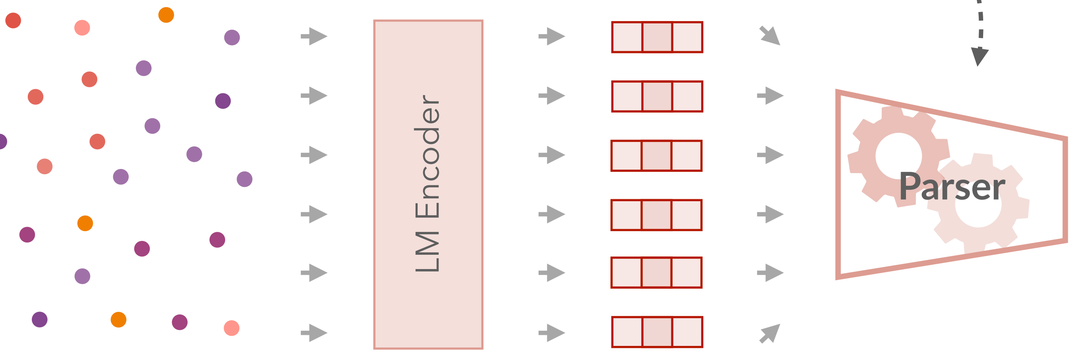
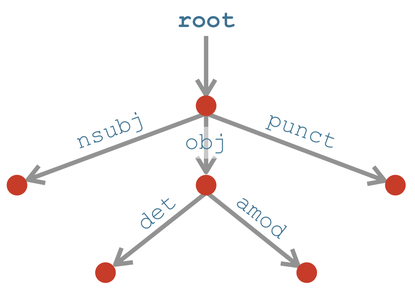
Probing for Labeled Dependency Trees

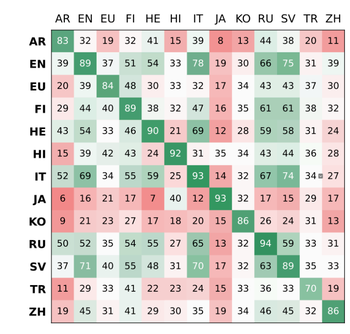
Max **Müller-Eberstein**, Rob **van der Goot** and Barbara **Plank**

Training a full biaffine attention parser (BAP) on mBERT: 183M parameters.





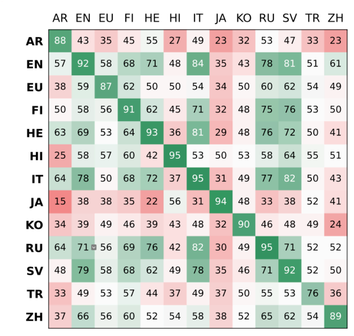
How accurately can these 169 transfer settings be predicted w/o training a full parser?



	AR	EN	EU	FI	HE	HI	IT	JA	KO	RU	SV	TR	ZH
AR	83	32	19	32	41	15	39	8	13	44	38	20	11
EN	39	89	37	51	54	33	78	19	30	66	75	31	39
EU	20	39	84	48	30	33	32	17	34	43	43	37	30
FI	29	44	40	89	38	32	47	16	35	61	61	38	32
HE	43	54	33	46	90	21	69	12	28	59	58	31	24
HI	15	39	42	43	24	92	31	35	34	43	44	36	28
IT	52	69	34	55	59	25	93	14	32	67	74	34	27
JA	6	16	21	17	7	40	12	93	32	17	15	29	17
KO	9	21	23	27	17	18	20	15	86	26	24	31	13
RU	50	52	35	54	55	27	65	13	32	94	59	33	31
SV	37	71	40	55	48	31	70	17	32	63	89	35	33
TR	11	29	33	41	22	23	24	15	33	36	33	70	19
ZH	19	45	31	41	29	30	35	19	34	46	45	32	86

