Identifying the Correlation Between Language Distance and Cross-Lingual Transfer in a Multilingual Representation Space

Fred Philippy [1,2], Siwen Guo [1], Shohreh Haddadan [1]

[1] Zortify Labs, Zortify S.A., Luxembourg
[2] SnT, University of Luxembourg, Luxembourg

SIGTYP Workshop
EACL 2023 – Dubrovnik, Croatia
06.05.2023

zortify

01

Introduction

Cross-Lingual Transfer

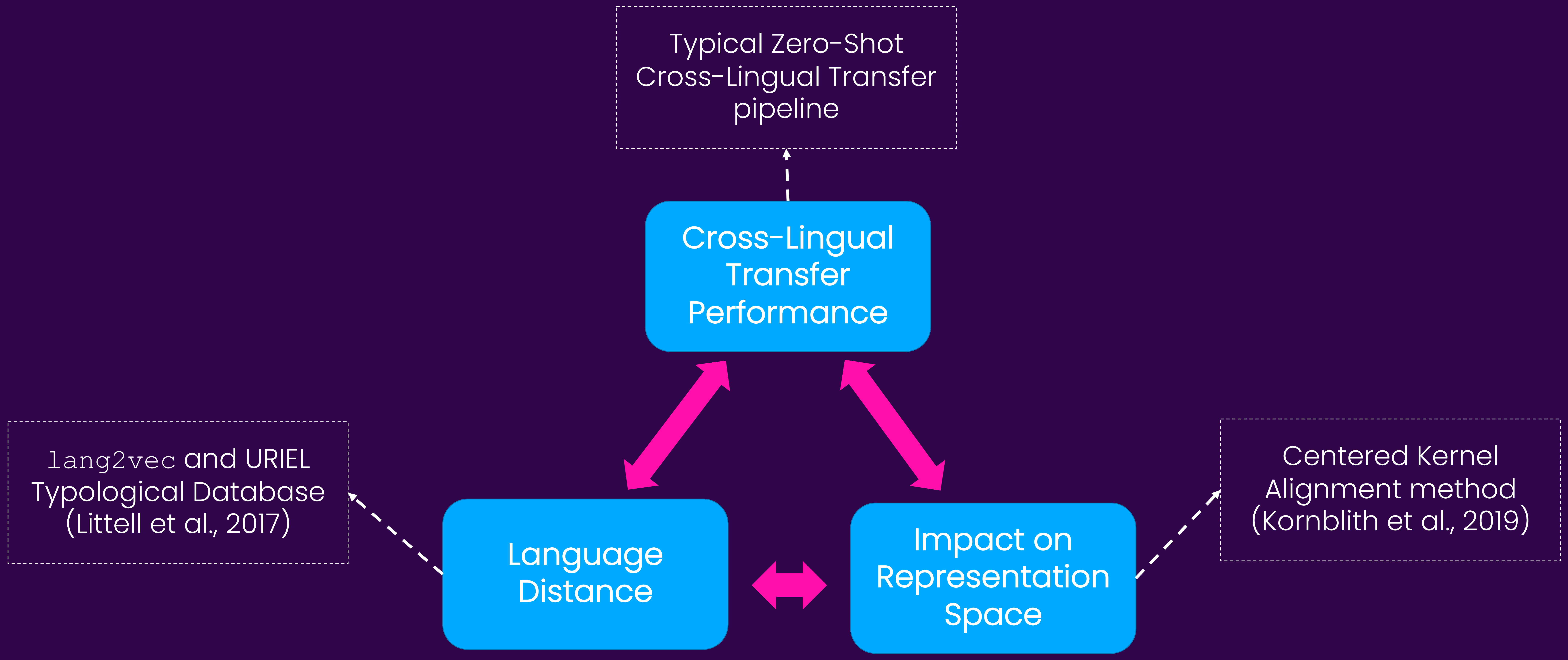
The Method and Motivation of Cross-Lingual Transfer in NLP

