

# CL-MoNoise: Cross-lingual Lexical Normalization.

Rob van der Goot

IT UNIVERSITY OF CPH



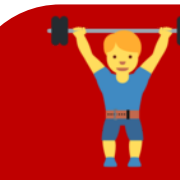
## Lexical Normalization

Lang.	Language name	Normalization example
DA	Danish	De skarpe lamper gjorde destromindre ek bedre . De skarpe lamper gjorde destro mindre ikke bedre .
DE	German	ogäj isch hätts auch dwwidern könn Okay ich hätte es auch twittern können
EN	English	u hve to let ppl decide what dey want to do you have to let people decide what they want to do
ES	Spanish	@username cuuxamee sii peroov veen yaa eem @username escúchame sí pero ven ya eh
HR	Croatian	svi frendovi mi nešto rade , veceras san osta sam . svi frendovi mi nešto rade , večeras sam ostao sam .
ID-EN	Indonesian-English	pdhal not fully bcs those ppl jg sih . padahal not fully because those people juga sih .
IT	Italian	a Roma è cosí primavera che sembra gia giov a Roma è così primavera che sembra già giovedì
NL	Dutch	Kga me wss trg rolle vant lachn Ik ga me waarschijnlijk terug rollen van het lachen
SL	Slovenian	jst bi tud najdu kovanec vreden veliko denarja . jaz bi tudi našel kovanec vreden veliko denarja .
SR	Serbian	komunalci kace pocne kaznjavanje ? komunalci kad počne kažnjavanje ?
TR	Turkish	He o dediyin suala cvb verdim He o dediğin suale cevap verdim
TR-DE	Turkish-German	@username Yerimm senii , damkee schatzymm :-* @username Yerim seni , danke Schatzym :-*



