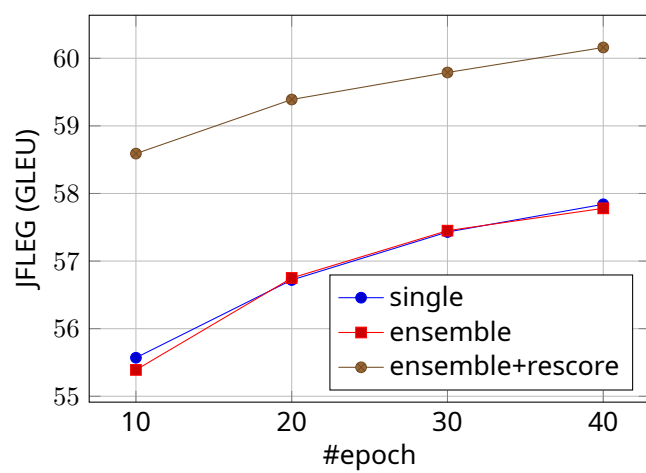
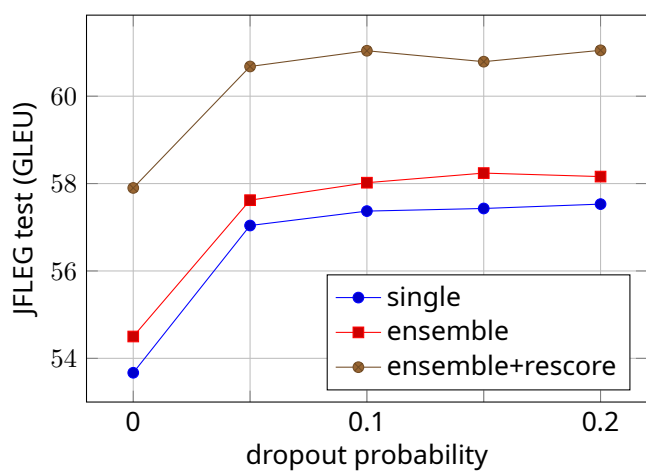
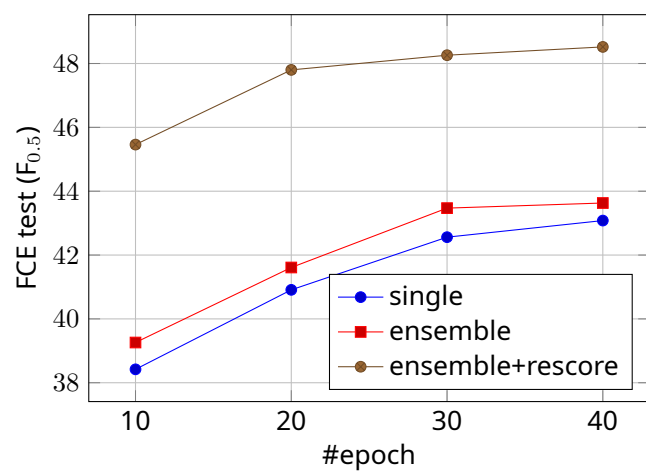
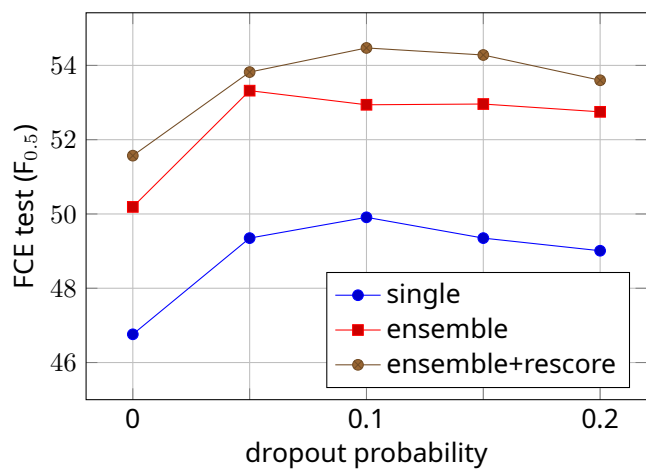