BAP

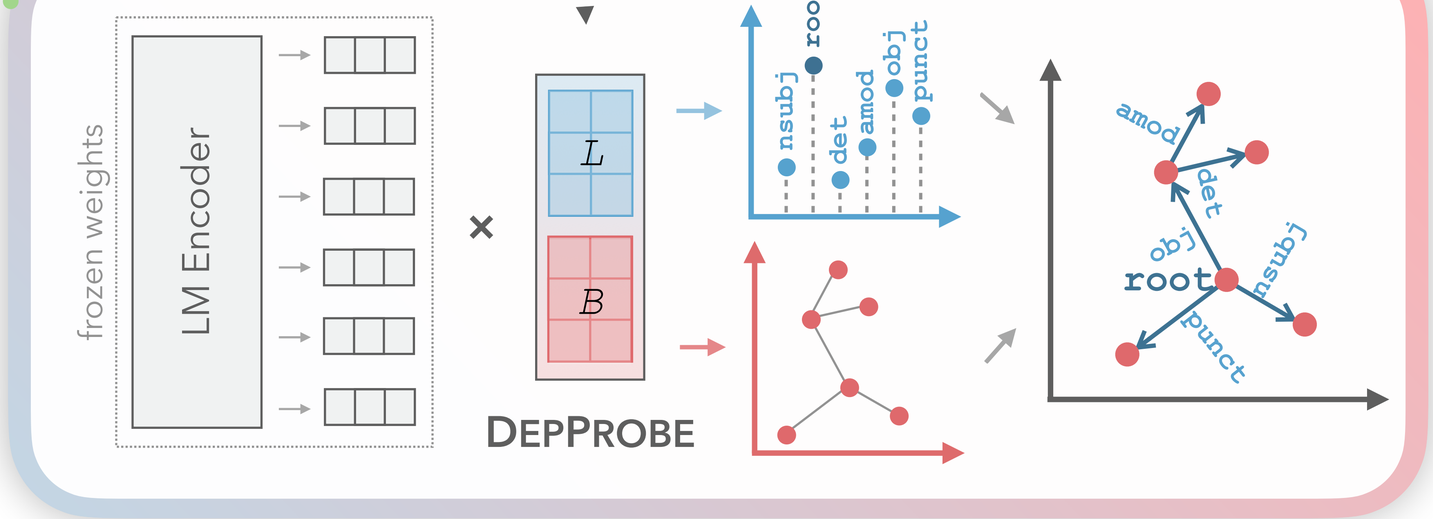
Dozat and Manning (2017)



	AR	EN	EU	FI	HE	HI	IT	JA	KO	RU	SV	TR	ZH
AR	88	43	35	45	55	27	49	23	32	53	47	33	23
EN	57	92	58	68	71	48	84	35	43	78	81	51	61
EU	38	59	87	62	50	50	54	34	50	60	62	54	49
FI	50	58	56	91	62	45	71	32	48	75	76	53	50
HE	63	69	53	64	93	36	81	29	48	76	72	50	41
HI	25	58	57	60	42	95	53	50	53	58	64	55	51
IT	64	78	50	68	72	37	95	31	49	77	82	50	43
JA	15	38	38	35	22	56	31	94	48	33	38	52	41
KO	34	39	49	46	39	43	48	32	90	46	48	49	24
RU	64	71	56	69	76	42	82	30	49	95	71	52	52
SV	48	79	58	68	62	49	78	35	46	71	92	52	50
TR	33	49	53	57	44	37	49	37	50	55	53	76	36
ZH	37	66	56	60	52	54	58	38	52	65	62	54	89

METHOD

Extracting labeled + directed dependency trees w/ DEPPROBE: <3 matrices (127k parameters).