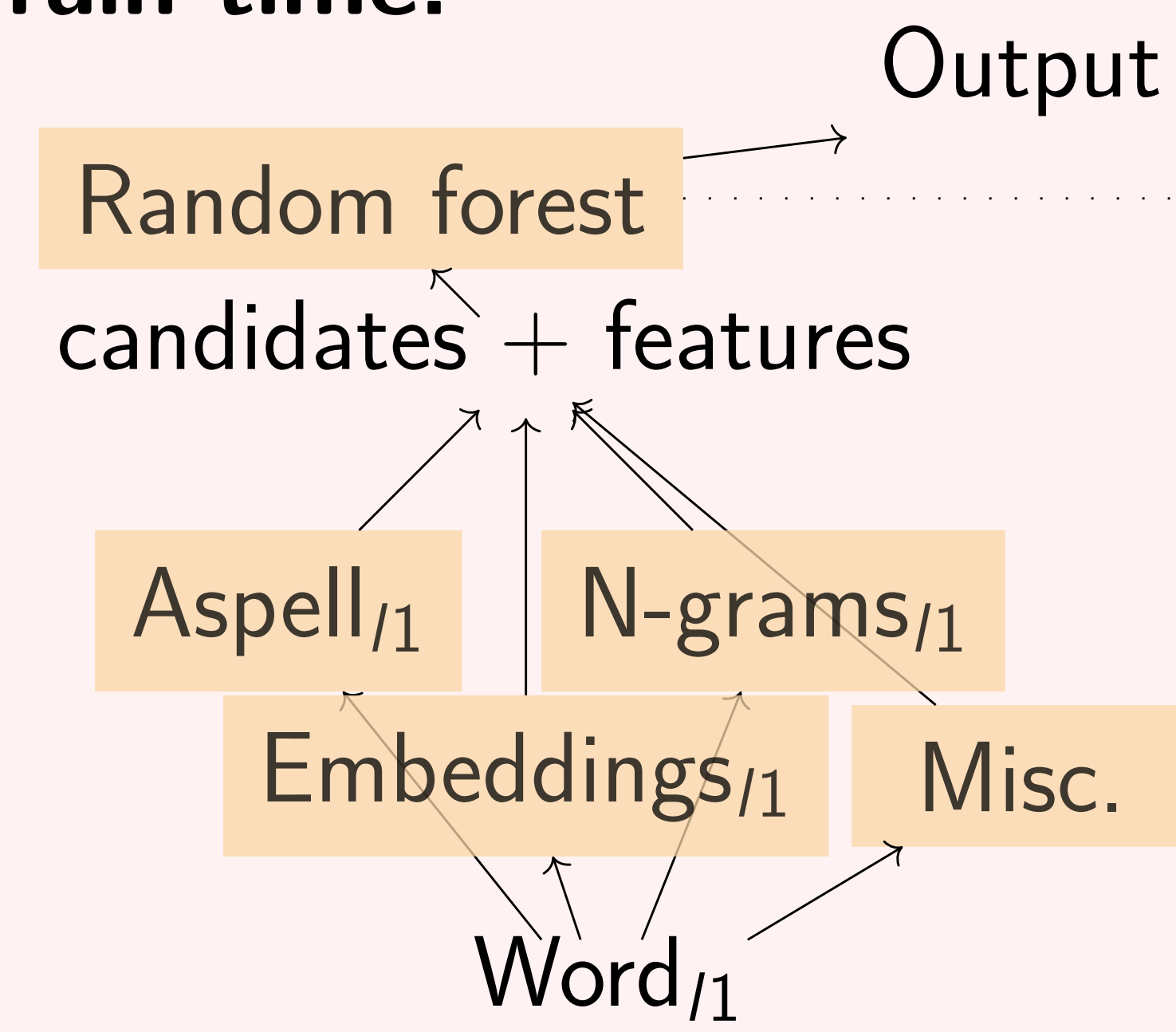
## Matching

tgt	src	src
	+look	-look
DA	SL	SL
DE	TR-DE	TR
EN	IT	TR
ES	IT	TR-DE
HR	SR	SR
ID-EN	TR	TR
IT	EN	TR
NL	TR	TR
SL	SR	HR
SR	HR	HR
TR	TR-DE	TR-DE
TR-DE	TR	TR

