

Recent Developments



Joss Moorkens & Yota Georgakopoulou

MT Summit XVI



Table of contents

- The TraMOOC Project
- NMT systems for TraMOOC
- Comparative Evaluation of Neural MT and Phrase-Based SMT
- Crowdsourced evaluations (explicit & implicit)
- Task-based evaluations

Joss Moorkens, Sheila Castilho, Federico Gaspari, Andy Way (DCU/ADAPT) – Ireland
Yota Georgakopoulou, Maria Gialama (Deluxe Media) – Greece/United Kingdom
Rico Sennrich, Antonio Valerio Miceli Barone (University of Edinburgh) - United Kingdom
Valia Kordoni, Markus Egg, Maja Popović (Humboldt University of Berlin) - Germany
Vilemini Sosoni (Ionian University, Corfu) - Greece
Iris Hendrickx (Radboud University Nijmegen) – The Netherlands
Menno van Zaanen (Tilburg University) – The Netherlands



9月2017年

Joss Moorkens & Yota Georgakopoulou



- **Reliable Machine Translation (MT) for Massive Open Online Courses (MOOCs)**
- The main expected outcome is a **high-quality semi-automated 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



9月2017年

Joss Moorkens & Yota Georgakopoulou



- Create domain-specific SMT NMT engines – 3 iterations
- Crowdsourced evaluation of MT quality
- Explicit and implicit evaluation stages
- Task-based evaluations
- Free and premium platform due 2018



9月2017年

Joss Moorkens & Yota Georgakopoulou



- 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

- 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.

- 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) HPI



9月2017年

Joss Moorkens & Yota Georgakopoulou



Which MT paradigm?

- Project had originally planned to compare Syntax-Based and Phrase-Based SMT
- Comparative Evaluation of Neural MT (Nematus) and Phrase-Based SMT (Moses)
- English to German, Greek, Portuguese, and Russian
- MT engines trained on open and educational data



9月2017年

Joss Moorkens & Yota Georgakopoulou



Chinese→English				German→English			
#	Ave %	Ave z	System	#	Ave %	Ave z	System
1	73.2	0.209	SogouKnowing-nmt	1	78.2	0.213	online-B
	73.8	0.208	uedin-nmt		76.6	0.169	online-A
	72.3	0.184	xmunmt		76.6	0.165	KIT
4	69.9	0.113	online-B	76.6	0.162	uedin-nmt	
	70.4	0.109	online-A	75.8	0.131	RWTH-nmt-ensem	
	69.8	0.079	NRC	74.5	0.098	SYSTRAN	
7	67.9	0.023	jhu-nmt	7	72.9	0.029	LIUM-NMT
	66.9	-0.016	afri-mitll-opennmt	8	70.2	-0.058	TALP-UPC
	67.1	-0.026	CASICT-cons		69.8	-0.072	online-G
	65.4	-0.058	ROCMT		68.6	-0.103	C-3MA
11	64.3	-0.107	Oregon-State-Uni-S	11	64.1	-0.260	online-F
12	61.7	-0.209	PROMT-SMT	English→German			
	61.2	-0.265	NMT-Ave-Multi-Cs	#	Ave %	Ave z	System
	60.0	-0.276	UU-HNMT	1	72.9	0.257	LMU-nmt-reranked
	59.6	-0.279	online-F	2	70.2	0.158	online-B
59.3	-0.305	online-G	69.8		0.139	uedin-nmt	
					68.9	0.092	SYSTRAN
					66.9	0.035	LMU-nmt-single
					66.7	0.022	KIT
					66.4	0.015	xmu
					66.6	0.006	LIUM-NMT
				66.0	-0.003	RWTH-nmt-ensem	
English→Chinese							
#	Ave %	Ave z	System				
1	73.2	0.208	SogouKnowing-nmt				
	72.5	0.178	uedin-nmt				
	72.0	0.165	xmunmt				
4	69.8	0.065	online-B				



- Main strength of NMT is grammatical improvements, but possible degradation in lexical transfer (Neubig, Morishita, Nakamura 2015)
- Output conditioned on full source text and target history
- Some problems:
 - Networks have fixed vocabulary → poor translation of rare/unknown words
 - Models are trained on parallel data; how do we use monolingual data?
 - Recent solutions:
 - Subword models allow translation of rare/unknown words (Sennrich, Birch, Haddow 2016a)
 - Train on back-translated monolingual data (Sennrich, Birch, Haddow 2016b)



- 4 datasets (250 segments) from EN MOOC data translated into German, Greek, Portuguese, and Russian using TraMOOC engine prototype 2
- PB-SMT/NMT mixed, random task order
- 2-4 professional translators in Deluxe Media
- **Detailed results presented by Sheila Castilho in Research Track and in proceedings of MT Summit XVI**



9月2017年

Joss Moorkens & Yota Georgakopoulou

11

- PBSMT
 - Moses, MGIZA is used to train word alignments, and KenLM is used for language model training and scoring (Huck and Birch 2015)
- NMT Tools Used:
 - Nematus: <https://github.com/rsennrich/nematus>
 - Amun: <https://github.com/amunmt/amunmt> (for deploying the models)
- 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