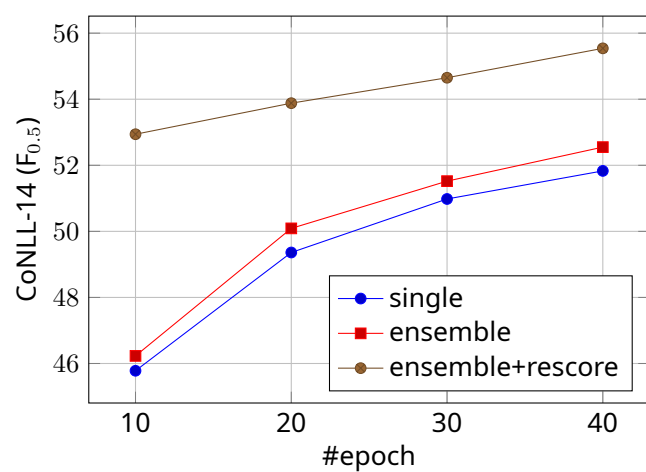
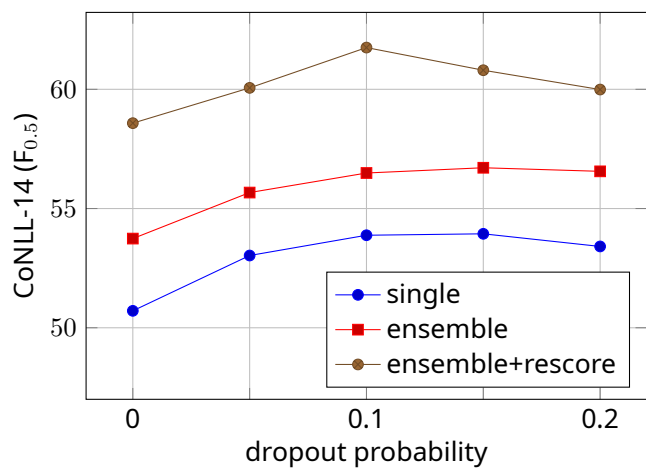
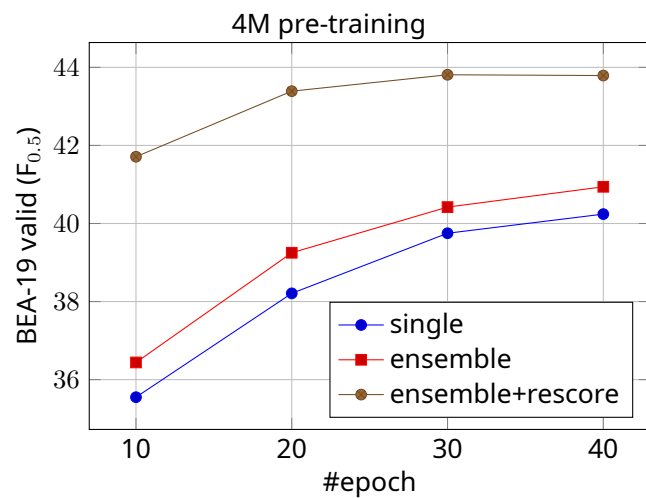
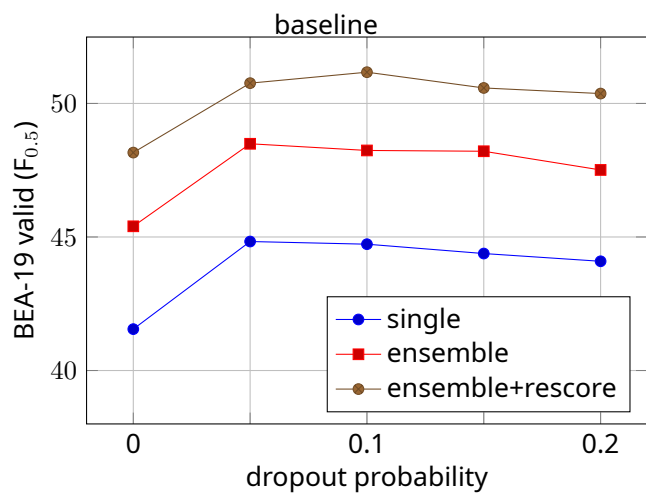


