Centre for Translation Studies
UNIVERSITY OF SURREY
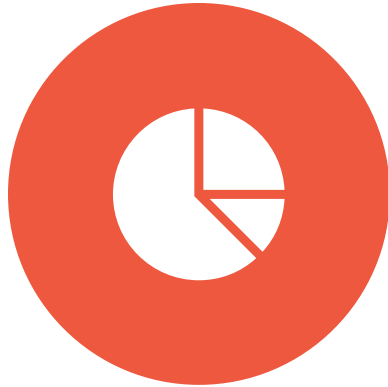
Challenges in Translation of Sentiments and Emotions in User-Generated Content

Constantin Orasan

Centre for Translations Studies
University of Surrey, UK
C.Orasan@surrey.ac.uk

While we are getting ready to start please open a browser and go to pollev.com/corasan432
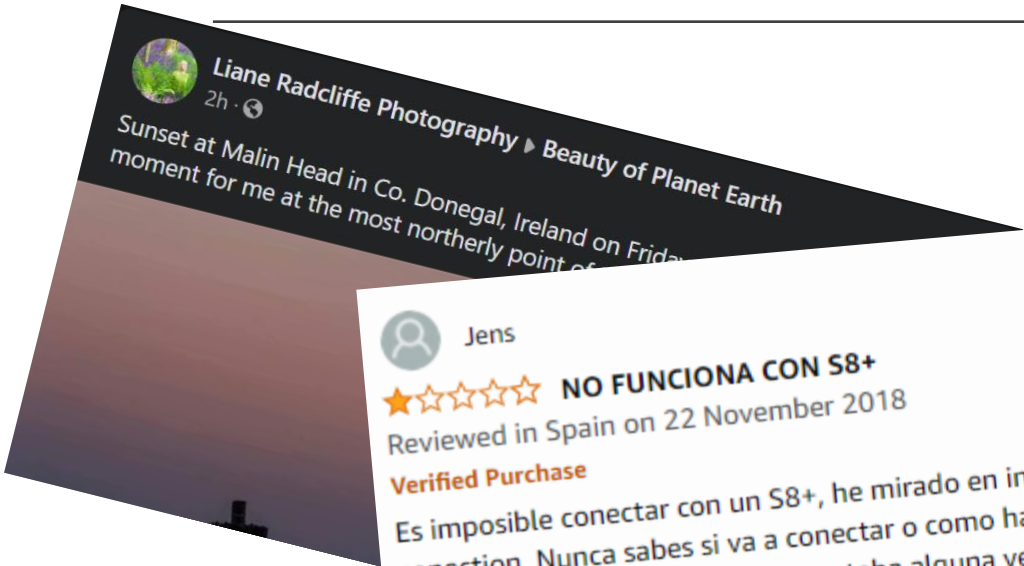
Structure

WHY TRANSLATE UGC

TRANSLATION OF SENTIMENTS

LOOKING FURTHER INTO TRANSLATION OF EMOTIONS AND "MESSY" DATA

What is User Generated Content (UGC)

Liane Radcliffe Photography ▶ Beauty of Planet Earth
2h · 🌐
Sunset at Malin Head in Co. Donegal, Ireland on Frida
moment for me at the most northerly point of

Jens

⭐☆☆☆☆ **NO FUNCIONA CON S8+**
Reviewed in Spain on 22 November 2018
Verified Purchase
Es imposible conectar con un S8+, he mirado en internet y al parece tiene muchisimos defect
conection. Nunca sabes si va a conectar o como hacerle conectar. Necesité instalar y desinsta
veces para empezar, despues conectaba alguna vez, pero ahora solo conecta con el cargado
pena que un producto tiene un fallo tan grande. Yo tenia entendido que esto era un produ
Samsung y aun asi o funciona con uno de los telefones Samsung mas vendidos.

