Effect of residual connection on perplexity evolution for 4-layer network Residual train perplexity Residual val perplexity Regular train perplexity 10<sup>2</sup> Regular val perplexity **Perplexity** 10<sup>1</sup> 10<sup>3</sup> 10<sup>4</sup> 10<sup>5</sup> **Iterations**