

Supplementary Materials For: When and Why are Pre-trained Word Embeddings Useful for Neural Machine Translation?

Ye Qi, Devendra Singh Sachan, Matthieu Felix, Sarguna Janani Padmanabhan, Graham Neubig

Language Technologies Institute, Carnegie Mellon University

{yeq, dsachan, matthief, sjpadman}@andrew.cmu.edu, gneubig@cs.cmu.edu

Language	Vocab	Pre-trained	Ratio
GL	18,258	14,702	0.80
PT	41,327	33,315	0.79
AZ	18,567	13,124	0.70
TR	202,520	103,135	0.50
BE	14,561	11,143	0.76
RU	172,682	143,821	0.81
ES	36,105	33,087	0.90
FR	35,474	32,873	0.77
IT	38,247	31,933	0.83
RU	62,488	58,081	0.89
HE	61,936	51,552	0.83

Table 1: The first column shows the vocabulary size of each language used in our experiments for translation to English. The second column shows the number of words for each language whose pre-trained vectors were present in `fastText` embeddings. The third column shows the ratio of the number of words with pre-trained embeddings to vocabulary size. The statistics shown in the second half of the table are computed from a dataset consisting of 40,000 sentences for each language which is used for experiments in §5.

1 Pre-trained Embedding Coverage

The ratio of the number of pre-trained word embedding over the vocabulary size will indicate how many words will be able to take advantage of pre-trained embeddings. Thus, this is a key factor connected to the efficacy of pre-trained word embedding. Table 1 reports these statistics of the embeddings used in our experiments.

2 Example Translations

The selected sentences in Table 2 further demonstrate the superiority of models using pre-trained embeddings in generating sentences that are more grammatical for source sentences that consist of rarer vocabulary and less frequent phrases.

source	(<i>risos</i>) e é que chris é un grande avogado , pero non sabía case nada sobre <u>lexislación de patentes</u> e absolutamente nada sobre xenética .
reference	(<i>laughter</i>) now chris is a <u>really brilliant lawyer</u> , but he knew almost nothing about <u>patent law</u> and certainly nothing about genetics .
bi:std	(<i>laughter</i>) and i 'm not a little bit of a little bit of a little bit of and (<i>laughter</i>) and i 'm going to be able to be a lot of years .
bi:pre	(<i>laughter</i>) and it 's what i 'm got a <u>great teacher</u> , but i did n't know it was nothing about the <u>economic</u> and nothing about the genetic test .
bi:pre-align	(<i>laughter</i>) and it 's a <u>big bit of a gift</u> , but i did n't know anything about the <u>political spills</u> and not anything about any diversity .
multi:std	(<i>laughter</i>) and it 's been a <u>big little bit likely</u> , but i did n't know that there 's nothing about the <u>fuel of diarrhea</u> and absolutely nothing about genetic .
multi:pre	(<i>laughter</i>) and it 's that chris is a <u>big lawyer</u> , but i did n't know anything about the <u>patent</u> and absolutely anything about genetic .
multi:pre-align	(<i>laughter</i>) and chris is a <u>big lawyer</u> , but i did n't know almost anything about <u>patent legislation</u> and absolutely nothing about genetic .
source	“ este traballo chámase “ ” a <u>sombra do sol</u> ” ” . <i>foi case como unha folla de papel</i> , coma un recortable dun debuxo infantil dunha mancha de petróleo ou un sol . ”
reference	“ so this is a work called “ ” the <u>sun shadow</u> . ” ” and <i>it was almost like a sheet of paper</i> , like a cutout of a childlike drawing of an oil spill or a sun . ”
bi:std	“ it 's a person , “ ” the <u>second is a man</u> , a man 's a high prize , a country , a person 's a bit of a year of a year . ”
bi:pre	“ this is called “ ” the <u>beautiful side of the sun</u> ” ” <i>it was almost like a piece of a real environment</i> , like a piece of a blue of a piece of depression . ”
bi:pre-align	“ this might be called “ ” the <u>mountain of the sun</u> ” ” ” <i>it was almost a “ ” a piece of the middle of the middle of a piece of the oil or a sun , ” ”</i>
multi:std	“ this work is called “ ” the <u>sun 's line</u> . ” ” <i>it was almost like a whole new day</i> , as a letter of a child 's year 's 20 or a sun . ”
multi:pre	“ this work is called “ ” the <u>shadow of the sun</u> . ” ” <i>it was almost a role of paper</i> , like a quarter of a child 's child 's black fashion called a sun . ”
multi:pre-align	“ this work is called “ ” the <u>shadow of the sun</u> . ” ” <i>it was almost a paper production</i> , like a moment of a child 's piece of oil or a sun . ”

Table 2: Example translations of GL → EN.