

Leveraging Principal Parts for Morphological Inflection

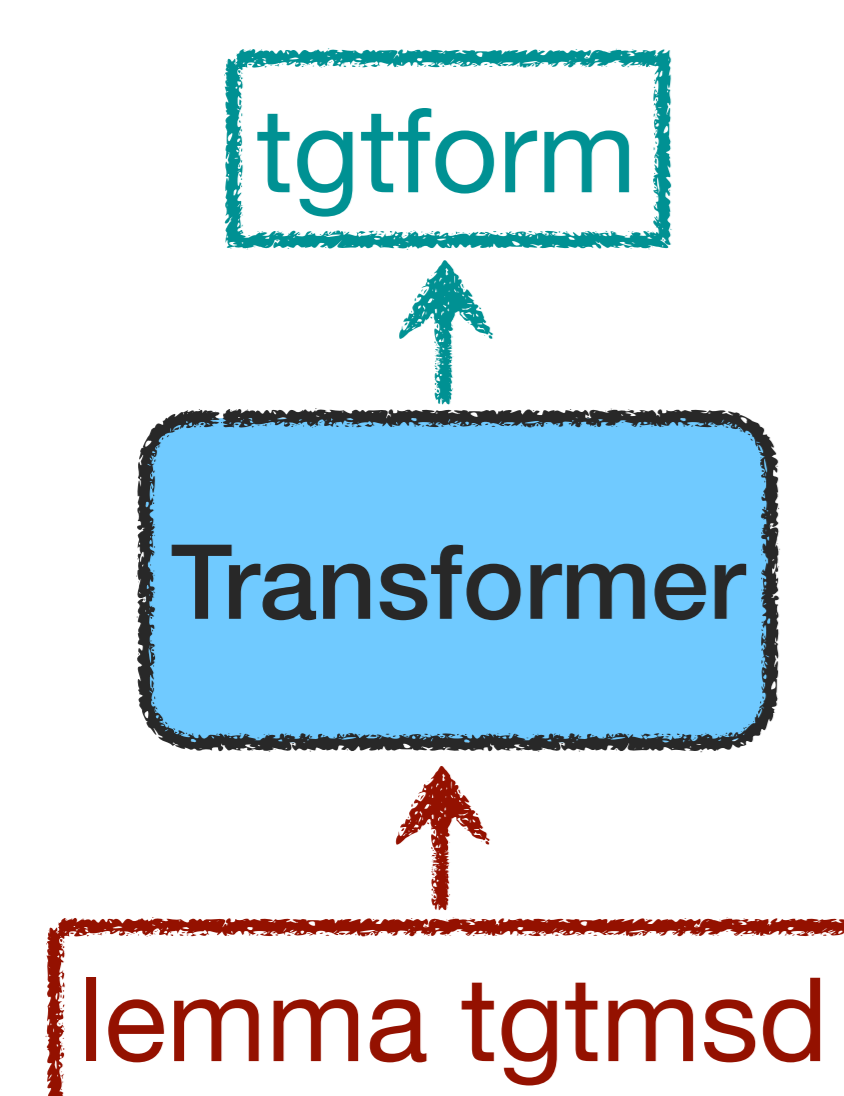
Ling Liu, Mans Hulden

firstname.lastname@colorado.edu

https://github.com/LINGuistLIU/principal_parts_for_inflection 

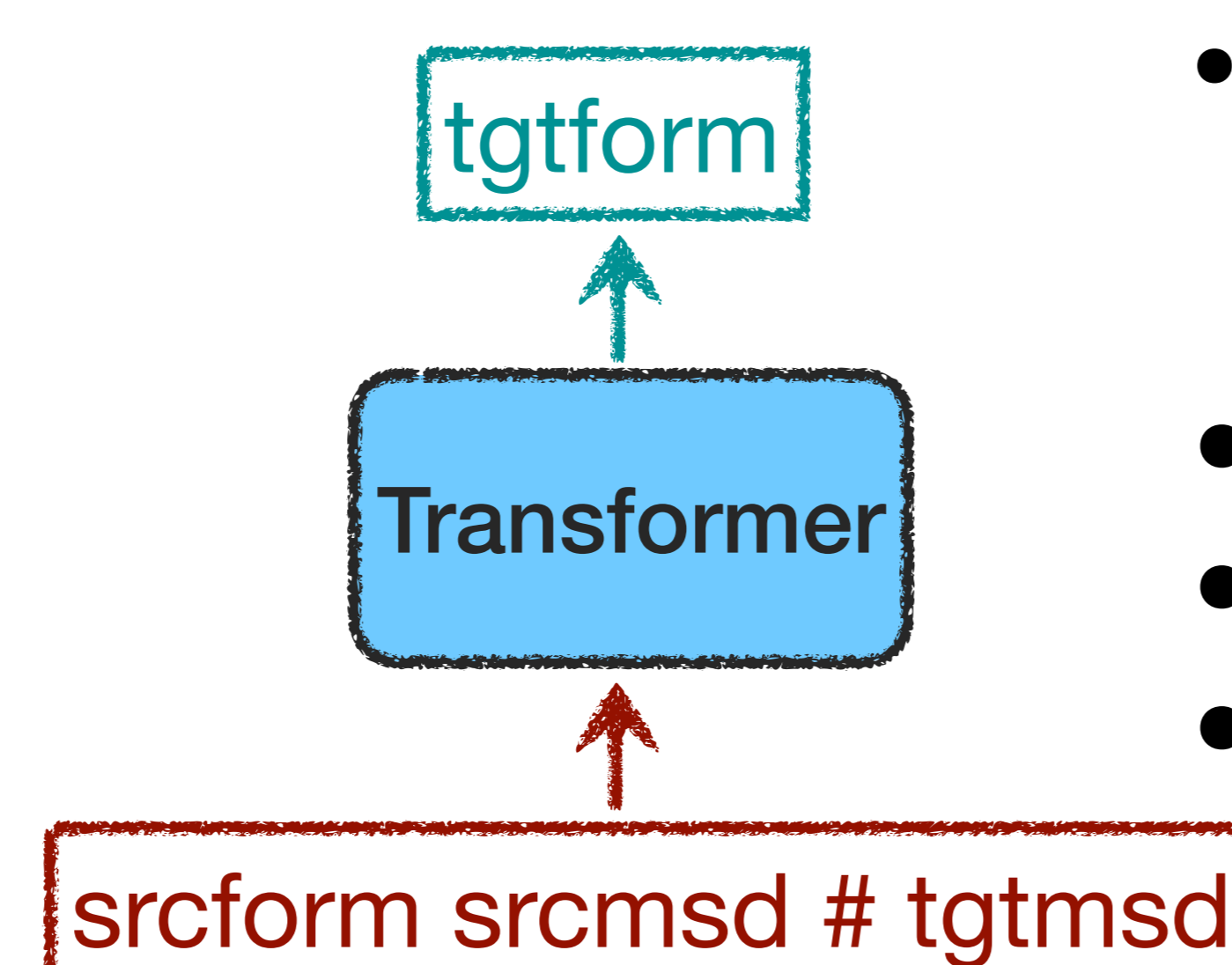
Model architectures

(a) baseline (Fairseq) model



