# Sent2vec

**First ideas:**

- Average pre-trained word vectors *(word2vec, GloVe, etc)*.
- Maybe use TF-IDF weights for averaging.

**Sent2vec:**

- Learn sentence embedding as a sum of sub-sentence units:

$$sim(u,s) = \frac{1}{|G_s|} \sum_{g \in G_s} \langle \phi_u, \theta_g \rangle$$

where $G_s$ is a set of word n-grams for the sentence *s*.

**Code and embeddings:** https://github.com/epfml/sent2vec

Pagliardini et. al. Unsupervised Learning of Sentence Embeddings using Compositional n-Gram Features, 2017.