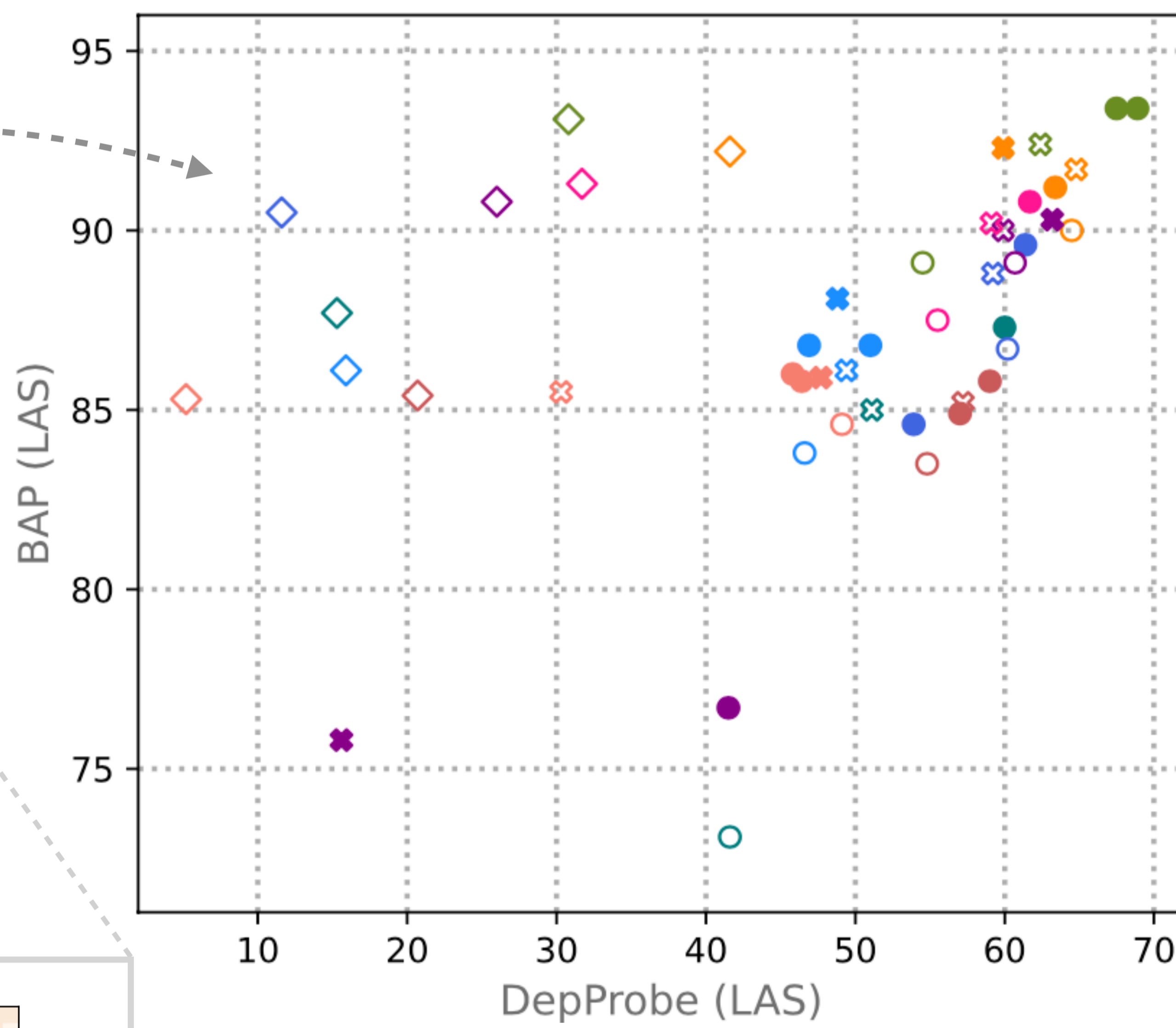
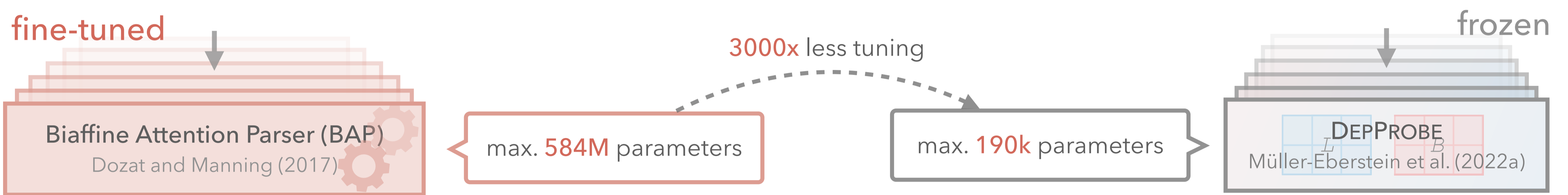


Sort by Structure: Language Model Ranking as Dependency Probing

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

How can we know which LM will perform best w/o tuning?

