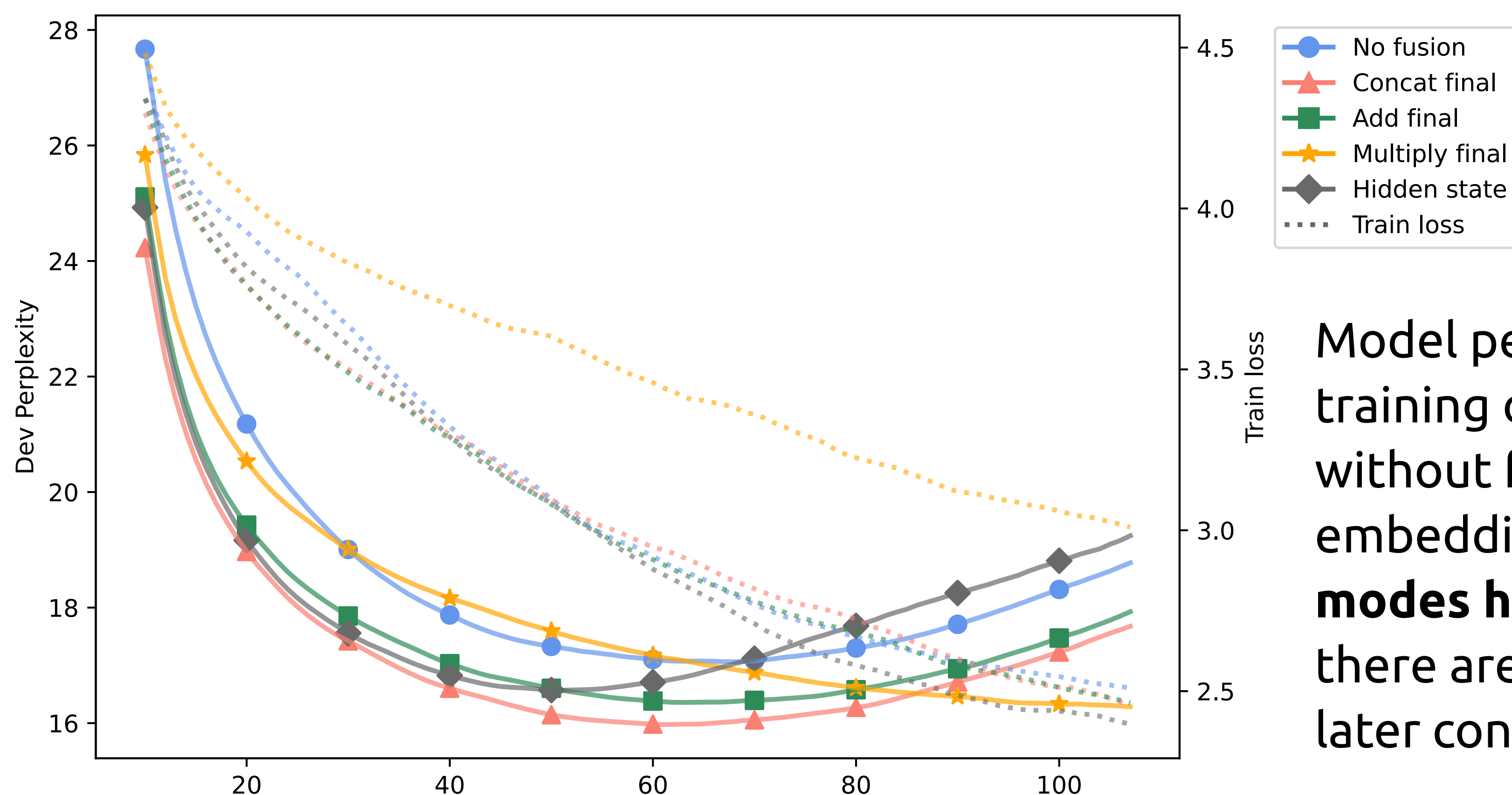
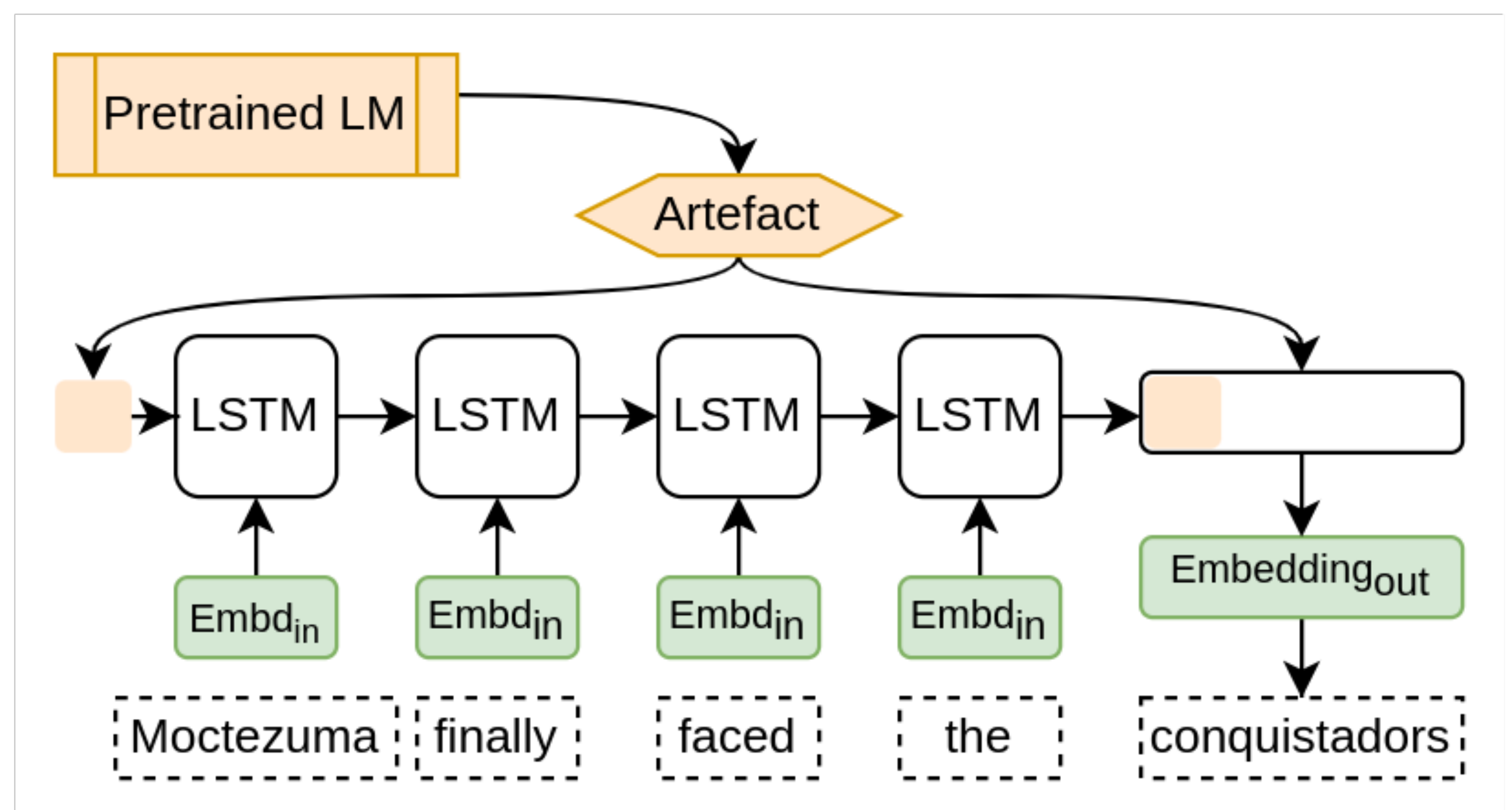


Fusing Sentence Embeddings Into LSTM-based Autoregressive Language Models

Vilém Zouhar, Marius Mosbach, Dietrich Klakow, 2022 (in review)

Example architecture of an LSTM language model enhanced with **fusion from a pretrained LM** illustrated on a sample sentence. Knowledge of *Moctezuma* helps in the next word prediction.



Model performance during training of models with and without fusion of sentence embeddings. **All fusion modes help reliably** but there are differences (e.g. later convergence).

How does fusion actually work?

Hypothesis: it moves the computation closer (in any intermediate vector space) to the correct output.

Future work: Why does early fusion fail if, *in theory*, the model has access to all information sooner.

