RQ1: Which is the best lexical input (character, word, sentence, thought) to generate language representations for a translation task?

1) Literature on the generation of **character** embeddings:

   Zalando Paper: “Contextual String Embeddings for Sequence Labeling”

   Google Brain Team: “Exploring the limits of Language Modeling”

   Github repo:
   https://github.com/tensorflow/models/blob/master/research/lm_1b/
   ReadMe says:
   The code supports 4 evaluation modes:

   - Given provided dataset, calculate the model's perplexity.
   - Given a prefix sentence, predict the next words.
   - Dump the softmax embedding, character-level CNN word embeddings.
   - Give a sentence, dump the embedding from the LSTM state.

2) Literature on the generation of **word** embeddings:

   Most obvious: Mikolovs’ “Efficient Estimation of Word Representations in vector Space”

   But I guess best is Peters’ “Semi-supervised sequence tagging with bidirectional language models”

3) Literature on the generation of **sentence** embeddings:

   Peters’ “Deep contextualized word representations”

   ELMo Embeddings:
   <[…]each token is assigned a representation that is a function of the entire input sentence[…]>

   Google AI Language Team: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”

4) Literature on the generation of **thought** embeddings (or rather document embeddings):

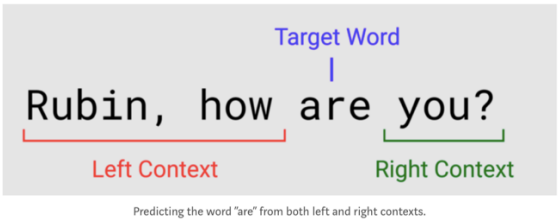
   Overview: https://en.wikipedia.org/wiki/Thought_vector

   Mikolovs’ “Distributed Representations of Sentences and Documents”

   doc2vec – model

RQ2: Which is the best Language Model (bi-directional, one-directional, etc.) to use for generating language representations applied to a translation task?

I. Literature on using Bi-directional Language Models:



Predicting the word "are" from both left and right contexts.

Peters' "Semi-supervised sequence tagging with bidirectional language models"

Zalando Paper: "Contextual String Embeddings for Sequence Labeling"

Mikolovs' "Efficient Estimation of Word Representations in vector Space"

Peters' "Deep contextualized word representations"

Google AI Language: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding"

II. Literature on using One-directional Language Models:

Google Brain Team: "Exploring the limits of Language Modeling"

Mikolovs' "Distributed Representations of Sentences and Documents"

The *DAN-Encoder* of Google's "Universal Sentence Encoder"

Stanfords' "GloVe: Global Vectors for Word Representation"