Arabic	English	Finnish	Anc. Greek	Hebrew	Korean	Russian	Swedish	Chinese
mBERT	mBERT	mBERT	mBERT	mBERT	mBERT	mBERT	mBERT	mBERT
XLM-R	XLM-R	XLM-R	XLM-R	XLM-R	XLM-R	XLM-R	XLM-R	XLM-R
RemBERT	RemBERT	RemBERT	RemBERT	RemBERT	RemBERT	RemBERT	RemBERT	RemBERT
AraBERT	BERT	BERT-FI	BERT-GRC	χ -BERT	BERT-KO	RuBERT	BERT-SV	BERT-ZH
BERT-AR	RoBERTA	BERT-fi	BERT-EL		RoBERTA-KO	RuBERTa		BERT-ZH WWM
					BERT-KOR	RoBERTA-RU		RoBERTA-ZH WWM

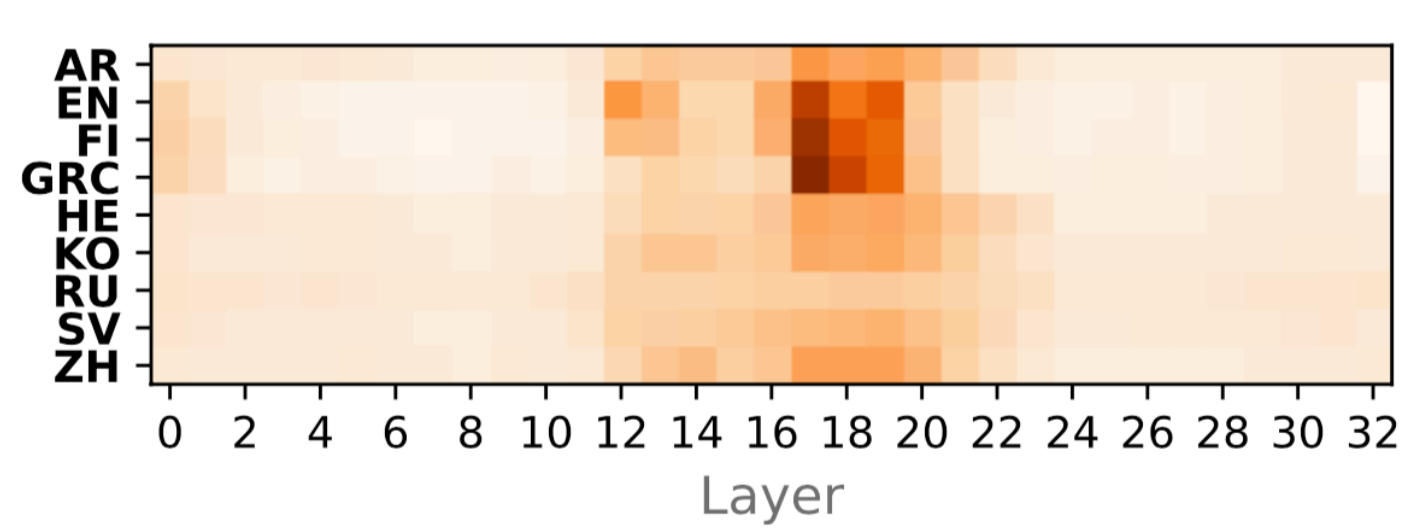
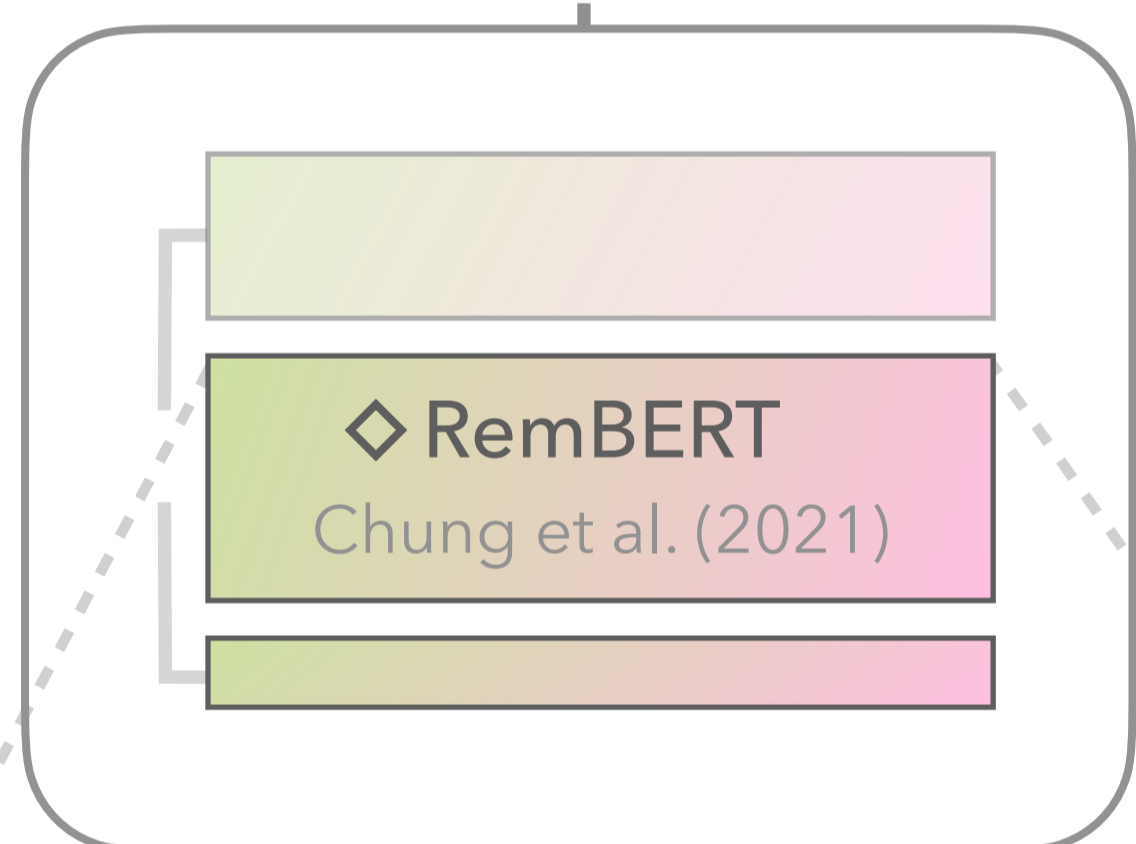