- Typical cross-lingual transfer steps
  1. Pre-train multilingual model
  2. Fine-tune model on labelled data in source language
  3. Evaluate/apply fine-tuned model in target language
- Enables better performance for low-resource languages
- Different factors (data size, model architecture, language similarity, etc.) impact cross-lingual transfer performance
- Fine-tuning affects cross-lingual alignment (Singh et al., 2019; Muller et al., 2021)
  → <u>relative impact</u> on representation space
- We focus on the <u>absolute impact</u> on the representation space of a (target) language

zortify

02 Correlation Analysis

Correlation Analysis



zortify

Correlation Analysis: Results

Relationship Between Language Distance and Cross-Lingual Transfer Performance

- All distance metrics correlate with cross-lingual transfer performance

Language Distance	Pearson	Spearman
Syntactic	-0.3193**	-0.4683**
Geographic	-0.3178**	-0.3198**
Inventory	-0.1706*	-0.1329*
Genetic	-0.3364**	-0.3935**
Phonology	-0.2075**	-0.2659**

(*$p < 0.05$, and **$p < 0.01$, two-tailed)

zortify

Correlation Analysis: Results

Relationship Between the Impact on the Representation Space and Cross-Lingual Transfer Performance

Layer	Pearson	Spearman
1	0.2779*	0.3233*
2	0.2456*	0.2639*
3	0.5277*	0.5926*
4	0.3585*	0.3411*
5	-0.009	0.0669
6	0.1033	0.1969
7	0.2945*	0.3500*
8	0.3004*	0.3517*
9	0.4209*	0.4583*
10	0.6088*	0.6532*
11	0.7110*	0.7525*
12	0.5731*	0.5901*
ALL	0.4343*	0.5026*

(*$p < 0.01$, two-tailed)

- Cross-lingual transfer performance correlates with impact on the representation space of the target language. This correlation tends to be stronger in the deeper layers of the model.

zortify

Correlation Analysis: Results

Relationship Between the Impact on the Representation Space and Language Distance

- Almost no significant correlation between representation space impact and **inventory** or **phonological** distance
- **Geographic** and **syntactic** distance mostly show significant correlation values at the last layers
- **Genetic** distance correlates significantly across all layers with the impact on the representation space.

Layer	SYN	GEO	INV	GEN	PHON
1	-0.176*	-0.222**	0.016	-0.19**	-0.186**
2	-0.1	-0.104	0.021	-0.197**	-0.067
3	-0.073	0.054	-0.03	-0.14*	0.005
4	0.051	-0.143*	-0.055	-0.282**	-0.027
5	0.159*	-0.105	-0.028	-0.251**	0.068
6	0.074	-0.118	0.014	-0.202**	0.019
7	-0.001	-0.148*	-0.002	-0.222**	-0.007
8	-0.068	-0.093	-0.015	-0.195**	-0.035
9	-0.107	-0.151*	0.001	-0.245**	-0.051
10	-0.184**	-0.168*	0.033	-0.279**	-0.034
11	-0.262**	-0.175*	0.032	-0.326**	-0.066
12	-0.17*	-0.167*	0.032	-0.291**	-0.047
AVG	-0.091	-0.177*	0.003	-0.307**	-0.045