5 people found this helpful

Report abuse

Translate review to English

DomB #FBPE 💙 @LoveEU67 · 2h
I used to give Wetherspoon's credit for its democratic pricing and half-
decent curries. Not now. #Brexiter owner Tim Martin has been instrumental
in creating the labour and supply shortages now affecting him. He deserves
everything he gets. #BrexitChaos

Beer shortages at Wetherspoons after Brexit and Covid hit supply chai...
Pints of Carling and Coors are unavailable in some branches, with the
problem due to a lack of lorry drivers.
🔗 metro.co.uk

💬 23   🔁 53   ♡ 247

Uses of UGC

Take decisions

Assess the mood

Gain new information

But UGC …

Is messy

Not easy to process

At times unreliable

Not always in a language we understand,

… but too much to be translated manually

Very often used to expresses sentiments, emotions, opinions

Have you ever used MT for UGC?

Please go to **pollev.com/corasan432** to answer this question

Why do we care about translating sentiments and emotions in user generated content?

Non-English sentiment analysis

Sentiment analysis is important

Option 1: Develop an ML system from scratch: we need resources (NLP components, word embeddings), annotated datasets

Option 2: Use some language specific lexical resources/rely on an English lexical resource (e.g. SentiWordNet) and work with linguists to produce a lexicon-based system

Option 3: Machine translate the texts into English and use an English sentiment analyser

… there are some other options like zero shot learning

The same approaches can also be used for emotion detection

MT-based sentiment analysis

Input text	Machine translation to English	English sentiment analysis

Very easy to develop because we can reuse existing APIs and in theory we do not care that much about the quality of MT

The performance of the sentiment analyser does not necessarily reduce a lot when MT is used vs a mono-lingual SA, but the size of the reduction is different from one source language to another (4-6% for French and Spanish, but 20% reduction for Japanese) [1]

A MT-based sentiment analyser may perform better than a native SA method [2]

Customised MT engines will lead to better results (the engine used by bookings.com may be a custom one)

[1] Poncelas, A., Lohar, P., Way, A., & Hadley, J. (2020). The Impact of Indirect Machine Translation on Sentiment Classification. *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*, *1*, 78–88. https://aclanthology.org/2020.amta-research.7/
[2] Araújo, M., Pereira, A., & Benevenuto, F. (2020). A comparative study of machine translation for multilingual sentence-level sentiment analysis. *Information Sciences*, *512*, 1078–1102. https://doi.org/10.1016/j.ins.2019.10.031

MT-based sentiment analysis

Sentiment analysis systems rely very much on sentiment bearing words to assign a score, so it's vital to translate them correctly

Small noise results in big changes:

- MT: "IN THE ATTIC OBLIQUE WALLS, THE ROOM VERY NARROW FOR OUR HEIGHT, WE PERMANENTLY LOVED OUR SHOULDERS OR HEAD." 0.79/0.05/0.16

- Correct: "IN THE ATTIC OBLIQUE WALLS, THE ROOM VERY NARROW FOR OUR HEIGHT, WE PERMANENTLY HIT OUR SHOULDERS OR HEAD." 0/0.93/0.07

- Scores for alternatives: INJURED 0/0.02/0.98   BANGED 0/0.72/0.28

- Original: LA MANSARDA PERETI OBLICI,CAMERA FOARTE STRIMTA PENTRU INALTIMEA NOASTRA,NE LOVEAM PERMANENT CU UMERII SAU CAPUL.

MT-based sentiment analysis

For the big picture, it does not matter as long as the MT engine that works **reasonably well**, we have a **fairly accurate** sentiment analyser and we apply them to **large quantities of data**

"Translation quality is not completely correlated to the performance of the classifier" [1] - but do we measure quality in the right way?

For the end user one mistake can matter, especially if it is systematic

Do we trust MT?

Reviewed: 22 November 2020

9.0

pleasant and comfortable

☺ · Excellent location for trips arriving by train in Sighisoara, less than five minutes from the train station .. For me the main advantage was the location and the fact that I always heard the trains in the room of the pension. Half an hour walk to the f and bathroom. Very comfortable bed. A big adv fridge in the room.

☹ · I have no negative mentions.

Translated by Google - Show original

Reviewed: 22 November 2020

9.0

placut si confortabil

☺ · Excelent situat pentru calatorii ce ajung cu trenul in Sighisoara, la nici cinci minute de gara de mers pe jos.. Pentru mine atuul principal a fost localizarea si faptul ca auzeam mereu trenurile din camera pensiunii. Juma de ora de mers pe jos pana la cetate. Mic magazin alimentar in apropiere. Camera si baia foarte curate. Patul foarte confortabil. Un mare avantaj pentru sederea pe timpul verii a fost minifrigiderul din camera.