9月2017年

Joss Moorkens & Yota Georgakopoulou

12

- 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



9月2017年

Joss Moorkens & Yota Georgakopoulou



- 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



9月2017年

Joss Moorkens & Yota Georgakopoulou



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



9月2017年

Joss Moorkens & Yota Georgakopoulou



- 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
 - “Errors are more difficult to spot”
- Based on the pace of improvement of NMT however, TraMOOC moved to NMT exclusively



9月2017年

Joss Moorkens & Yota Georgakopoulou



Using the Crowdfunder platform for all 11 language pairs:

- *Clear instructions available during the entire translation procedure.*
- *Test Questions to validate the accuracy of the participants' input.*
- *Post-editing question should be displayed first, hiding the rest of the questions to avoid influencing the contributors' judgment.*
- *Fluency for ST and TT, adequacy and error mark-up for TT*
- *Multiple error mark-up supported.*

For QA and language coverage, 5-10% expert evaluation by DME



9月2017年

Joss Moorkens & Yota Georgakopoulou

17

- Post-editing (expert and crowd): “Make changes in the translation if there are errors in grammar, meaning or spelling”
 - Basic rules regarding spelling apply. If there are any typos or slight grammatical/syntactic mistakes in the original, please fix them in the translation
 - Do not implement corrections for stylistic reasons only
 - No need to restructure sentences only to improve the natural flow of the text

English Sentence	Translated Sentence
OK. At a high level, yes.	Добре. На високо ниво, да.
Question 1 Fix errors in grammar, meaning or spelling making as few changes as possible.	
<input type="text" value="Добре. На високо ниво, да."/>	
<input type="button" value="Finished Post-editing"/>	

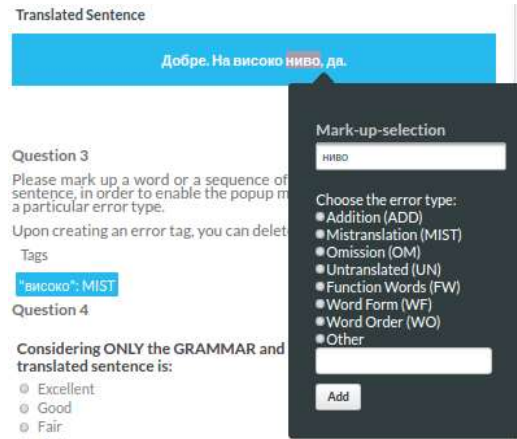


9月2017年

Joss Moorkens & Yota Georgakopoulou

18

- 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**



9月2017年

Joss Moorkens & Yota Georgakopoulou



- Unforeseen delays:
 - Crowdsourcing contracts
 - Change of MOOC partner
- Crowd behaviour issues →

Crowd behaviour issue	Solution(s)
Malicious behaviour	Constant monitoring, manual and automated
Use of Google Translate	Source language text is an image. Manual check with Google Translate feature in Chrome.
BR performing EU-PT tasks	Target specific countries
No change, yet low score on quality	Popup alerts
Poor coverage/ low contributor flow	Increase HIT payment; expand geographical reach & channel; decrease contributor level; decrease text question difficulty



9月2017年

Joss Moorkens & Yota Georgakopoulou



Malicious behaviour	Solutions
Blank translations	Change tactics for test questions, binary evaluation answers, distributed randomly
Random symbols	Increase the minimum time per page
Repetitive answers	Increase contributors' level
Other language characters	Constant manual and script-based (automated) monitoring: Python scripts for blanks, Latin characters in non-Latin languages, etc.
Multiple malicious accounts	Customised alerts scripts (blanks, length, time per page, etc.); flag malicious contributors; ban specific channels



9月2017年

Joss Moorkens & Yota Georgakopoulou



Underway: Crowdsourced implicit evaluations

Implicit evaluation: Annotation of entities, topics and terms in the source and target texts

- Generate a thesaurus of tag-sets that allows for the **implicit evaluation** of MT output through the comparison of the source and target tag-sets

Activities:

1. Entity annotation via Wikification
2. Topic & sentiment annotation

Course Title: Business and Negotiation

Comment

This is such a poor transcript. I've completed a few MOOC courses in the past and they all had transcripts that helped students quickly review materials or search keywords. This doesn't look like a transcript, at least not a useful one, and doesn't look like subtitles either. I can't understand what it is and how could anyone use it. I had to go over and over some lines a few times in order to understand what they are saying. The subject was boring and the text was annoying. Misleading thread title :)

1) Is the topic of this comment mainly about:

The course

The translation

Something else

2) Does the post express a sentiment towards this topic that is:

Positive (happy, excited, enthusiastic, complimentary)

Neutral/I don't know (unrelated comments, mixed feelings)

Negative (hate, anger, sadness, frustration)

Positive Neutral Negative



9月2017年

Joss Moorkens & Yota Georgakopoulou



- openHPI - European MOOC platform plus TraMOOC API
 - Launched by the Hasso Plattner Institute (HPI) for Digital Engineering in Potsdam, Germany
- Users will be able to switch between the original course language and automatically translated content
- Users will be able to request translation for specific forum contributions
- Feedback via surveys on the translation content and the integration of the translation tools into the openHPI platform



9月2017年

Joss Moorkens & Yota Georgakopoulou



ありがとうございました



ありがとう



9月2017年

Joss Moorkens & Yota Georgakopoulou

