| Language | Train Words (k) | | Dev Types | Dev Tokens |
	Sent	Dict	% OOV	% OOV
Tamil	335	77	44	25
Telugu	414	41	39	21
Bengali	240	7	37	18
Malayalam	263	151	6	3
Hindi	659	n/a	34	11
Urdu	616	116	23	6