☹ · Nu am mentiuni negative.

Show translation

MT of UGC for "human consumption"

If sentiment/emotions matter in the automatic translation, we should make sure MT engines translate these phenomena correctly

Reviewed: 26 August 2021

Superb

9.0

☺ ·
цент

☹ ·
Това

Show

Reviewed: 26 August 2021

Superb

9.0

☺ · The location is ideal for both business and leisure stays. Even though it's downtown, it's quiet and peaceful. Hygiene is high-level, with a delicate scent, even in the hallway.

☹ · We had to wait until after the check-in time, because the room wasn't ready. That was the only thing, everything else is okay.

Translated by **Booking.com** - Show original

👍 Helpful   👎 Not helpful

Reviewed: 3 August 2019

Passable

5.0

☺ ·

☹ ·
Топ

Show

Reviewed: 3 August 2019

Passable

5.0

☺ · Position and view.

☹ · It's absurd for second pillows to be in the shoe locker.

Hot water is not available.

Translated by **Booking.com** - Show original

1 person found this review helpful.   👍 Helpful   👎 Not helpful

- We rely on UGC to take decisions when we book services or purchase products

- The *Can't read, Won't buy* report by Common Sense Advisory from 2020 emphasised the importance of multilingual sites:
  - 40% won't buy in another language.
  - 65% prefer content in their native language.
  - 73% want reviews of products in their language.
  - 65% prefer content in their language – even if it's of poor quality (i.e. MT content)

- Automatic translation of reviews has become a common feature for websites like Amazon, Bookings, eBay, …

- but how good is this translation?

How to develop an MT engine
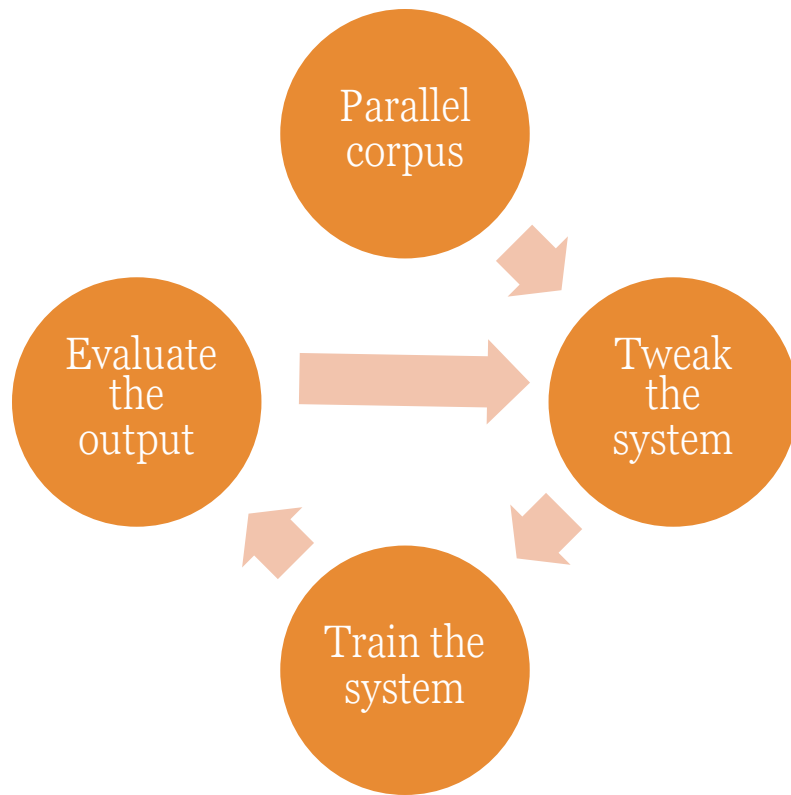
The evaluation is based on automatic metrics like BLEU, METEOR, BERTscore

These metrics do not capture the sentiment

REF: *The novel is great, the only flow is its last part*

HYP1: *The novel is* **terrible***, the only flow is its last part*
HYP2: *The* **story** *is great, the only flow is its last part*

