

Summary

- Transformer is a great replacement for RNN.
 - Constant-time processing.
 - (CNNs *can* be comparable, but Gehring et al. (2016) was kind of missed.)
- Explicit syntax can be useful.
 - Many options how to include it.
 - Some gains hard to reproduce.
 - Dummy information *can be equally useful*.
 - Transformer *seems to learn syntax for free*.