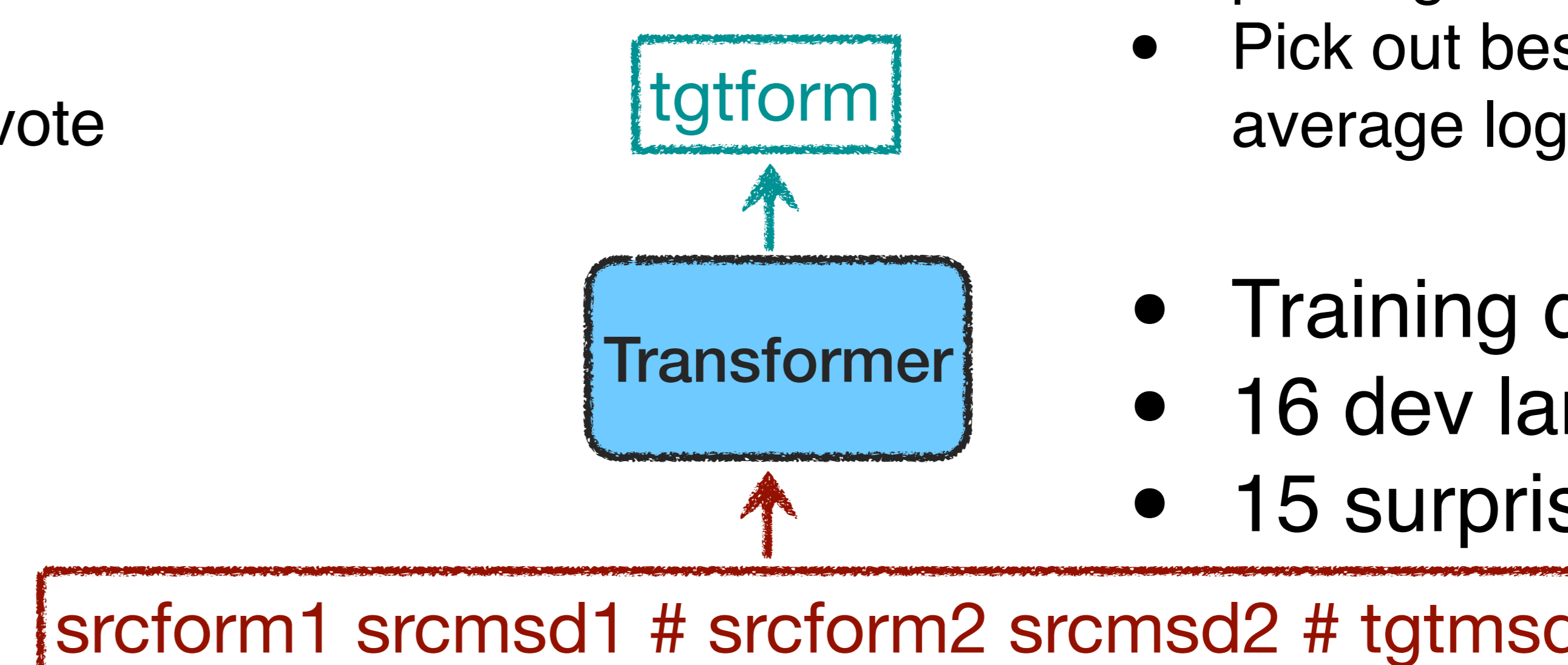
- Use **only the lemma** as the source form to predict the target form
- Training data > 5k
- 21 dev languages
- 23 surprise languages

(b) 1-src model



- Use **each single given slot** in the same paradigm to predict the target form
- Pick out best prediction by average log-likelihood/majority vote
