1. Efficient estimation of word representations in vector space <https://arxiv.org/abs/1301.3781>
2. Distributed representations of words and phrases <https://arxiv.org/abs/1310.4546>
3. Deep contextualized word representations <https://arxiv.org/abs/1802.05365>
4. Attention is all you need <https://arxiv.org/abs/1706.03762>
5. Neural machine translation by jointly learning to align and translate <https://arxiv.org/abs/1409.0473>
6. Toward controlled generation of text <https://arxiv.org/abs/1703.00955>
7. Content preserving text generation with attribute controls <https://arxiv.org/abs/1811.01135>
8. Plug and play <https://arxiv.org/abs/1912.02164>
9. Transition-Based Dependency Parsing with Stack Long Short-Term Memory <https://arxiv.org/pdf/1505.08075.pdf>
10. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding <https://arxiv.org/abs/1810.04805>