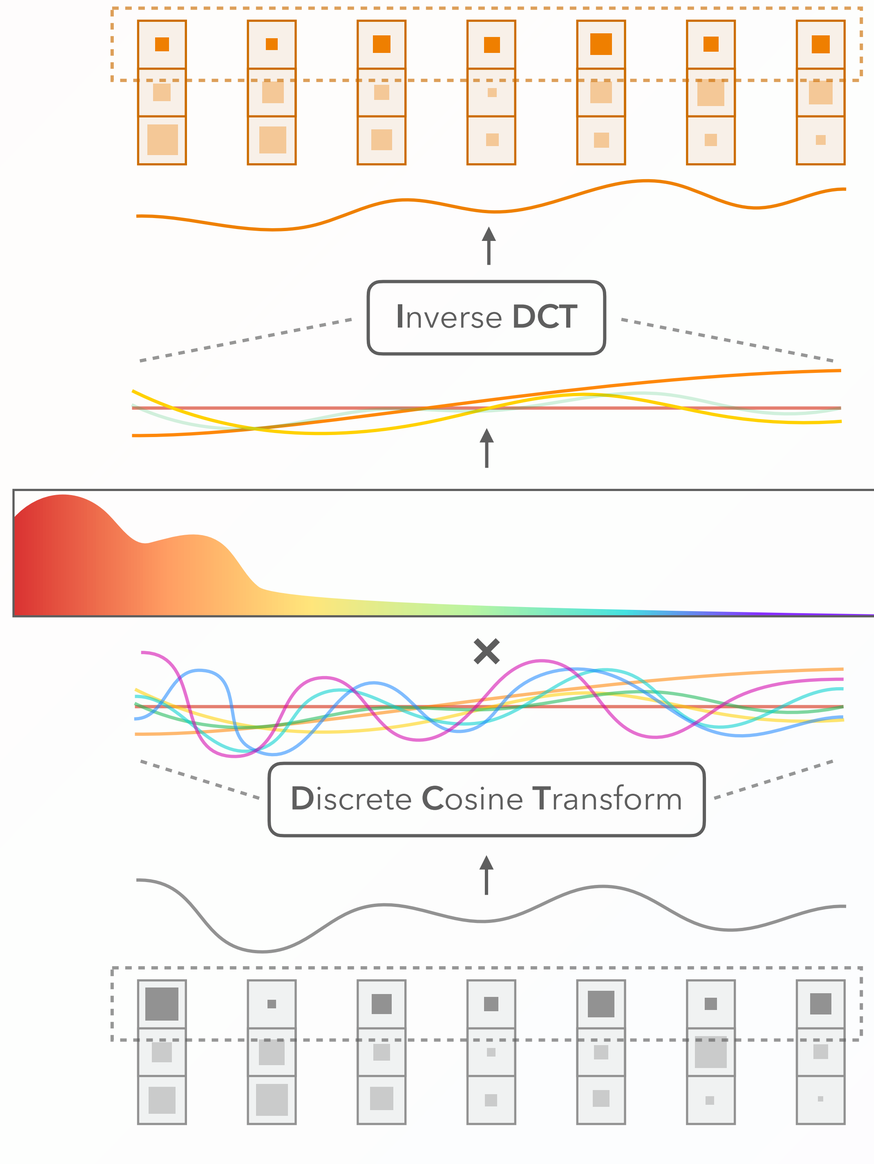
Spectral Probing

Max **Müller-Eberstein**, Rob **van der Goot** and Barbara **Plank**

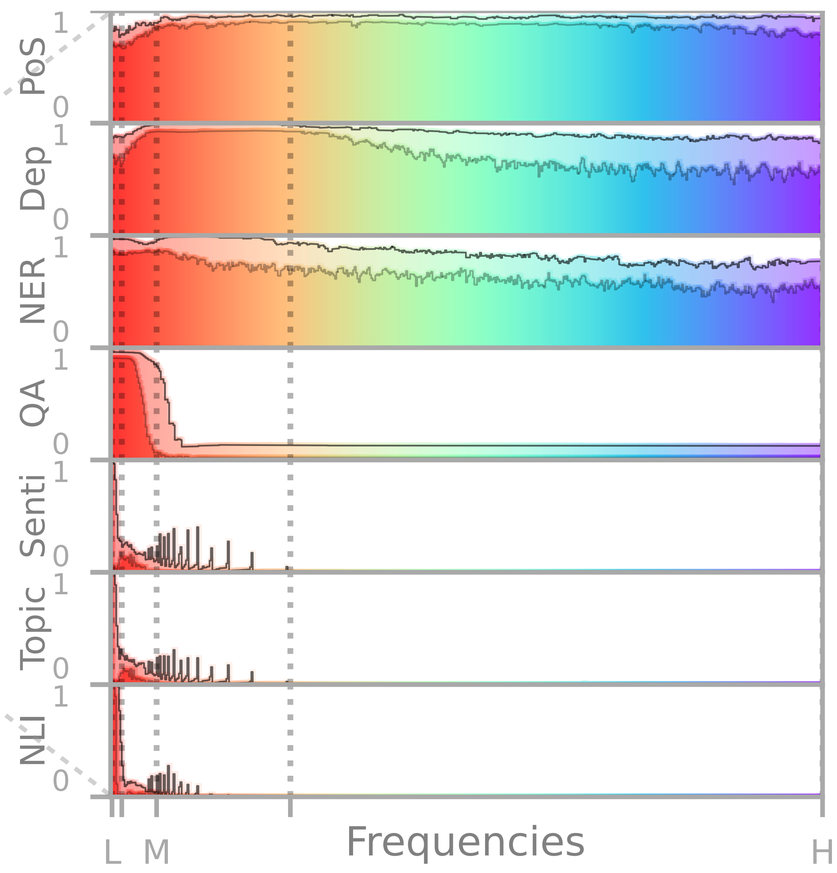
At what **timescales** are task and language-specific contextualized information encoded?

Method

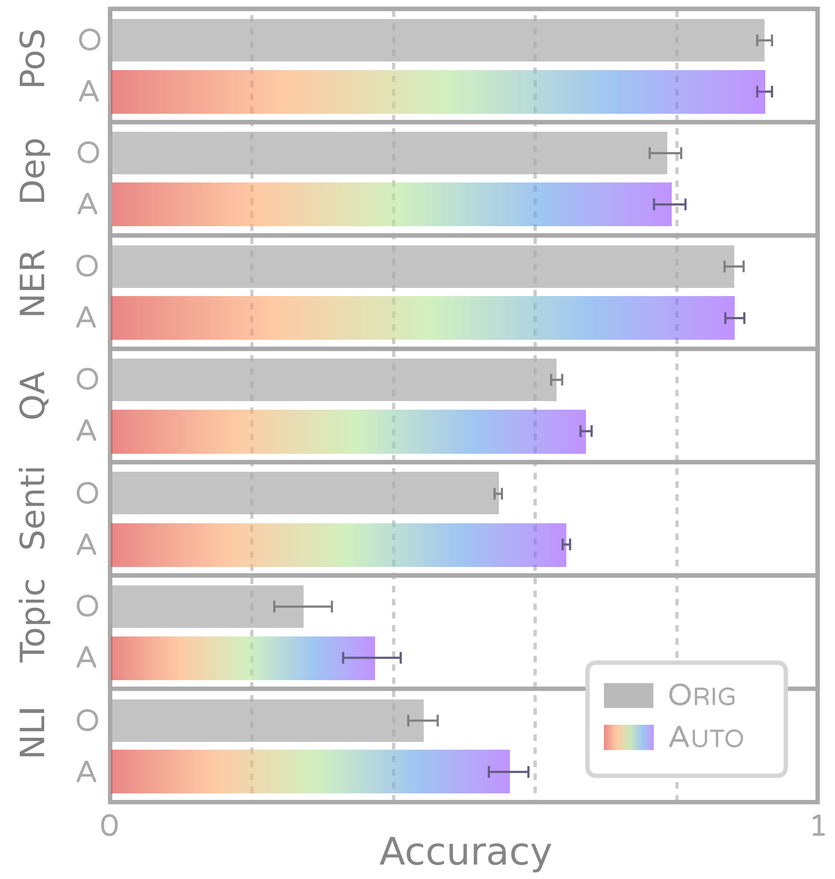