- Training data < 5k
- 24 dev languages
- 23 surprise languages

(c) 2-src model



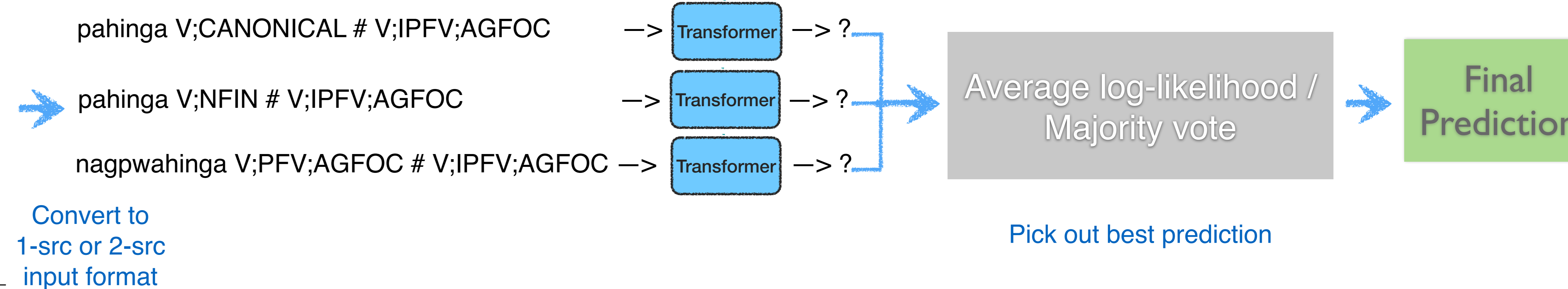
- Use **each pair of given slots** in the same paradigm to predict the target form
- Pick out best prediction by average log-likelihood/majority vote
- Training data < 2k
- 16 dev languages
- 15 surprise languages

