

Evolution of Representations in Cross-lingual fine-tuning

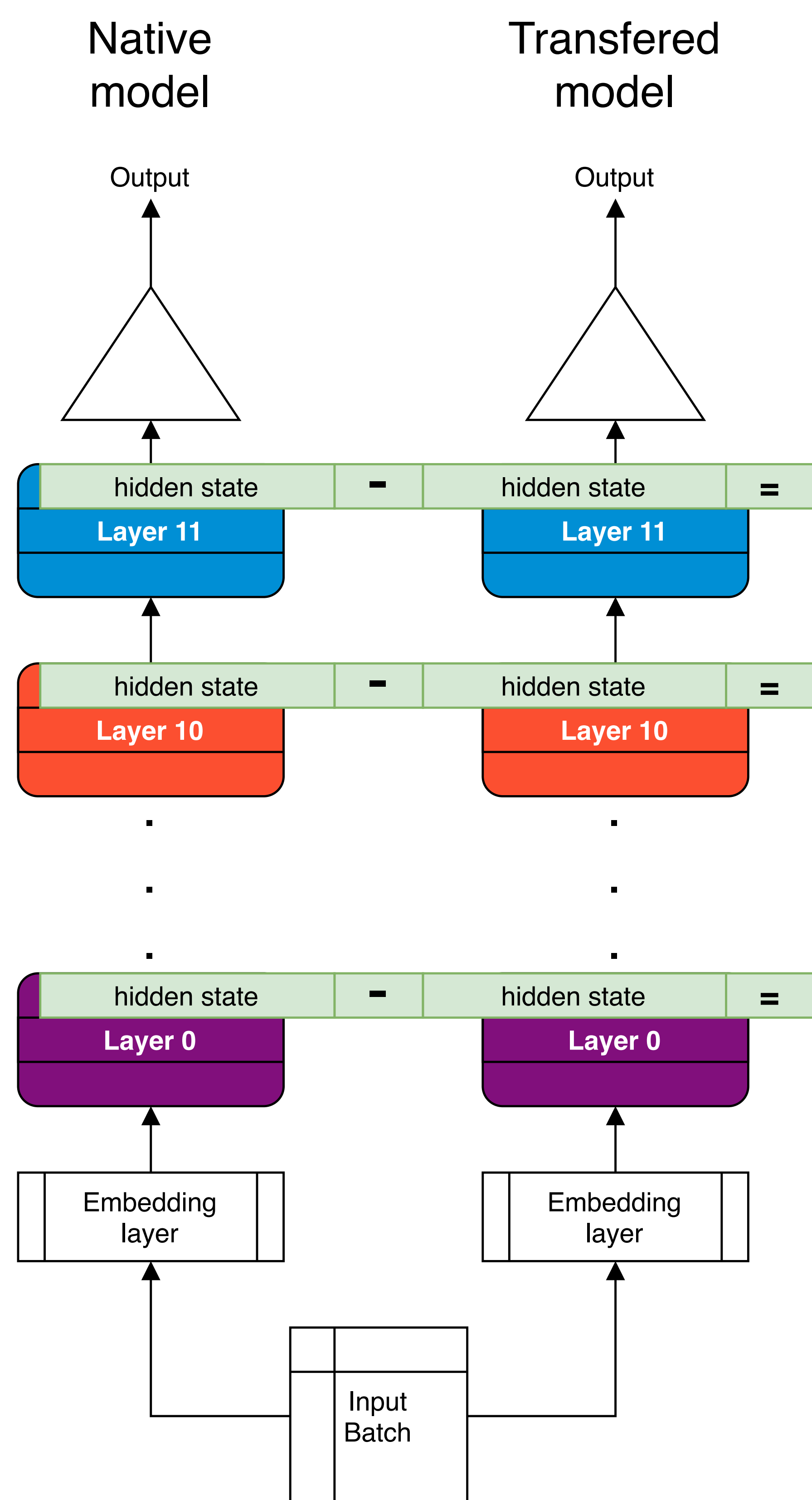
Aman Hussain, Emil Dudev
Universiteit van Amsterdam

