GRADUAL LANGUAGE MODEL ADAPTATION USING FINE-GRAINED TYPOLOGY

AALBORG UNIVERSITET
DEPARTMENT OF COMPUTER SCIENCE

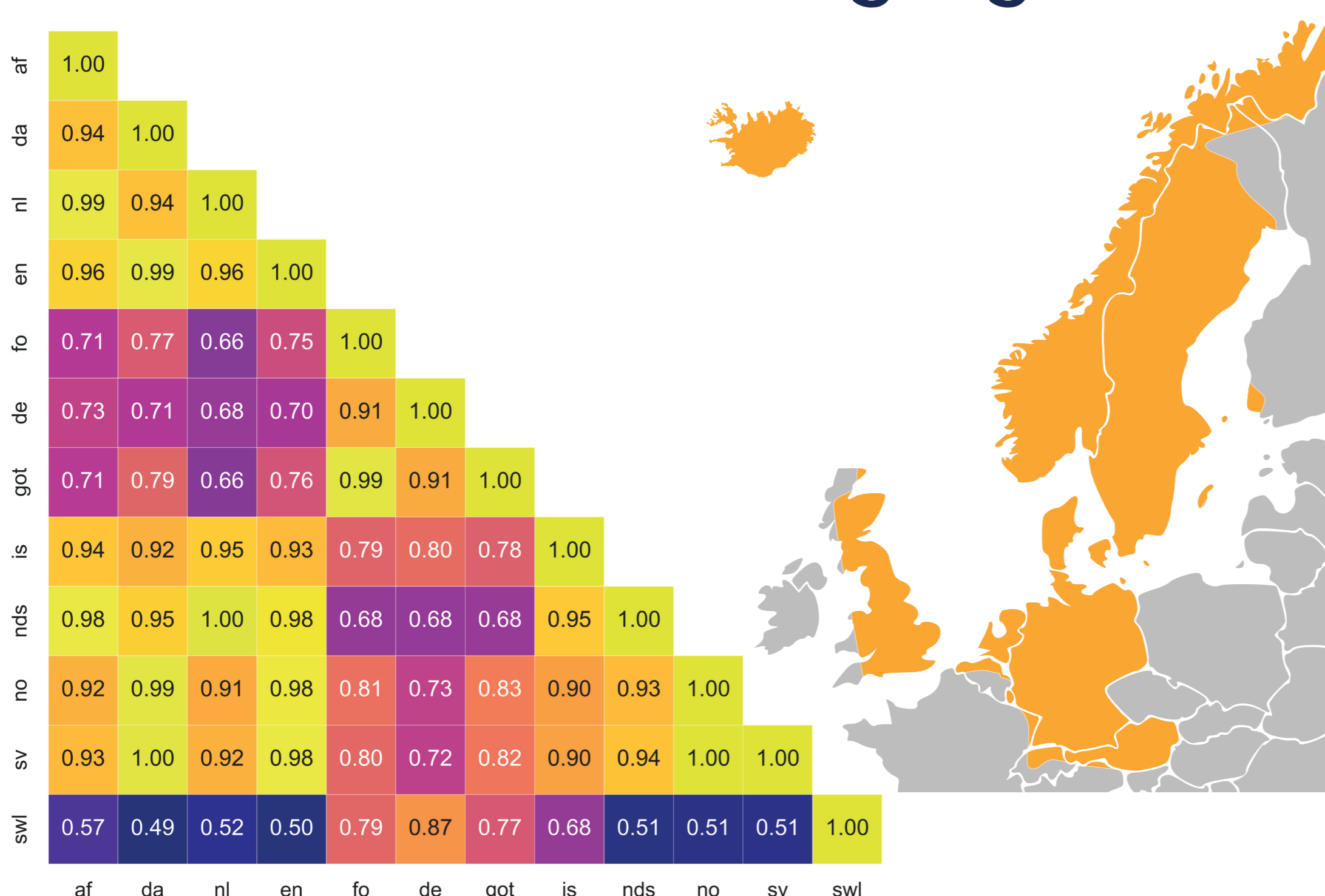**Marcell Richard Fekete**, Johannes Bjerva

✉ mrfe@cs.aau.dk

❶ Introduction

- Most languages lack the data to allow pretraining well-performing language models [1]

- Multilingual models may struggle when it comes to novel languages

- Monolingual and multilingual **model adaptation techniques** often do not consider the relation of languages beyond in-family membership [2]

- Using information about typological similarity between languages, **cross-lingual transfer** may be facilitated to the benefit of model adaptation to **resource-poor languages**

❷ Typological similarity

- Typological factors between languages seem to influence the success of **cross-lingual transfer**

- Typological properties in prior work are mostly derived from resources such as WALS [3]

- Such resources do not allow granularity in the quantification of typological properties [4]

- Instead we derive **fine-grained typological properties** using structural vectors [5]

- Structural vectors have been created from **counts of dependency links** taken from the Germanic subset of Universal Dependencies

Typological similarity of Germanic languages

	af	da	nl	en	fo	de	got	is	nds	no	sv	swl
af	1.00											
da	0.94	1.00										
nl	0.99	0.94	1.00									
en	0.96	0.99	0.96	1.00								
fo	0.71	0.77	0.66	0.75	1.00							
de	0.73	0.71	0.68	0.70	0.91	1.00						
got	0.71	0.79	0.66	0.76	0.99	0.91	1.00					
is	0.94	0.92	0.95	0.93	0.79	0.80	0.78	1.00				
nds	0.98	0.95	1.00	0.98	0.68	0.68	0.68	0.95	1.00			
no	0.92	0.99	0.91	0.98	0.81	0.73	0.83	0.90	0.93	1.00		
sv	0.93	1.00	0.92	0.98	0.80	0.72	0.82	0.90	0.94	1.00	1.00	
swl	0.57	0.49	0.52	0.50	0.79	0.87	0.77	0.68	0.51	0.51	0.51	1.00