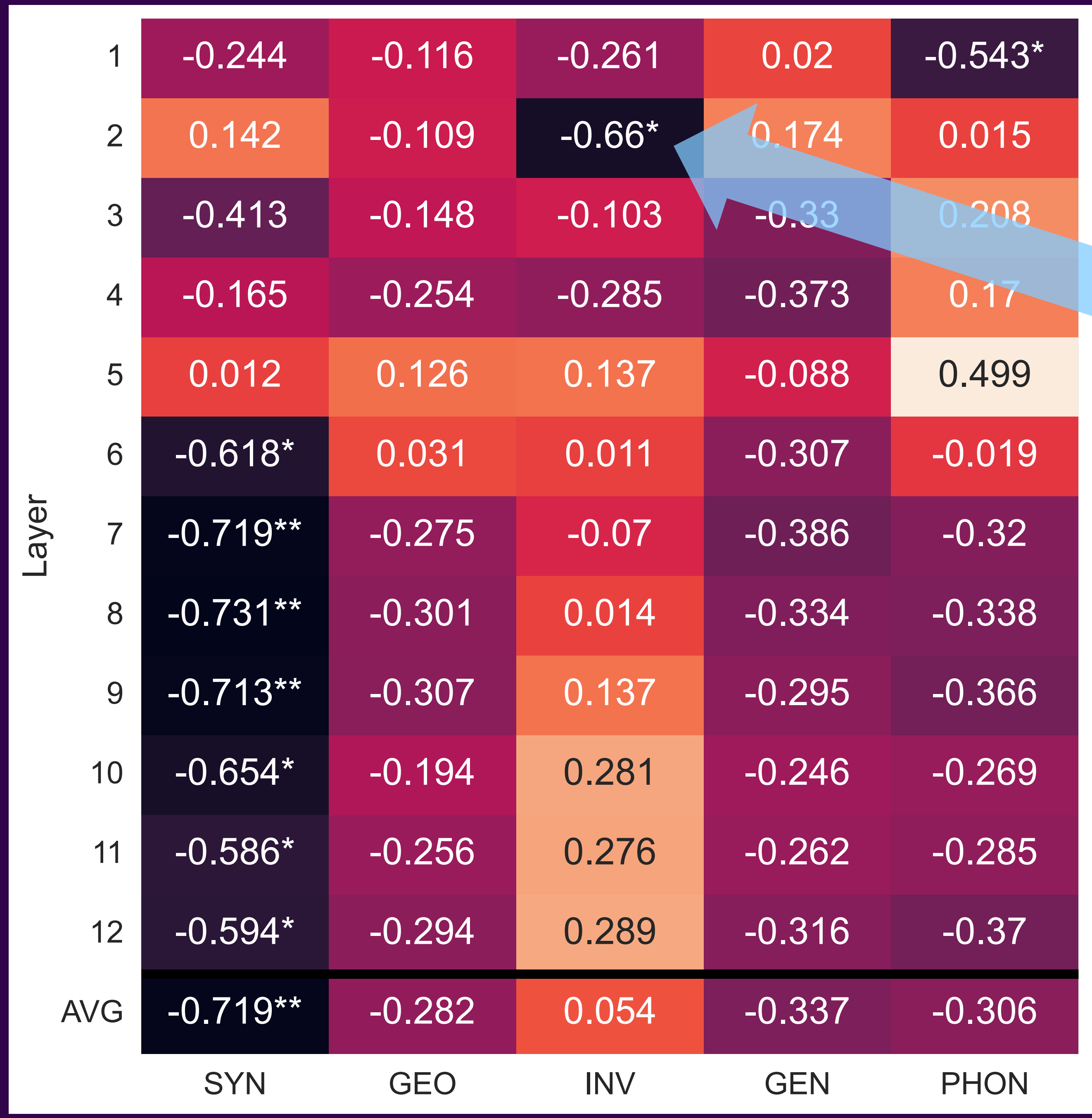
(*$p < 0.05$, and **$p < 0.01$, two-tailed)

03

Does Selective Layer Freezing Allow to Improve Transfer to Linguistically Distant Languages?

Experiment A

Reducing the transfer gap based on a single language distance metric



Frozen Layers	SYN	GEO	INV	GEN	PHON	Transfer Perf.
{}	-0.7354	-0.5109	-0.4907	-0.6116	-0.5776	66.70
{2}	-0.7310	-0.5109	<u>-0.4791</u> (↑)	-0.6009	-0.5791	66.53
{5}	-0.7438	-0.5053	-0.4897	-0.6148	-0.5896	66.77
{1,2,6}	-0.7325	-0.5000	-0,4846	-0.6065	-0.5666	66.75