Procedures

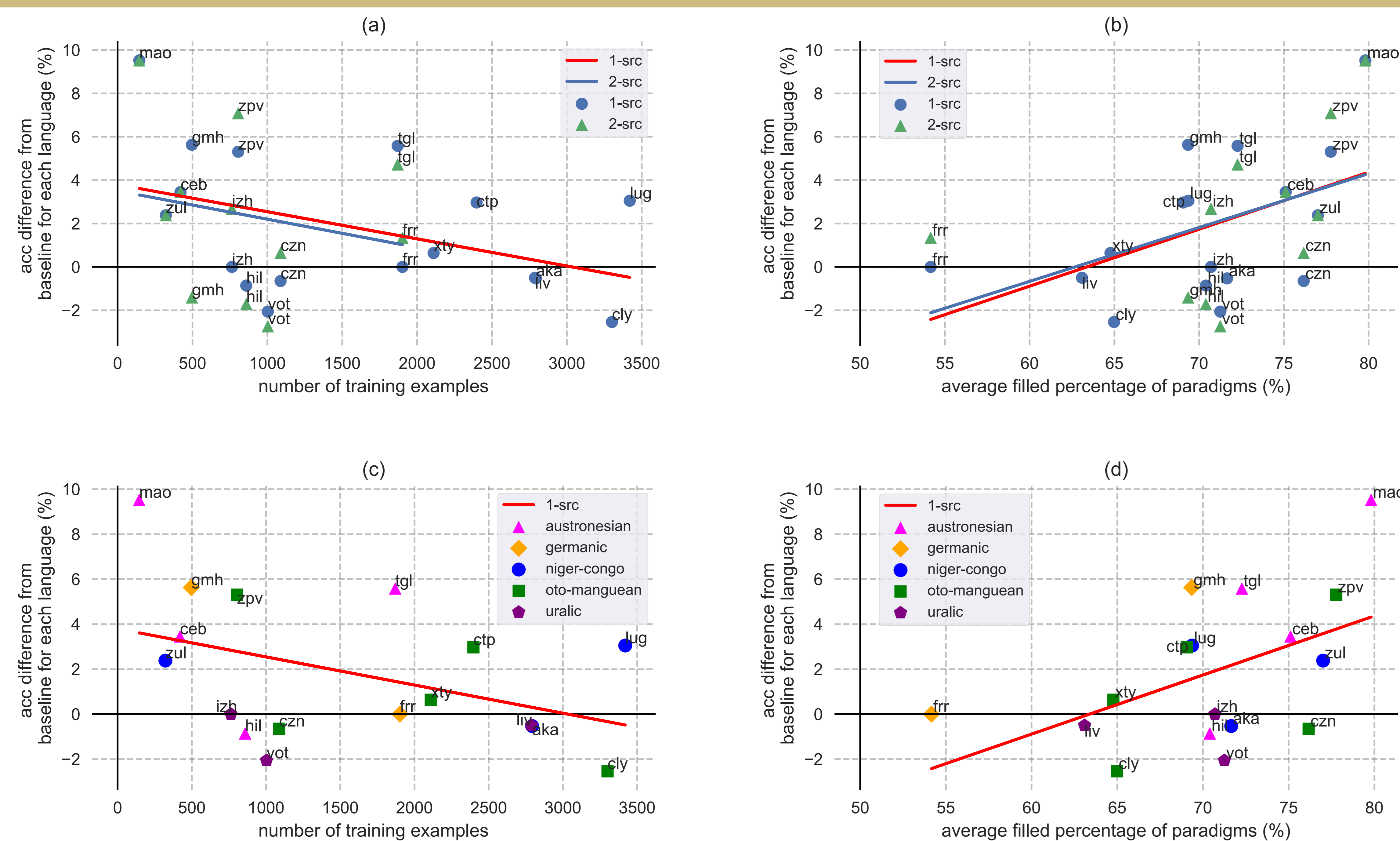
	Train	Dev	Test	?
	libang pahinga tinda	magpapahinga	pahinga	...
	naglibang nagpahinga ititinda			...
	V;IPFV;AGFOC V;PFV;AGFOC V;PFOC;LGSPEC1	V;AGFOC;LGSPEC1	V;IPFV;AGFOC	...

Organize into paradigms

ID	MSD	Lexeme1	Lexeme2	Lexeme3
1	V;CANONICAL	pahinga	bayad	pukpok
2	V;AGFOC;LGSPEC1	-	magbabayad	manumukpok
3	V;IPFV;AGFOC	?	nagbabayad	namumukpok
4	V;IPFV;PFOC	*	binabayaran	pinupukpok
5	V;NFIN	pahinga	bayad	pukpok
6	V;PFOC;LGSPEC1	*	babayaran	pupukpukin
7	V;PFV;AGFOC	nagpahinga	nagbayad	namukpok
8	V;PFV;PFOC	*	binayaran	pinukpok



Results: 1-src/2-src vs trm-single baseline



Results: 1-src/2-src vs best performance of all baselines