Spectral Probing automatically learns task frequency profiles which match linguistic intuitions and are cross-lingually consistent.

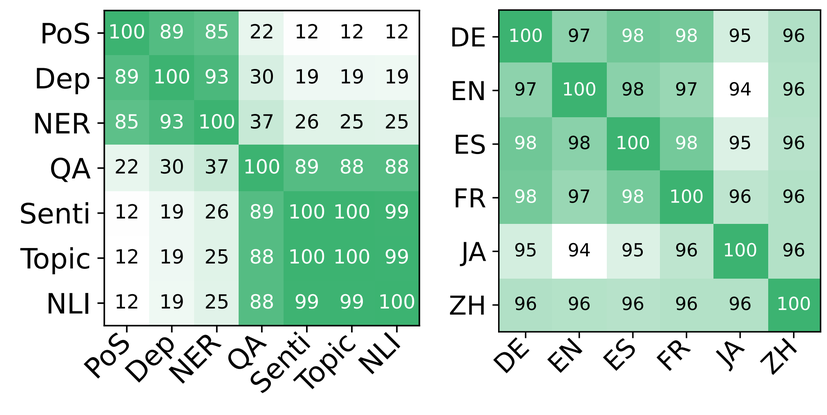
Multilingual Experiments



Spectral Profiles of 7 tasks (weight per frequency), with lower/upper bounds across 6 languages.