We need to take the sentiment into account when developing the system

Incorporating sentiment in evaluation metrics

We should have a way to incorporate how well the sentiment expressed in the source is transferred to the target in the evaluation metric

Sentiment-Aware Measure (SAM) was presented earlier today [3]

SAM is a factor for standard MT evaluation metrics

Accounts for the 'sentiment-closeness' between the hypotheses and references

[3] Saadany, H., Orasan, C., Mohamed, E. and Tantavy A (2021). Sentiment-Aware Measure (SAM) for Evaluating Sentiment Transfer by Machine Translation Systems. *Proceedings of Recent Advances in Natural Language Processing, pages 1221–1230, Sep 1–3, 2021.* https://doi.org/10.26615/978-954-452-072-4_138
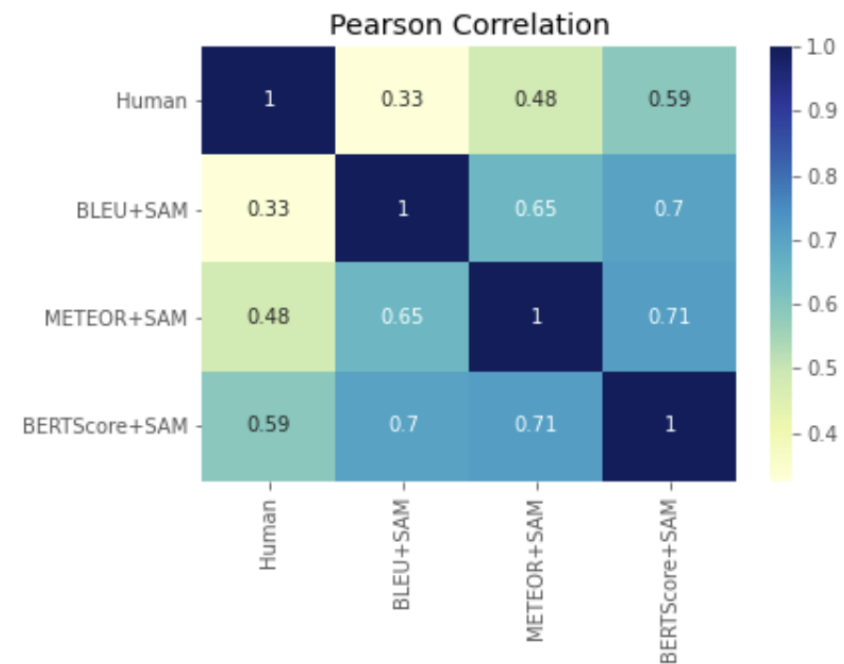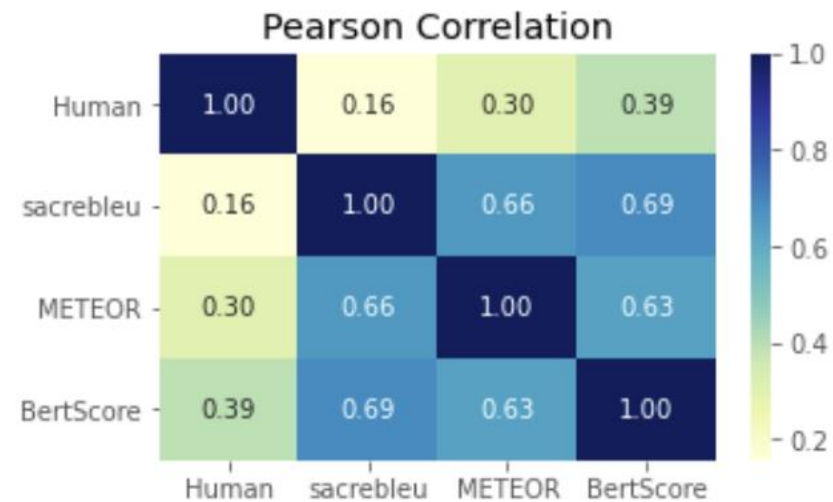
SAM evaluation

