Comparative Evaluation of Neural Machine Translation Quality in Arabic<>English Translation

AsLing: TC 42 2020

19 November

Khaled ben Milad (Alamary), Swansea University

Neural Machine Translation and Less-Resourced Languages

- Neural machine translation has shown a high level of quality in many language pairs(Bojar et al., <u>2016)</u>
- Success is limited to language pairs with availability of large amounts of parallel data (Koehn and Knowles, 2017)
- Arabic is a low resource language

Related Studies

- Abdelaal and Alazzawie (2020)
- I. Investigated Google's output
- II. From Arabic to English
- III. High adequacy and some errors in semantic fluency
- Al-mahasees (<u>2020</u>)
- I. Google, Bing, in addition to the Sakhr hybrid
- II. Conducted twice, in 2016 and 2017
- III. Google outperformed the other systems in terms of adequate and fluent output over the two years in both directions

Methodology

- Ten texts were extracted from LDC2004T18: test one consists of 117 sentences in Arabic, while test two includes 109 sentences in English.
- Sentences length ranging from 19-39 words.
- Sentences translated using Bing NMT, Google NMT and Yandex NMT, as well as the commercial interactive and adaptive Lilt tool.

Survey

- Subset of four source sentences, each paired with its four MT systems' outputs in each direction, was distributed in random order in an online questionnaire
- Participants were asked to rate adequacy and fluency on a four-point Likert scale, according to TAUS quality criteria

Evaluators

- Responses: 10 complete; 23 partial
- Advanced students
- Respondents' first language: Arabic
- Duration: one month
- Completion time: about 40 mins

ST: Orejuela seemed quite calm as he was <u>being led to</u> the American plane that would take him to Miami in Florida.
REF: وبدا اوريخويلا هادئا عندما اقتيد الى الطائرة الاميركية التي ستنقله الى ميامي في فلوريدا
Transliteration/ wabada awrykhwyla hadia eindama aqtid 'iilaa alttayirat al'amirkiat alty satunqiluh 'iilaa mayami fi flurida /

Bing NMT

- هادئة جدا كما كان يجري نقله إلى Orejuela بدا
 الطائرة الأمريكية التي من شأنها أن تأخذه إلى ميامي في
 فلوريدا
- **Back translation**: Orejuela looked very quiet as he was being <u>transported</u> to an American plane that would take him to Miami, Florida.

Google NMT

- هادئًا تمامًا لأنه كان يقود إلى Orejuela بدا الطائرة الأمريكية التي ستأخذه إلى ميامي في فلوريدا
- **Back translation**: Orejuela seemed completely calm as he was <u>driving</u> to the American plane that would take him to Miami, Florida.

ST: Orejuela seemed quite calm as he was <u>being led to</u> the American plane that would take him to Miami in Florida.
REF: وبدا اوريخويلا هادئا عندما اقتيد الى الطائرة الاميركية التي ستنقله الى ميامي في فلوريدا Transliteration/ wabada awrykhwyla hadia eindama aqtid 'iilaa alttayirat al'amirkiat alty satunqiluh 'iilaa mayami fi flurida /

Lilt

- بدا هادئاً تماماً بينما كان يقود إلى Orejuela الطائرة الأمريكية التي ستأخذه إلى (ميامي) في ((فلوريدا
- Back translation: Orejuela seemed completely calm while <u>driving</u> to the American plane that would take him to Miami, Florida.

Yandix NMT

- أوريخويلا بدت هادئة تماما كما كان يجري أدى إلى الطائرة الأمريكية التي من شأنها أن تأخذه إلى ميامي في ولاية فلوريدا
- **Back translation**: Orejuela seemed perfectly calm as he <u>drove</u> to the American plane that would take him to Miami, Florida.)

Human Judgment Findings 1

Overall Rating of Adequacy and Fluency for NMT Systems - en to ar Translation



Human Judgment Findings 2

Overall Rating of Adequacy and Fluency for NMT Systems - ar to en translation





ModernMT Results

During the period of free translation service offered due to COVID-19 - in May 2020.

 ModernMT's output received a better BLEU score than Lilt and Yandex in terms of Arabic to English translation, while it scored the best regarding English to Arabic translation, in which proper names were translated.

Summary

- NMT systems produced more adequate translation than fluent translation in both directions.
- Producing fluent translation into Arabic was more difficult than into English.
- Google NMT was rated by evaluators the most adequate and fluent system in both translation directions, while Bing NMT achieved the best BLEU score.

References

- Abdelaal Noureldin and Alazzawie Abdulkhaliq (2020) <u>Machine Translation: The Case of Arabic-English Translation of News Texts.</u>
- Al Mahasees (2020) Diachronic Evaluation of Google Translate, Microsoft Translator and Sakhr in English<> Arabic Translation
- <u>https://www.bing.com/translator</u>
- <u>https://translate.google.co.uk/</u>
- <u>https://www.lilt.com/</u>
- <u>https://www.modernmt.com/</u>
- <u>https://translate.yandex.com/</u>
- Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. (2014). <u>On the</u> properties of neural machine translation: Encoder-decoder approaches. In Proceedings of SSST-8.
- Ond^{*}rej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Aurelie Neveol, Mariana Neves, Martin Popel, Matt Post, Raphael Rubino, Carolina Scarton, Lucia Specia, Marco Turchi, Karin Verspoor, and Marcos Zampieri. (2016). <u>Findings of the 2016</u> <u>conference on machine translation.</u> In Proceedings of the First Conference on Machine Translation. Association for Computational Linguistics, Berlin, Germany, pages 131–198
- Sutskever Ily, Vinyals Oriol, and Quoc. V. Le. (2014). <u>Sequence to sequence learning with neural</u> <u>networks.</u> Technical report, arXiv preprint arXiv:1409.3215,
- Philipp Koehn and Rebecca Knowles. (2017). <u>Six challenges for neural machine translation.</u> In Proceedings of the First Workshop on Neural Machine Translation, pages 28–39, Vancouver, August. Association for Computational Linguistics.