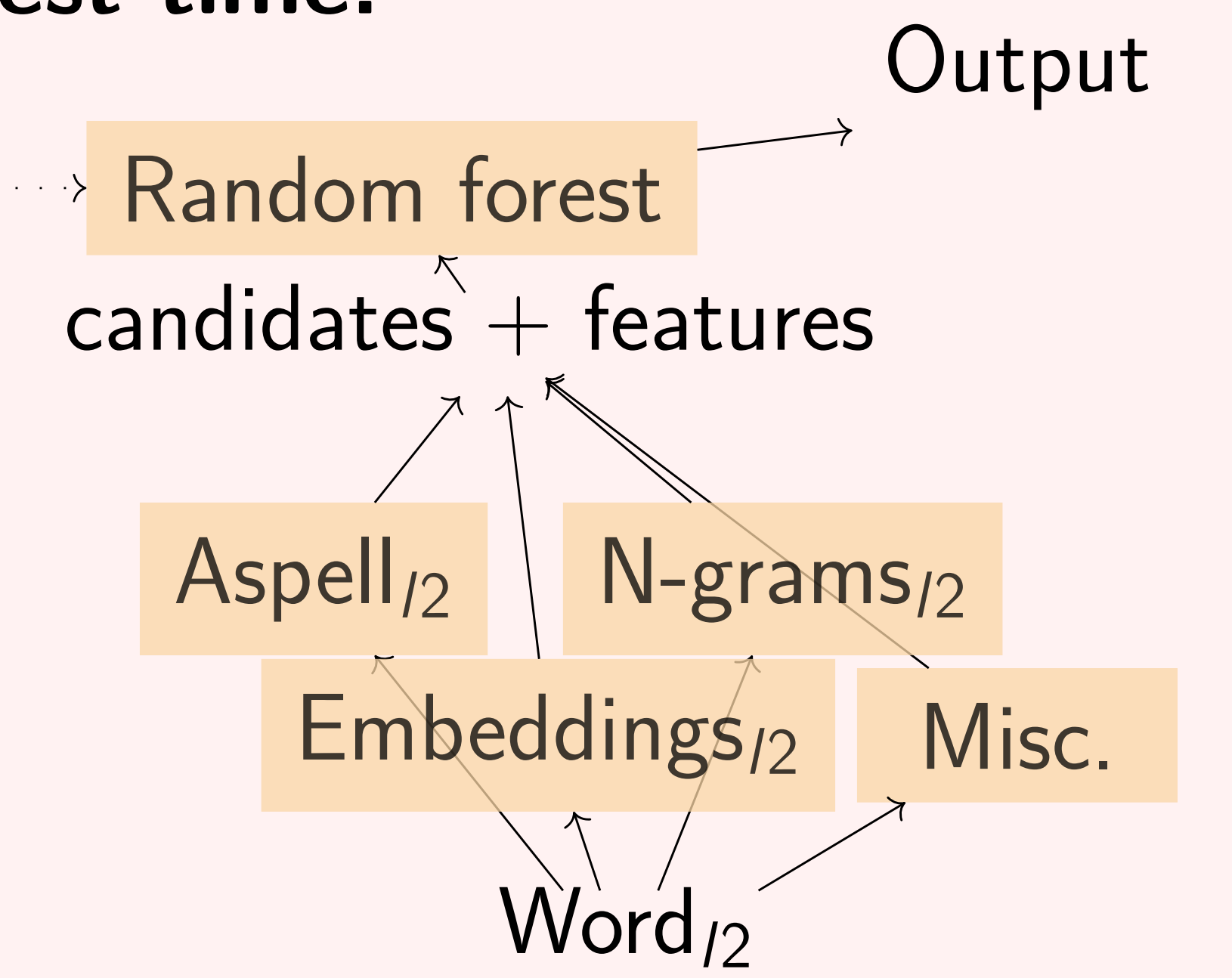


## CL-MoNoise

Train time:

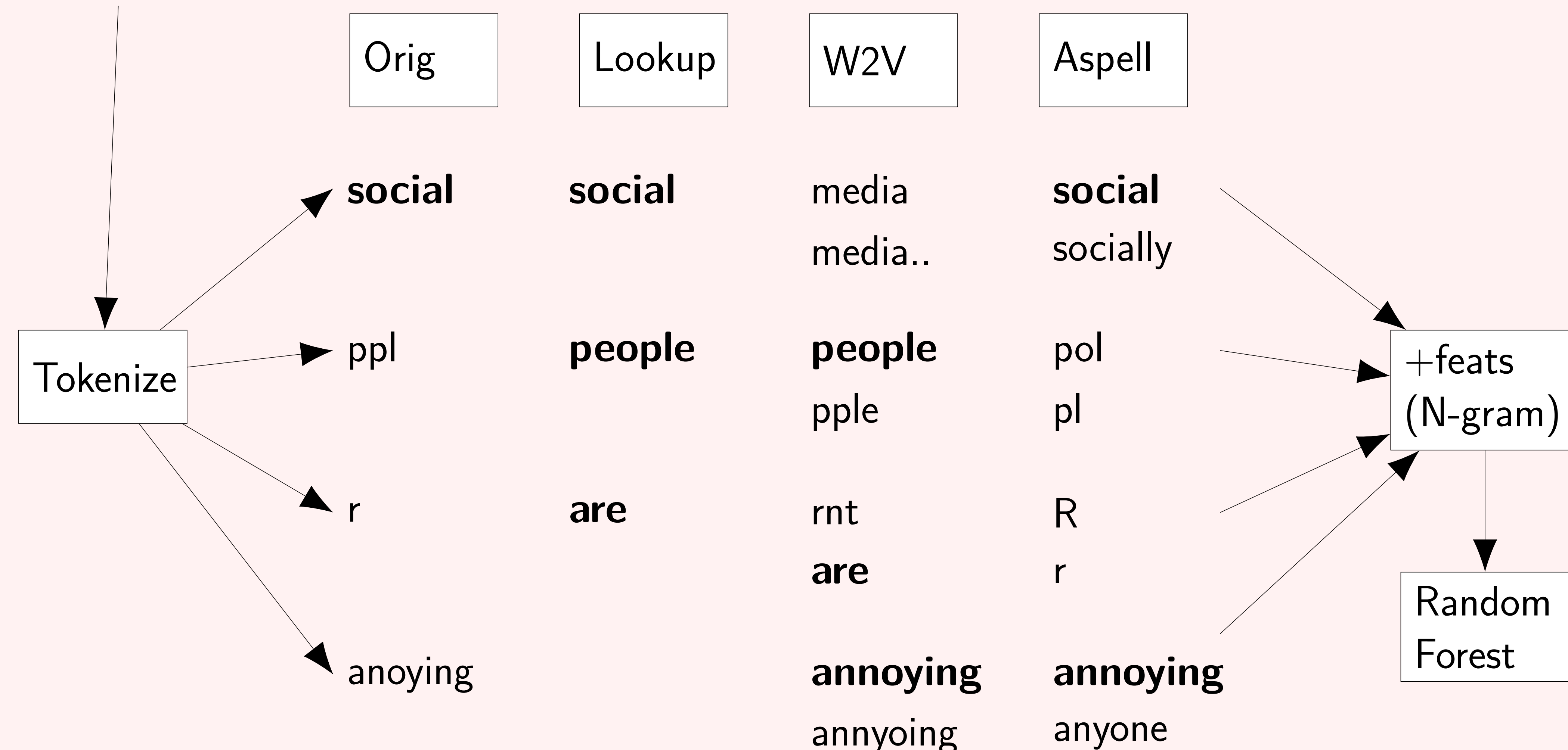


Test time:

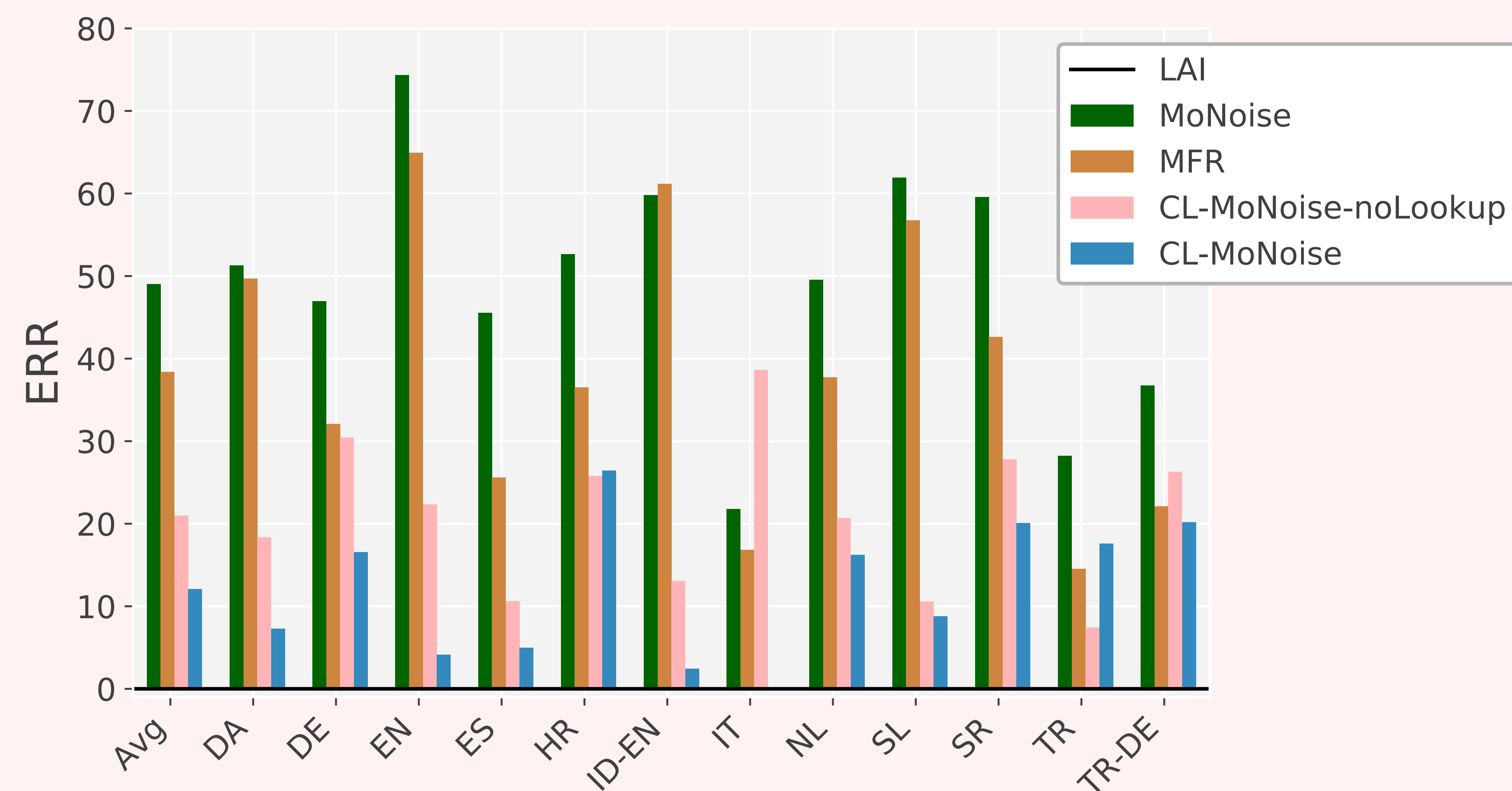


## MoNoise

social ppl r anoying



## Results



## Conclusion

- First cross-lingual lexical normalization model
- Best transfer language is hard to predict in some cases
- Performance is much lower compared to MoNoise, but outperforms LAI
- <https://bitbucket.org/robvanderger/cl-monoise>