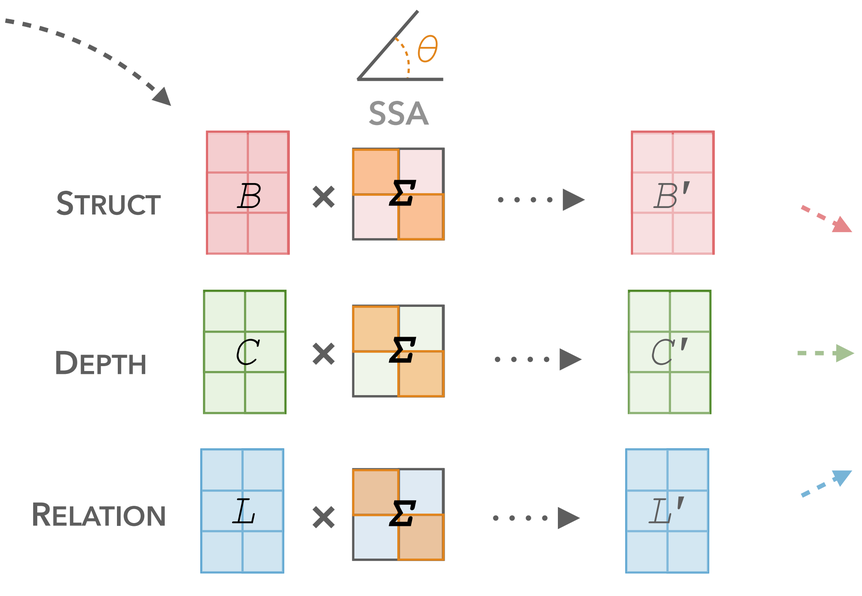


DEPPROBE selects the best source 94% of the time.

Model	LAS ρ	LAS τ_w	UAS ρ	UAS τ_w
LANG2VEC	.86	.72	.80	.70
DIRPROBE	–	–	.91	.81
DEPPROBE	**.97**	**.88**	.94	.85

Transfer Correlation with BAP. Pearson ρ and weighted Kendall's τw for BAP's LAS and UAS w.r.t. DIRPROBE's UAS (Kulmizev et al., 2020), DEPPROBE's UAS and LAS as well as lang2vec (Little et al., 2017) cosine similarity.

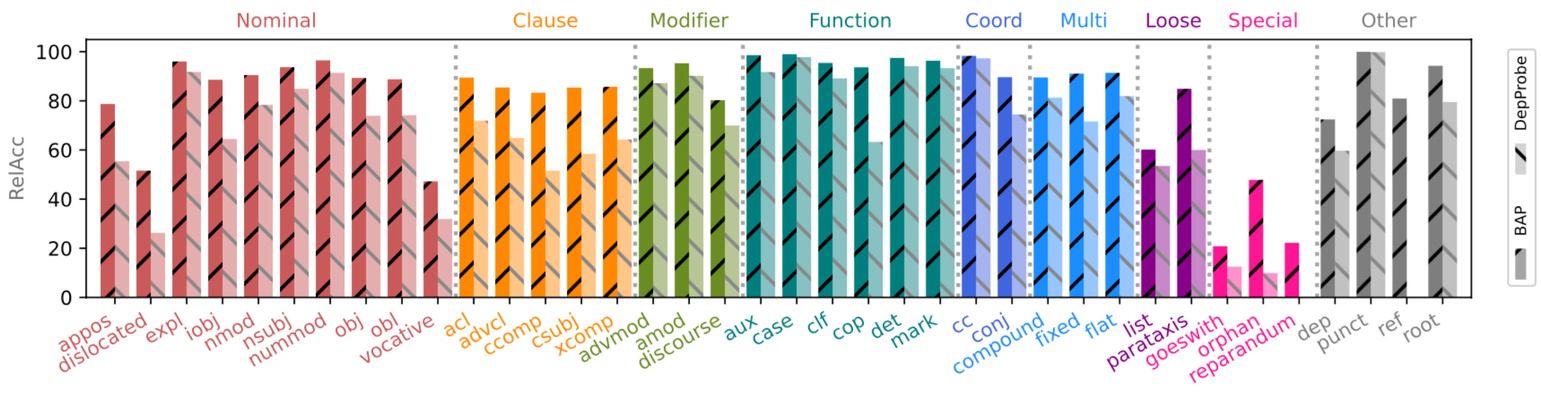
Each probe is just one matrix, so all relevant parameters across languages can be compared at once!



Which type of dependency information is relevant for cross-lingual transfer?

Probe Type	LAS ρ	LAS τ_w	UAS ρ	UAS τ_w
SSA-STRUCT	.68	.42	.60	.43
SSA-DEPTH	.62	.34	.53	.35
SSA-REL	**.73**	**.55**	.65	.53

SSA Correlation with BAP. ρ and τw w.r.t. subspace angles of structural (Hewitt and Manning, 2019; STRUCT), depth (DEPTH) and relational probes (REL).



Relation Accuracy of BAP and DEPPROBE compared for all 13 in-language targets, grouped according to the Universal Dependencies taxonomy (de Marneffe et al., 2014).

How much relational information do tuned/untuned embeddings contain?

PERSONADS.ME/X/ACL-2022

@mxmeij, @robvanderg and @barbara_plank

{mamy, robv, bapl}@itu.dk

IT-UNIVERSITETET I KØBENHAVN / north