Introduction

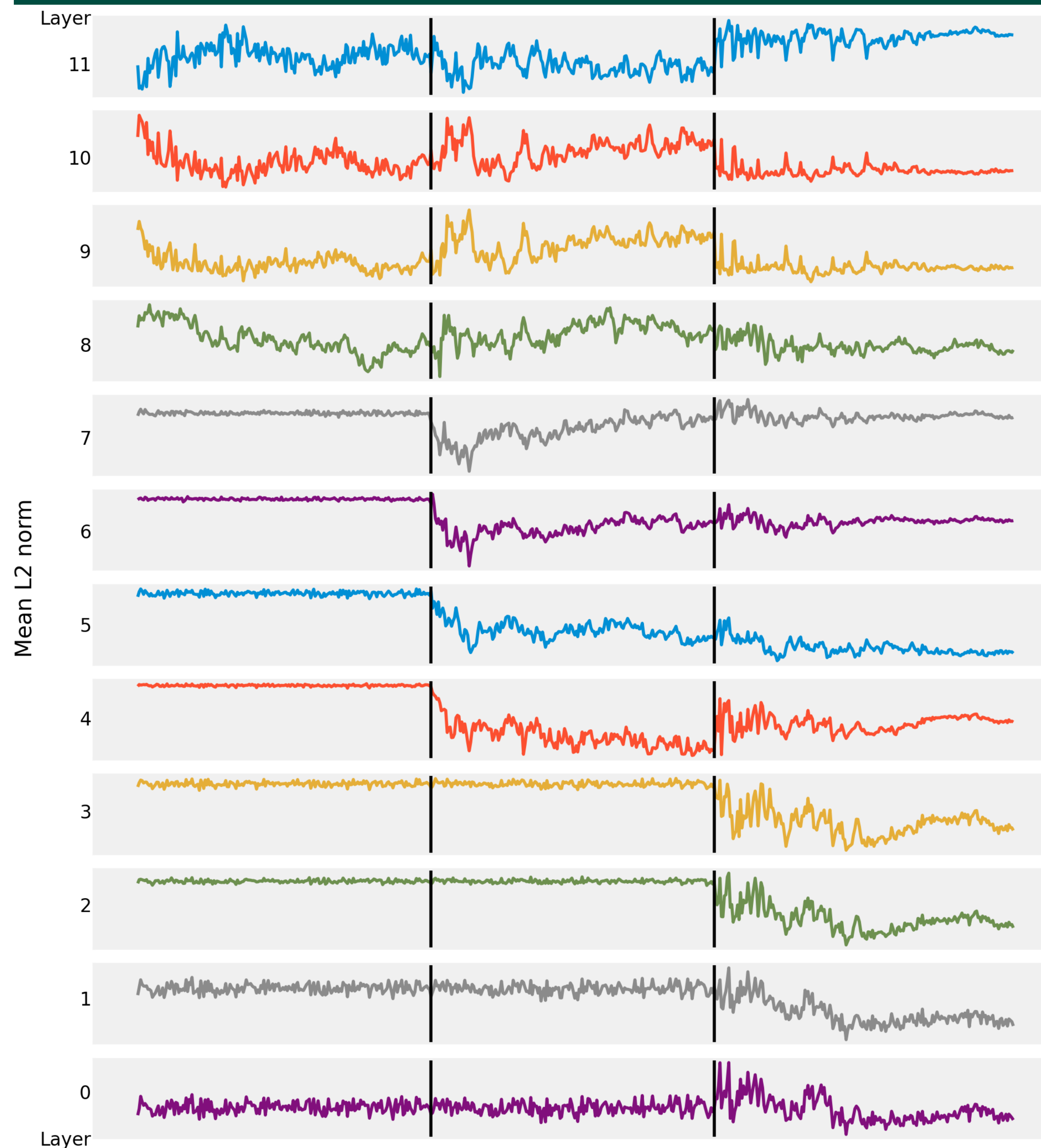
- Transferring multi-lingual BERT to Dutch
- **How representations evolve during fine-tuning?**
- Representation space: Native vs. Transferred