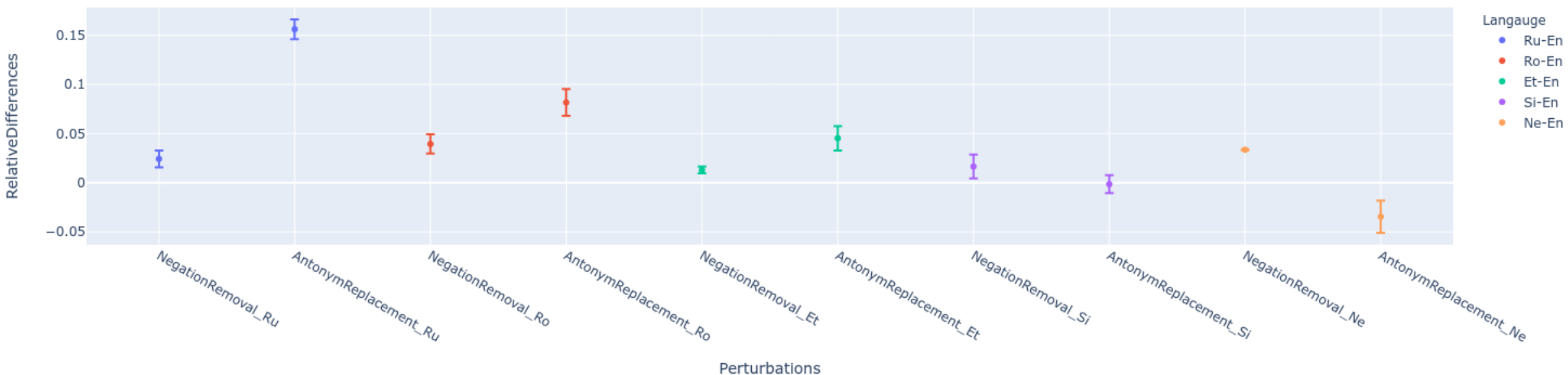Experiments with SacreBLEU, METEOR and BERTScore

400 translated tweets annotated with a score 0 to 10 to reflect the quality of the translation

SAM improves the Pearson and Kendall correlations

It has a number of limitations, and it is language specific (English)

MPPs and MAPs

change the polarity of sentence (negation or antonyms)

- The QE score could be used to inform the user
- It may be difficult to interpret it
- How much of the score is determined by the correct translation of the sentiment and how much of the rest?

[4] Berard, A., Calapodescu, I., Dymetman, M., Roux, C., Meunier, J.-L., & Nikoulina, V. (2019). Machine Translation of Restaurant Reviews: New Corpus for Domain Adaptation and Robustness. *Proceedings of the 3rd Workshop on Neural Generation and Translation*, 168–176. https://doi.org/10.18653/v1/D19-5617

[5] Tharindu Ranasinghe, Constantin Orasan and Ruslan Mitkov (2020) TransQuest: Translation Quality Estimation with Cross-lingual Transformers. In Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online), pp. 5070–5081, doi: 10.18653/v1/2020.coling-main.445

Sentiment preserving MT

How we can develop MT systems that preserve the sentiment of the source

Sentiment preserving MT

Develop different SMT systems each tuned to translate sentiments with a certain polarity [6]

Fine tuning of NMT models using in domain data [4]

Use reinforcement learning with curriculum learning to fine-tune an NMT system to preserve the sentiment [7]

… but there is no way to control specific phenomena

[6] Lohar, P., Afli, H., & Way, A. (2017). Maintaining Sentiment Polarity in Translation of User-Generated Content. *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 73–84. https://doi.org/10.1515/pralin-2017-0010

[7] Kumari, D., Chennabasavraj, S., Garera, N., & Ekbal, A. (2021). Sentiment Preservation in Review Translation using Curriculum-based Re-inforcement Framework. *Proceedings of the 18th Biennial Machine Translation Summit (Volume 1: MT Research Track)*, 150–162. https://aclanthology.org/2021.mtsummit-research.13/

Translation of book reviews from Arabic [8]

Corpus of book reviews written in Arabic (both Modern Standard Arabic and Dialectical Arabic)

Scores assigned by the authors of reviews

Automatically translated about 13,000 sentence to English using Google Translate and used Microsoft Azure Text Analytics to determine their sentiment

Selected 1,000 sentences where the sentiment of the translation disagreed the original sentiment to analyse the reasons for errors