- mBERT
- ⊗ XLM-R
- ◇ RemBERT
- *BERT
- * *RoBERTA

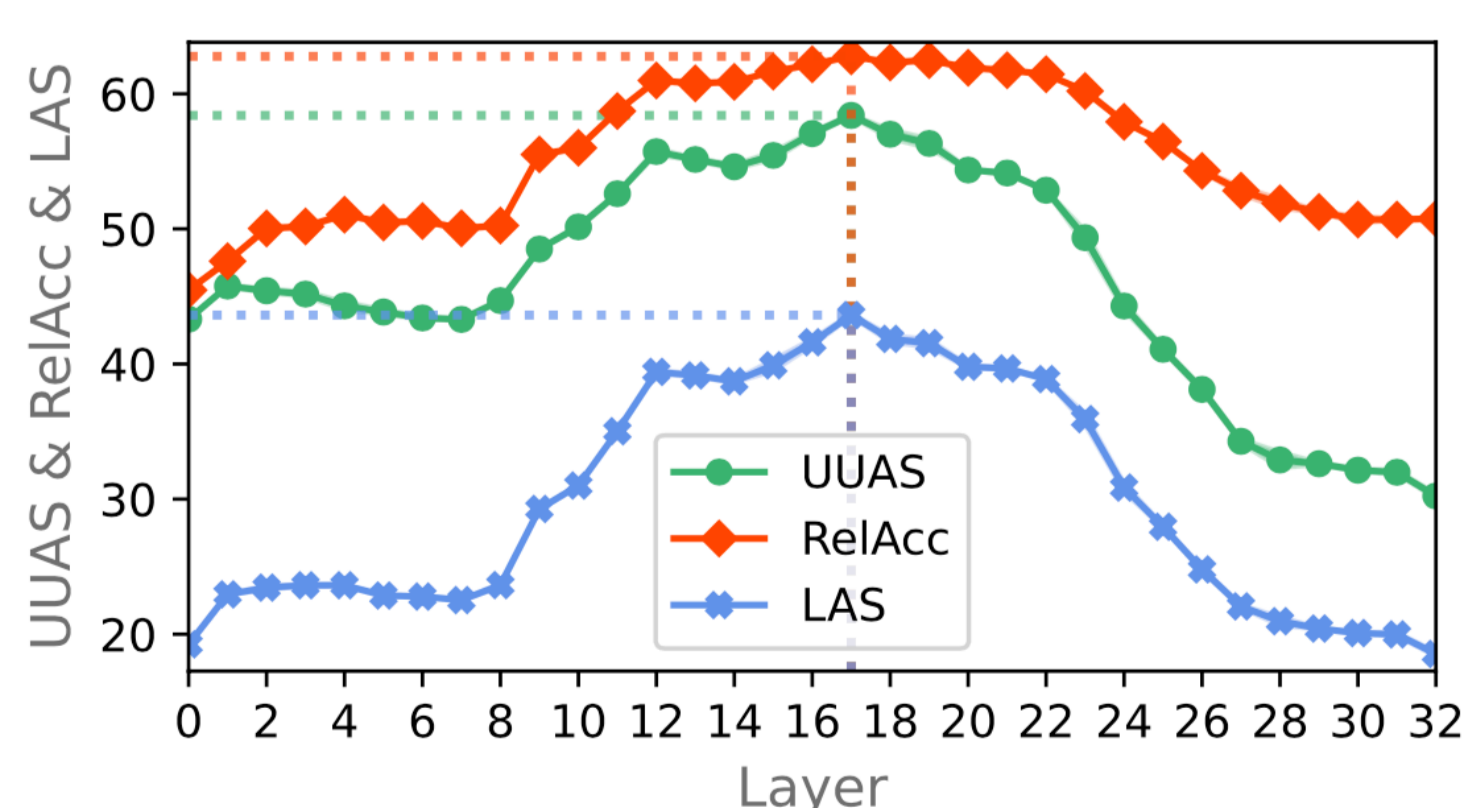
- AR
- EN
- FI
- GRC
- HE
- KO
- RU
- SV
- ZH

