

# Hybrid Word-Character Neural Machine Translation Pamela Toman, Sigtryggur Kjartansson

## **Motivation & Background**

- Words are **not** the primary level of meaning in all languages.
- Because traditional NMT is word-level, non-analytic languages tend to have lower-quality translations.
- Want a universalist architecture that performs well for all language pairs.



Focusing on Arabic: 5th largest language by number of speakers, understudied, and has a variety of clitics, affixes, spelling ambiguities, and the root-and-pattern morphology of Semitic languages.





30

Sequence Length

40

**Goal:** Improve Arabic-English Machine Translation. **Benchmark:** Almahairi et al. 2016<sup>1</sup> using BLEU metric. **Approach**: Neural Machine Translation model that achieves open vocabulary.

- 2016:<sup>2</sup>
  - Source Character-Level Encoder.  $\bigcirc$
  - Word-Level Sequence-to-Sequence with global Ο bilinear attention.<sup>3</sup>
  - Target Character-Level Generator. 0
- Backs off to the character level when word-level representations do not exist, in order to achieve open vocabulary using a small known vocabulary.



Department of Computer Science, Stanford University

### **Problem Statement**

#### **Method**

• We unite three models to create a hybrid word-character model, based on Luong and Manning

### **Experimental Evaluation & Results**

- Achieved BLEU score: 42.10
- Evaluate model by:
- Train with different Ο hidden states sizes.
- Train with different  $\bigcirc$ vocabulary sizes.
- Visualize embeddings 0 using t-SNE.
- Inspect attention Ο alignments.
- Use alternative  $\bigcirc$ frequency sampling techniques.



#### **Conclusion & Future Direction**

- Hybrid architecture combines the strength of both word- and character-based models; fast to train and offer high-quality translation; and achieves open vocabulary.
- Want to add a character-level attention mechanism and/or a convolutional layer to facilitate interactions with Arabic's complex morphology.

<sup>2</sup> Luong and Manning, 2016; arXiv:1604.00788

Acknowledgements: Ignacio Cases, Microsoft Azure, CS224N staff



<sup>&</sup>lt;sup>1</sup> Almahairi et al., 2016; arXiv:1606.02680

<sup>&</sup>lt;sup>3</sup> Luong, Pham and Manning, 2015; arXiv:1508.04025