[8] Saadany, H., & Orasan, C. (2020). Is it great or terrible? Preserving sentiment in neural machine translation of Arabic reviews. *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, 24-37. https://www.aclweb.org/anthology/2020.wanlp-1.3/

Error classification

Type of error	Automatic translation	Correct translation
Idiomatic	The book of <u>his blood is light</u>	The book is <u>funny</u>
Dialectical	and I felt that it was a <u>sweet</u> story	and I felt it was a <u>silly</u> story
Negation	I admire that the protagonist of the novel is weak in character	I didn't like that the protagonist of the novel has a weak character
Diacritic	The <u>envelope</u> of what I read	One of the <u>most entertaining</u> things I read
Contronyms	Wonderful narration and <u>rigid</u> style	Wonderful narration and <u>excellent</u> style
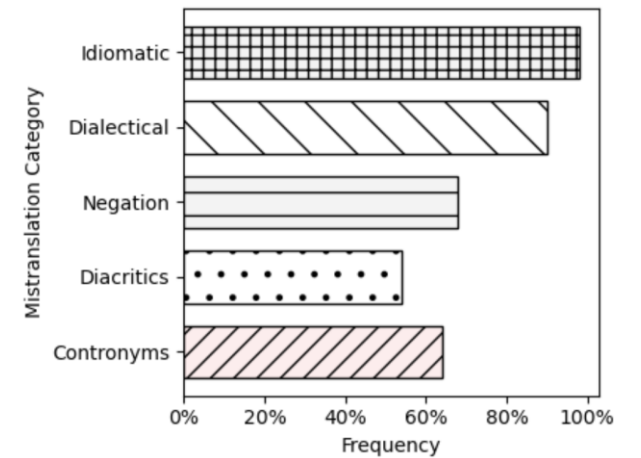
NMT system for contronyms

We cannot address all the problems at once and probably each problem should be addressed using a different approach

We trained an NMT engine on a general corpus and tuned it on synthetic data to translate contronyms correctly

Prepared a dataset
- Automatically translated reviews from Goodreads from Arabic to English
- Manually corrected contronyms that were not correctly translated (but did not correct the rest of the sentence)
- Annotated each contronym with its polarity given the context

NMT system for contronyms (II)

We used OpenNMT

Three models tried:

- seq2seq model
- Transformer 1 (tagging with sentiment)
- Transformer 2 (tagging with sentiment and pretrained with a vector model created for contronyms)

Method	BLEU	F1 – correct polarity	Sentence score distance
Seq2seq	33.9	0.55	0.24/0.44
Transformer 1	**38.77**	0.69	0.14/0.21
Transformer 2	37.14	**0.79**	**0.06/0.14**

NMT system for contronyms (III)

It is possible to develop MT systems that can address specific errors

But how does it affect the overall translation?

How our work applies to other language pairs (which language specific errors?)

How we can address several phenomena at once?

Moving to more difficult problems

Focus on emotions

Translate UGC from social media

What are emotions?

Dictionary definition = "strong feeling deriving from one's circu[mstances, mood, or relationships with] others."

Emotions usually have a polarity attached

NLP research has focused on Ekman's six major emotions: *sadness, joy, anger, disgust, fear and surprise*

There is quite a bit of disagreement whether these emotions are universal across cultures

We would like to translate UGC in order to understand what the authors say especially in situations of crisis, emergencies, etc. (but poor translation can lead to misunderstandings and/or an incomplete picture)

"emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal-directed, and adaptive" (Kleinginnas 1981, p. 355)

A BRITISH PERSON SAYS "*THAT'S A VERY BRAVE PROPOSAL*"

Positive or negative?

Please go to **pollev.com/corasan432** to vote

What the British say	What the British mean	What others understand
I hear what you say	I disagree and do not want to discuss it further	He accepts my point of view
With the greatest respect...	I think you are an idiot	He is listening to me
That's not bad	That's good	That's poor
That is a very brave proposal	You are insane	He thinks I have courage
Quite good	A bit disappointing	Quite good
I would suggest...	Do it or be prepared to justify yourself	Think about the idea, but do what you like
Oh, incidentally/ by the way	The primary purpose of our discussion is...	That is not very important

Are emotions universal?