❸ Gradual adaptation

- Instead of **direct adaptation** from source to target language, carry out **model adaptation in stages** using intermediate languages

- By ensuring a high degree of typological similarity between languages, the aim is **to catalyse cross-lingual transfer** across the model adaptation stages

- This may contribute to a **more data-efficient model adaptation** process

- Especially favouring **typologically diverse** and **resource-poor languages**

Comparison of direct and gradual adaptation

Direct Adaptation   Gradual Adaptation
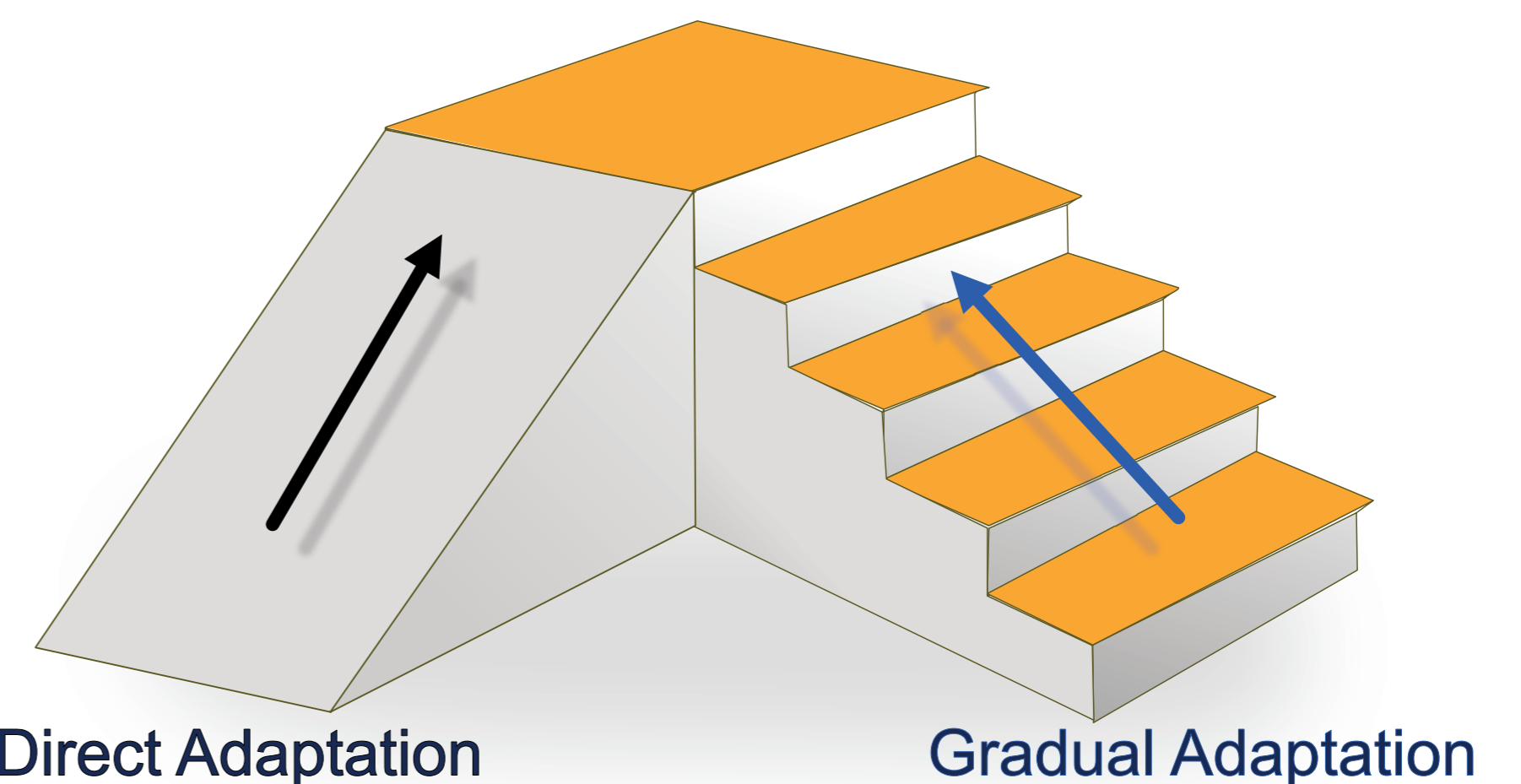
❹ Future work

- Extend the scope to all languages in Universal Dependencies

- Assess the success of gradual adaptation using various tasks such as POS-tagging, dependency parsing and more

- Apply gradual adaptation to different model adaptation techniques (e.g. adapters, retraining of lexical layer)

- Outcome: a flexible and promising framework of **typologically-informed model adaptation**

❺ References

[1] **Joshi et al.** (2021) *The State and Fate of Linguistic Diversity and Inclusion in the NLP World.*

[2] **Snæbjarnarson et al.** (2023) *Transfer to Low-Resource Language via Close Relatives: The Case Study on Faroese.*

[3] **Üstün et al.** (2022) *UDapter: Typology-based Language Adapters for Multilingual Dependency Parsing and Sequence Labeling.*

[4] **Ponti et al.** (2019) *Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing.*

[5] **Bjerva et al.** (2019) *What Do Language Representations Really Represent?*