Correlation between **cross-lingual transfer performance** and different language distance metrics after freezing different layers

zortify

Experiment B

Increasing the transfer gap based on a single language distance metric

Layer	SYN	GEO	INV	GEN	PHON
1	-0.244	-0.116	-0.261	0.02	-0.543*
2	0.142	-0.109	-0.66*	0.174	0.015
3	-0.413	-0.148	-0.103	-0.33	0.208
4	-0.165	-0.254	-0.285	-0.373	0.17
5	0.012	0.126	0.137	-0.088	0.499
6	-0.618*	0.031	0.011	-0.307	-0.019
7	-0.719**	-0.275	-0.07	-0.386	-0.32
8	-0.731**	-0.301	0.014	-0.334	-0.338
9	-0.713**	-0.307	0.137	-0.295	-0.366
10	-0.654*	-0.194	0.281	-0.246	-0.269
11	-0.586*	-0.256	0.276	-0.262	-0.285
12	-0.594*	-0.294	0.289	-0.316	-0.37
AVG	-0.719**	-0.282	0.054	-0.337	-0.306

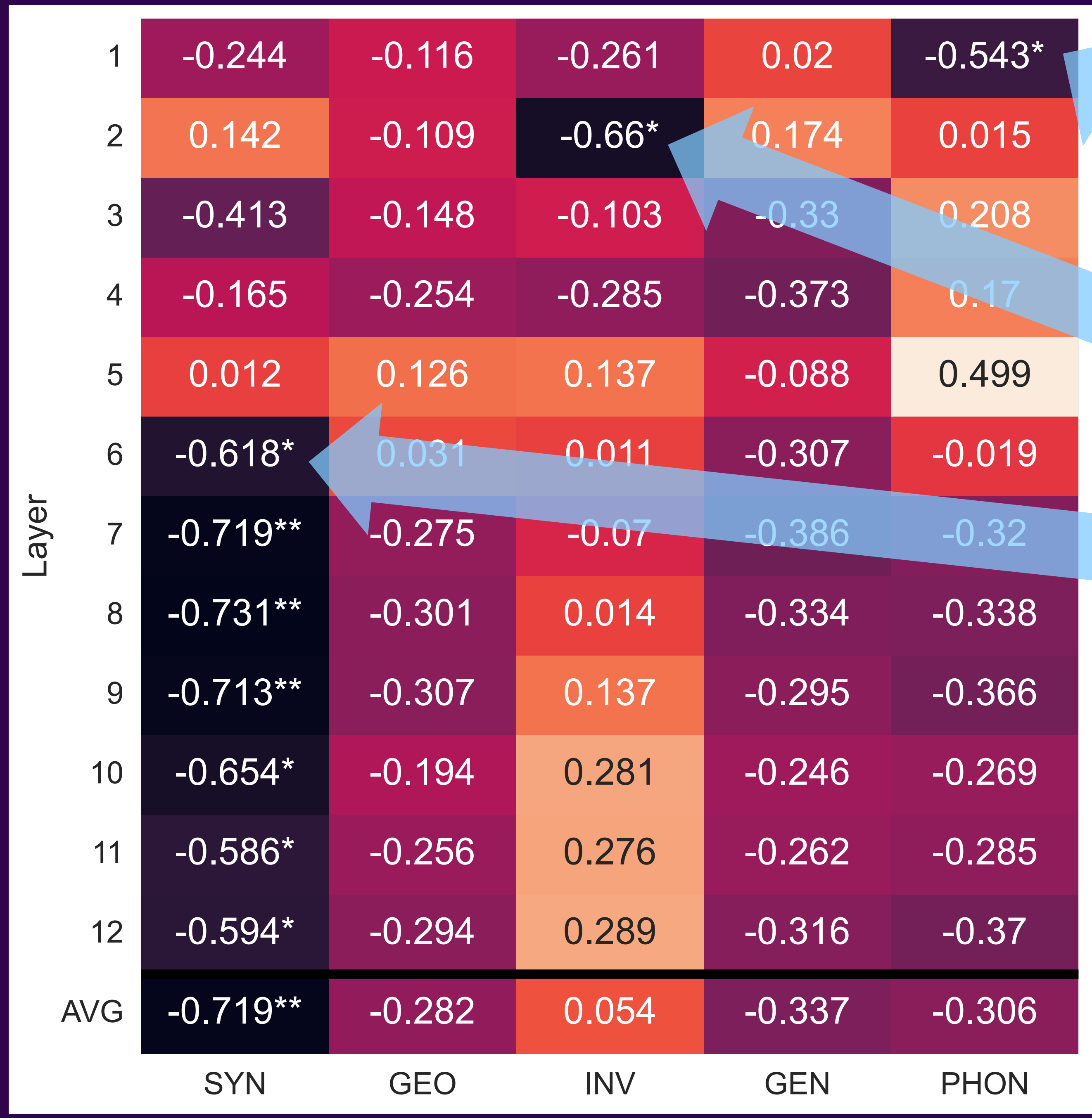
Frozen Layer	SYN	GEO	INV	GEN	PHON	Transfer Perf.
{}	-0.7354	-0.5109	-0.4907	-0.6116	-0.5776	66.70
{2}	-0.7310	-0.5109	-0.4791	-0.6009	-0.5791	66.53
{5}	-0.7438	-0.5053	-0.4897	-0.6148	<u>-0.5896</u> (↓)	66.77
{1,2,6}	-0.7325	-0.5000	-0,4846	-0.6065	-0.5666	66.75