UGC is messy (especially on social media)

Non standard grammar and spelling

Typos, grammatical mistakes

Use of #hashtags part of the text

Emoticons/emojis

Code switching

Use of caps

No punctuation

Still #devastated and #sad #Tariq's #TariqLive channel is no more and #ProfessorBlackTruth as well. #YouTubeFAIL #EpicFAIL! #TariqNasheed 😢

10:09 AM · Jul 12, 2017 · Twitter Web Client

Hes no even scared of em anymore

#IndiaAgainstIslamists #RadicalIslam possess the danger not only2 #India ,but also 2the world .Its time 4 uprooting the source f #terrorism

4:46 PM · Jul 11, 2017 · Twitter for Android

Can we translate emotions automatically?

We run a pilot study to see what kind of errors Google translate makes when translating emotions in Tweets

Used datasets from SemEval tasks on emotion and aggression detection
- Source languages: English, Arabic and Spanish
- Target languages: English, Romanian, Arabic, Spanish and Portuguese
- Focused on joy, fear, aggression, and anger (from the annotated data)

We identified mismatches between the original emotion and of the translation

Manually analysed 4 x 400 translations with English as source and 2 x 1,500 with target English

Reasons for errors

We focused on six reasons for wrong translation of emotion: hashtags, non-standard orthography, idiomatic expressions, polysemy and grammar (mainly negation structures)

We annotated a maximum of 2 types of errors per tweet

At times the source and target were such a mess that it was difficult to decide on the error

Target languages like Romanian (61%) and Arabic (41%) have higher percentage of tweets with mistranslated emotions

Hashtags

Hashtags contain lots of important information, but it is unclear when they need to be translated or not

- EN: "Are u #depressed #hypo #manic #lonely #bored #nofriends #needfreinds #friend I feel chatty I wanna help ppl or just #makefriend 's #dm #moms"
- RO: "Sunteți #deprimat #hipo #maniac # singur #bored #nofriends #needfreinds #friend Mă simt vorbăreț Vreau să ajut plpl sau doar #makefriend's #dm #moms"

- EN: It's lack of #faith that makes #people #afraid of #meeting #challenges #MuhammadAli
- RO: Lipsa de # credință face ca # oamenii să se teamă de # întâlnire # provocări #MuhammadAli

- EN: *#disappointed ES: #diseñado (designed)*

It may be possible to normalise the hashtags, but which ones?

Slang and dialectical expressions

Slang and swear words are not translated correctly, usually meaning loss of emotion (usually aggression or anger)

Examples:

Some offensive examples coming …

CENSORED
for your protection

Do we really have to translate UGC which is aggressive/abusive? In which contexts it is necessary to translate swear words?

Non-standard orthography

UGC uses creative abbreviations, unconventional orthography
- "watching sad bts video bc im sad. Iwannacryy" is mistranslated to all target languages
- Such words are usually out of vocabulary, so cannot be translated

A normalisation step may be able to help, but it can also lead to information loss

Some of the non-standard orthography gets corrected in the translation process

Idiomatic expressions/Polysemy/Negation

All these are problematic in any context, but Arabic dialectical negation causes more problems in our data

At times funny effects:

- "A trip to the dentist never gets easier" -> PT: "Uma ida ao dentista nunca foi tão fácil" ("A trip to the dentist has never been easier")

A solution could be to inform the MT engine of the emotion of the source

Limitations

The annotation scheme needs to be revisited and made more consistent

It's not always clear which is the main source of error

What we have now is very much a pilot, so it would be good to repeat the annotation

See how much we can use some automatic processing

A small scale experiment on Facebook comments lead to similar results

Conclusions

Translation of user-generated content is here to stay

There is a whole range of challenges that need to be addressed in order to translate correctly sentiments and emotions, and we may have the approaches necessary to address some of them

We need to have evaluation metrics that capture how well the sentiments/emotions are translated

Educate users of the limitations of translating UGC (related to the issue of MT literacy)

In many cases culture specific issues need to be considered when translating sentiments/emotions

Thank you for your attention!

If you are looking for a PhD studentship and the topic of this talk is interesting you, get in touch!!!