

# **1 Which datasets are used in training and evaluation (size, language, etc.)?**

WMT '14 contains the following English-French parallel corpora. Europarl with 61M words, news commentary with 5.5M words, UN (United Nations) with 421M words and two crawled corporas with sizes of 90M and 272.5M words. Total words is 850M words which was reduced to final corpus of size 348M words.

# **2 Which neural model architectures are evaluated in the paper?**

Most of the proposed neural machine translation models belong to a family of encoder-decoders. one is an RNN Encoder–Decoder, the other is the proposed model of the paper called as RNNsearch

# **3 Briefly summarize the results and the used metric. Did the neural attention improve the results compared to the previously existing models?**