Correlation between **cross-lingual transfer performance** and different language distance metrics after freezing different layers

zortify

Experiment C

Reducing the transfer gap based on multiple language distance metrics



Frozen Layer	SYN	GEO	INV	GEN	PHON	Transfer Perf.
{}	-0.7354	-0.5109	-0.4907	-0.6116	-0.5776	66.70
{2}	-0.7310	-0.5109	-0.4791	-0.6009	-0.5791	66.53
{5}	-0.7438	-0.5053	-0.4897	-0.6148	-0.5896	66.77
{1,2,6}	<u>-0.7325</u> (↑)	-0.5000	<u>-0.4846</u> (↑)	-0.6065	<u>-0.5666</u> (↑)	66.75

Correlation between **cross-lingual transfer performance** and different language distance metrics after freezing different layers

04 Conclusion

Conclusion

- New perspective on the representation space dynamics during cross-lingual transfer
- Inter-correlation between language distance, representation space impact and cross-lingual transfer performance
- Hypothesis: By selectively freezing layers, based on the observed correlations, languages that exhibit specific linguistic features can be targeted for better transfer

References

Patrick Littell, David R. Mortensen, Ke Lin, Katherine Kairis, Carlisle Turner, and Lori Levin. 2017. URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. In Proceed- ings of the 15th Conference of the European Chap- ter of the Association for Computational Linguistics: Volume 2, Short Papers, pages 8–14, Valencia, Spain. Association for Computational Linguistics.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, 361 and Geoffrey Hinton. 2019. Similarity of Neural 362 Network Representations Revisited. In Proceedings 363 of the 36th International Conference on Machine 364 Learning, pages 3519–3529. PMLR.

Jasdeep Singh, Bryan McCann, Richard Socher, and 426 Caiming Xiong. 2019. BERT is Not an Interlingua 427 and the Bias of Tokenization. In Proceedings of the 428 2nd Workshop on Deep Learning Approaches for 429 Low-Resource NLP (DeepLo 2019), pages 47–55, 430 Hong Kong, China. Association for Computational 431 Linguistics.

Benjamin Muller, Yanai Elazar, Benoît Sagot, and Djamé Seddah. 2021. First Align, then Predict: Un- derstanding the Cross-Lingual Ability of Multilin- gual BERT. In Proceedings of the 16th Conference of the European Chapter of the Association for Compu- tational Linguistics: Main Volume, pages 2214–2231, Online. Association for Computational Linguistics.

Thank you

Email:

fred@zortify.com