1. Neural Lattice Language Models (Buckman and Neubig 2018)
2. Quasi-Recurrent Neural Networks (Bradbury et al. 2016)
3. Learned in translation: Contextualized word vectors (McCann et al. 2017)
4. Deep contextualized word representations Blogpost (Peters et al. 2018)

# References

Bradbury, James, Stephen Merity, Caiming Xiong, and Richard Socher. 2016. "Quasi-Recurrent Neural Networks." *arXiv Preprint arXiv:1611.01576.*

Buckman, Jacob, and Graham Neubig. 2018. "Neural Lattice Language Models." *arXiv Preprint arXiv:1803.05071.*

McCann, Bryan, James Bradbury, Caiming Xiong, and Richard Socher. 2017. "Learned in Translation: Contextualized Word Vectors." In *Advances in Neural Information Processing Systems*, 6297–6308.

Peters, Matthew E, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. "Deep Contextualized Word Representations." *arXiv Preprint arXiv:1802.05365.*