Loss over time per dialect (with pre-training)