Experiment Setup

- Target language: Dutch
- Native Model: BERTje (Rijksuniversiteit Groningen)
- Transfer Model: BERT Multi-lingual

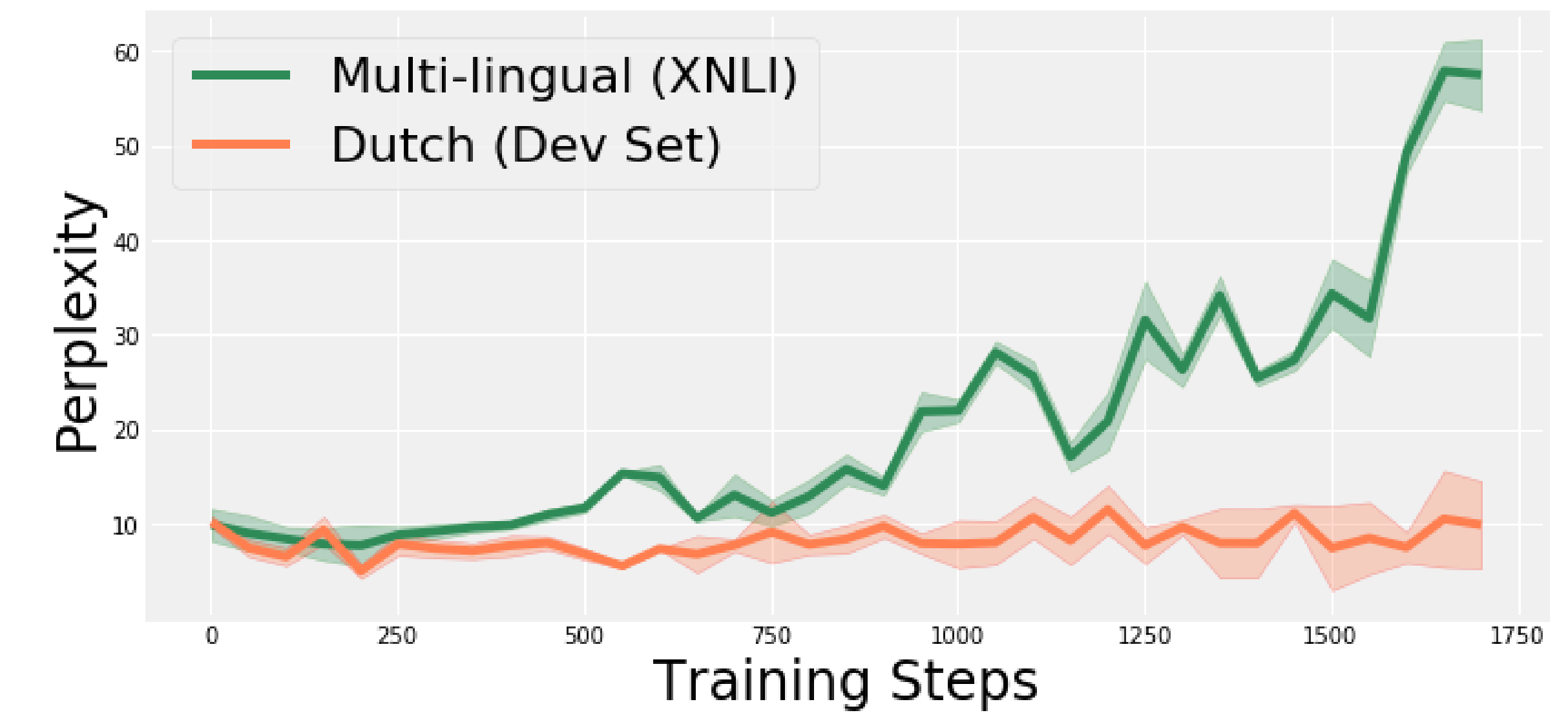


Transferred BERT learns a different representation subspace than Native BERT

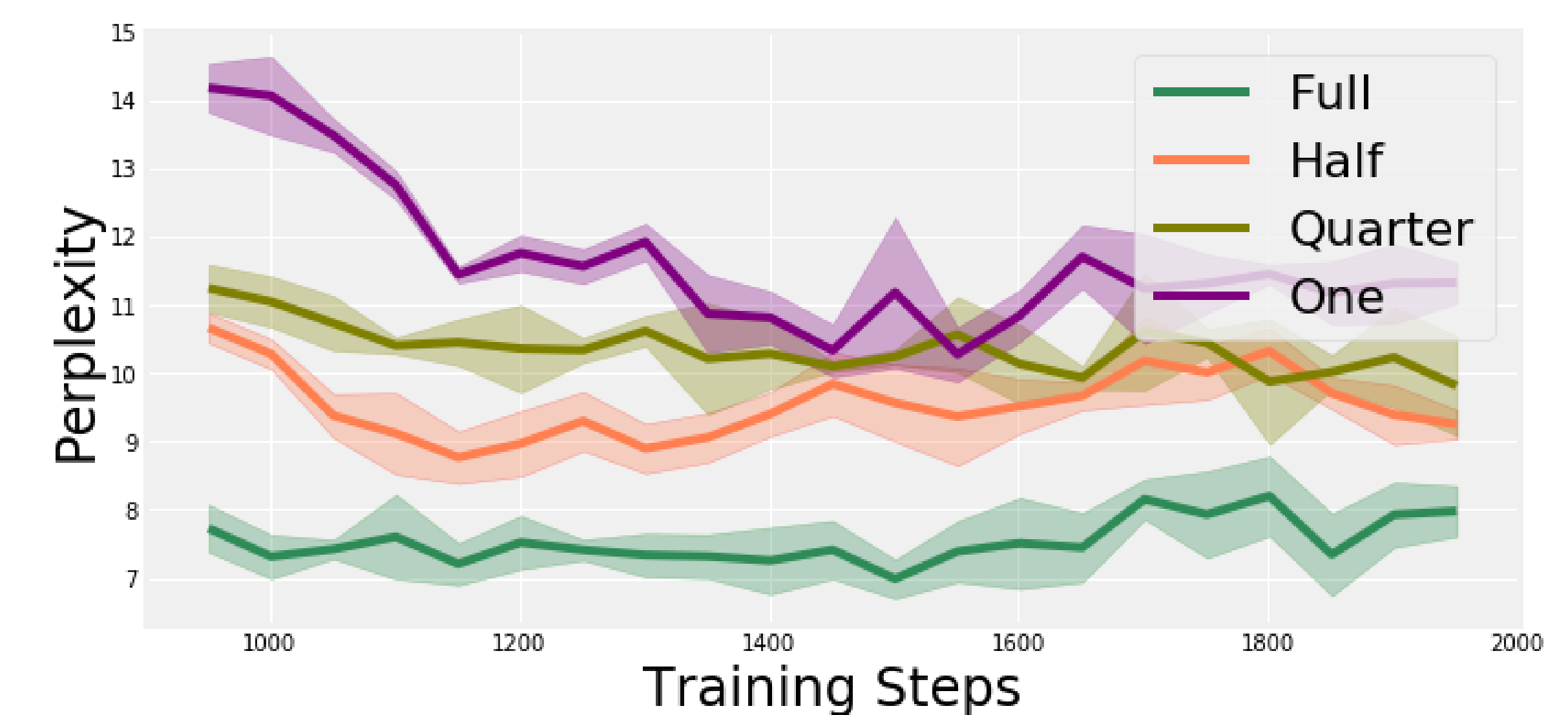


Attention Differences Compared to Native Model

Results



Multi-lingual Capacity during Fine-tuning



Comparison of Layer Freezing Approaches

Analysis

- Do we overfit to our Dutch data? No!
- Tuning the last layer quickly leads to divergence
- Multiple good representations
- Multiple local minima

Conclusion

- Multi-lingual BERT learns to focus only on Dutch
- Better to fine-tune all of BERT (Empirically)
- Dutch-tuned BERT divergent from Native Model