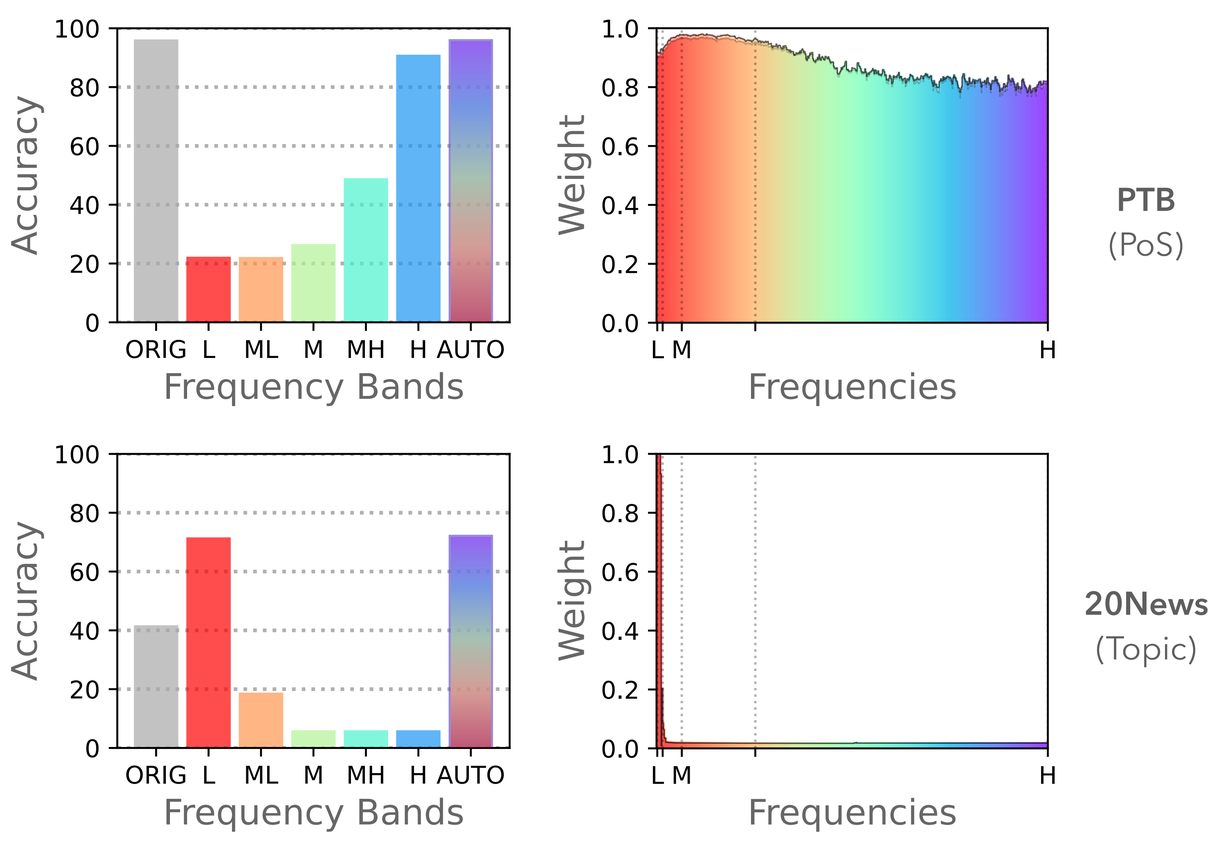
Task-wise Accuracy using original and automatically filtered embeddings (mean over languages).



Filter Overlap across Tasks/Languages as measured in percentage-normalized L1 distance.

Monolingual Experiments

(Reproducing and Extending Tamkin et al., 2020)



Monolingual Results on PTB and 20News. Accuracy of unfiltered (ORIG), low (L), mid-low (ML), mid (M), mid-high (MH), high (H), and the spectral probe's automatic filters (AUTO) with frequency weightings.

PERSONADS.ME/X/EMNLP-2022

mamy@itu.dk, robv@itu.dk, b.plank@lmu.de

IT-UNIVERSITETET I KØBENHAVN / north × LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN / nlp