LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

LMU

CENTRUM FÜR INFORMATIONS- UND SPRACHVERARBEITUNG
STUDIENGANG COMPUTERLINGUISTIK

Thesis proposal

Topic:	**Leveraging conceptual language similarity for cross-lingual transfer learning**
Supervisor:	Haotian Ye
Examiner:	Hinrich Schütze
Level:	BSc

Summary: Traditionally, languages have been categorized based on typologies such as phylogenetic (genealogical) relationships or geographical proximities. [1] suggests that languages *conceptualize* the world differently, dividing it into distinct *concepts*. This divergence offers valuable insights into language similarity. On a related note, language similarity has been shown to be a useful factor for cross-lingual transfer learning [3][4]. In this project, we would like to explore whether conceptual similarity complements existing language similarity measures and positively contributes to cross-lingual transfer performance. We will select transfer languages based on conceptual language vectors [1][2] and evaluate their effectiveness in transferring knowledge to target languages across a range of multilingual datasets.

Requirements: Good programming and data processing skills (preferably using Python), good knowledge of a DL framework (preferably PyTorch) and Transformers, enthusiasm and readiness to publish positive results at a conference/workshop.

References:

[1] Yihong Liu et al. (July 2023). "A Crosslingual Investigation of Conceptualization in 1335 Languages". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, pp. 12969–13000. DOI: 10.18653/v1/2023.acl-long.726. URL: https://aclanthology.org/2023.acl-long.726

[2] Haotian Ye, Yihong Liu, and Hinrich Schütze (2023). "A study of conceptual language similarity: comparison and evaluation". In: *arXiv preprint arXiv:2305.13401*

[3] Yu-Hsiang Lin et al. (July 2019). "Choosing Transfer Languages for Cross-Lingual Learning". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Ed. by Anna Korhonen, David Traum, and Lluís Màrquez. Florence, Italy: Association for Computational Linguistics, pp. 3125–3135. DOI: 10.18653/v1/P19-1301. URL: https://aclanthology.org/P19-1301

[4] Peiqin Lin et al. (Mar. 2024). "mPLM-Sim: Better Cross-Lingual Similarity and Transfer in Multilingual Pretrained Language Models". In: *Findings of the Association for Computational Linguistics: EACL 2024*. Ed. by Yvette Graham and Matthew Purver. St. Julian's, Malta: Association for Computational Linguistics, pp. 276–310. URL: https://aclanthology.org/2024.findings-eacl.20