DEPPROBE's Predictive Power

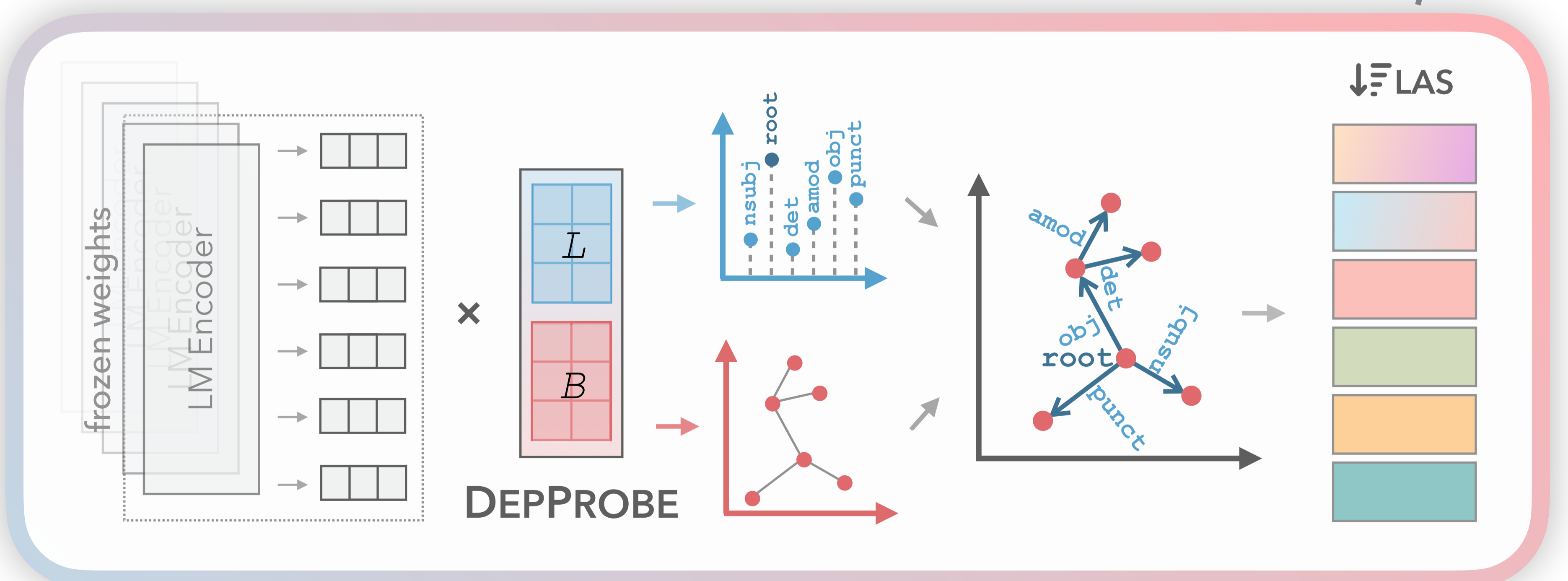
$$\mathcal{T}_w .58 \rightarrow 79\%$$



Per-language α of RemBERT Layers for DEPPROBE (dark > light).



Dependency Information per RemBERT Layer via DEPPROBE's UUAS, RelAcc, LAS on EN-EWT.



Probing to rank LMs provides a quantitative method for encoder pre-selection.



PERSONADS.ME/X/NAACL-2022

@mxmeij, @robvanderg and @barbara_plank
{mamy, robv, bapl}@itu.dk

IT-UNIVERSITETET I KØBENHAVN

