	German		Greek		Portuguese		Russian	
	SMT	NMT	SMT	NMT	SMT	NMT	SMT	NMT
Segments without Issues	*61*	*189*	*90*	*168*	*197*	*236*	*101*	*195*
total no. of "Inflectional morphology"	732	608	443	307	404	378	695	506
total no. of "Word Order"	382	180	303	208	216	181	197	122
total no. of "Omission"	126	84	48	57	53	58	194	163
total no. of "Addition"	46	39	24	31	61	44	183	151
total no. of "Mistranslation"	401	323	459	483	348	342	385	404
Total number of issues	**1687**	**1234**	**1277**	**1086**	**1082**	**1003**	**1654**	**1346**

NMT/SMT Summary

In this study, using these language pairs, in this domain…

- Fluency is improved, word order errors are fewer using NMT
- Fewer segments require editing using NMT
- NMT produces fewer morphological errors
- No clear improvement for omission or mistranslation using NMT
- NMT for production: no great improvement in post-editing throughput
  - o "Errors are more difficult to spot"

Constraints

- Time-constraints
- Number of available translators
- Different platform

Crowdsourcing

- Evaluation prototype 2 (NMT)

- Crowdflower Platform
  - To start this month
  - External and Expert Crowd

- Adequacy & Fluency

- Source Evaluation

- Post-editing (expert and crowd): "Please correct words or phrases that are unintelligible, wrong, or ambiguous"
  - Consider how to time PE task for temporal effort

- Change the mark-up error type list (for expert group) so as to map onto DQF-MQM typology: **Addition, Mistranslation, Omission, Untranslated, Function Words, Word Form,** and **Word Order**.

Crowdsourcing

English Sentence

The Essay Question

Translated Sentence

Ερώτηση έκθεσης

Question 1

The English sentence is grammatically:
- ○ Flawless
- ○ Adequate
- ○ Incomprehesible

Question 2

The translated sentence convey the meaning of the original sentence:
- ○ Very well
- ○ Well
- ○ Adequately
- ○ Poorly
- ○ Very poorly

Question 3

The translated sentence is grammatically:
- ○ Flawless
- ○ Good
- ○ Adequate
- ○ Poor
- ○ Incomprehesible

Question 4

Please mark up a word or a sequence of words, from the translated sentence, in order to enable the popup menu, where you can provide a particular error type.

Upon creating an error tag, you can delete it, by clicking on it

Tags

Question 5

Please correct words or phrases that are unintelligible, wrong, or ambiguous.

Ερώτηση έκθεσης

Crowdsourcing

English Sentence

The Essay Question

Translated Sentence

Ερώτηση έκθεσης

Mark-up-selection

έκθεσης

Choose the error type:
- Addition (ADD)
- Mistranslation (MIST)
- Omission (OM)
- Untranslated (UN)
- Function Words (FW)
- Word Form (WF)
- Word Order (WO)

Add

Question 1

The English sentence is grammatically:
- Flawless
- Adequate
- Incomprehesible

Question 2

The translated sentence convey the m... sentence:
- Very well
- Well
- Adequately
- Poorly
- Very poorly

Question 3

The translated sentence is grammatically:
- Flawless
- Good
- Adequate
- Poor
- Incomprehesible

Question 4

Please mark up a word or a sequence o... sentence, in order to enable the po... provide a particular error type.

Upon creating an error tag, you can de...

Tags

Question 5

Please correct words or phrases that are unintelligible, wrong, or ambiguous.

Ερώτηση έκθεσης

Crowdsourcing

TraMOOC
Translation for Massive Open Online Courses

- Prototype 2 vs Prototype 3

Source Sentence:
The Essay Question

Drag each sentence according to the ranking you prefer.

1 Ερώτηση προς ανάπτυξη
2 Ερώτηση έκθεσης
3 Ερώτηση ανάπτυξης

Source Sentence:
The Essay Question

Drag each sentence according to the ranking you prefer.

1 Ερώτηση έκθεσης
2 Ερώτηση προς ανάπτυξη
3 Ερώτηση ανάπτυξης

- Unforeseen delays
- Crowdsourcing contracts
- Change of MOOC partner
  - Delays are part of most academic collaborations

- From on-going Crowdsourcing activity (translation):
  o Malicious behaviour
    - Blank translations
    - Random symbols
    - Repetitive answers
    - Other language characters

  o Use of Google Translate
  o BR performing EU-PT tasks

Crowdsourcing

- Specified Solutions (from on-going translation):
    - Allow copy/paste 5 characters long
    - Increase the minimum time per page
    - Increase contributors level (from 1 to 2)
    - Ban contributors from specific countries
    - Constant monitoring

What is still needed

- Specific set up for each language on the platform
  - Learn from the crowdsourcing translation task

- Test design for Post-editing and evaluation

www.tramooc.eu

Thank you!

Sheila Castilho