Neural and Symbolic Arabic Paraphasing with Automatic Evaluation

First Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Second Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Abstract

We present a tool for Arabic paraphrasing that yields good paraphrasing accuracy. We present and compare several methods for paraphrasing and obtaining monolingual parallel data. We also present first results on Arabic paraphrasing using neural methods. Additionally, we propose a new evaluation metric for paraphrasing that is shown to correlate highly with human judgement.

In order to extract paraphrases, we first obtained a parallel bilingual corpus using the EUROPARL dataset (Koehn, 2005). (Bannard and Callison-Burch, 2005)

We plan to explore other options for obtaining monolingual parallel data. One possible approach is to retreive headlines of news articles from different agencies covering the same event. We expect headlines describing the same event to have some degree of semantic similarity yet different surface realizations.
The sequence-to-sequence models required a relatively long time to run which limited the testing of other architectures and model options. We plan to conduct more expriments on different architectures and compare results.

Acknowledgments

References

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 597–604, Stroudsburg, PA, USA. Association for Computational Linguistics.

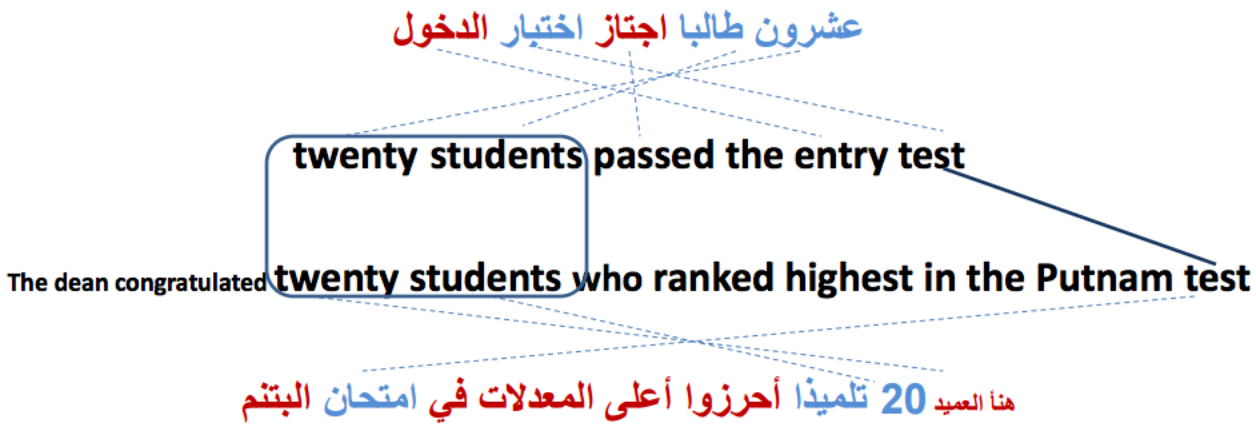

Figure 1: An example paraphrased sentence produced using the pivot method.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. http://www.statmt.org/europarl/.