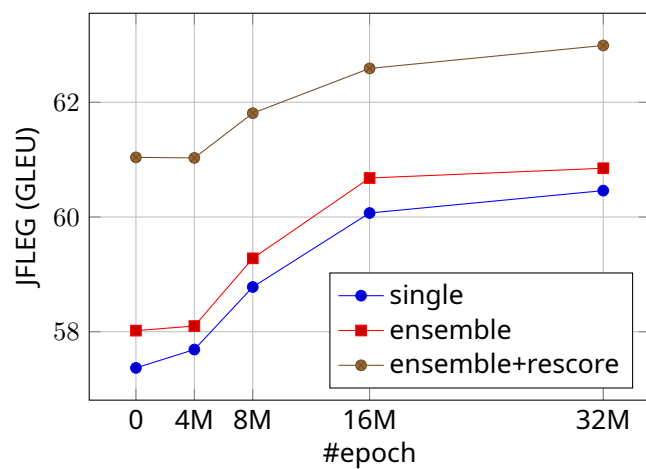
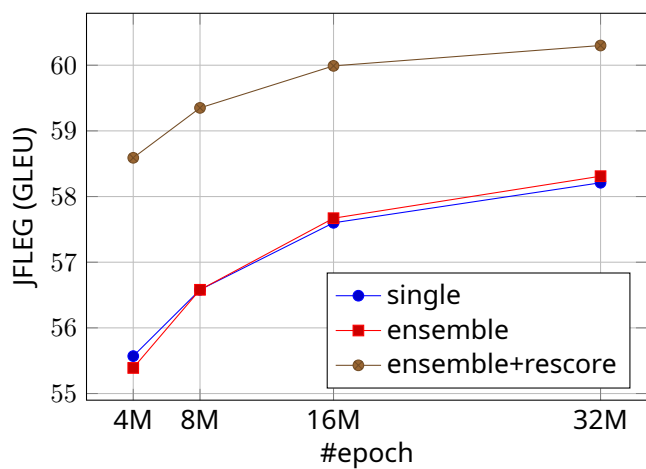
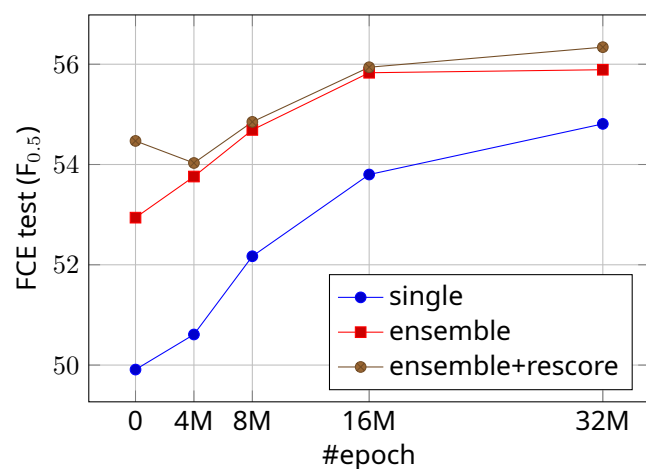
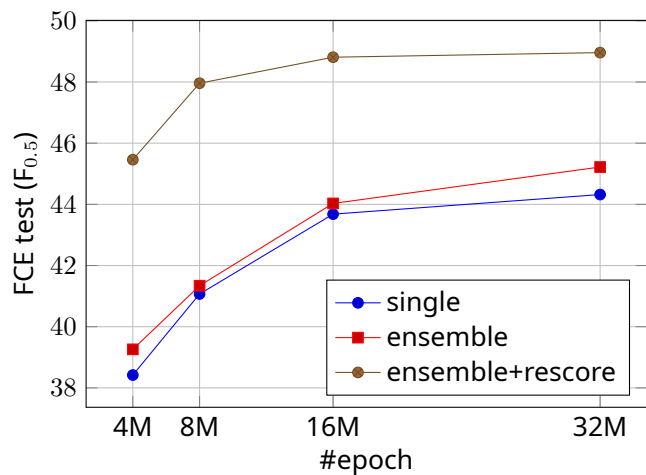
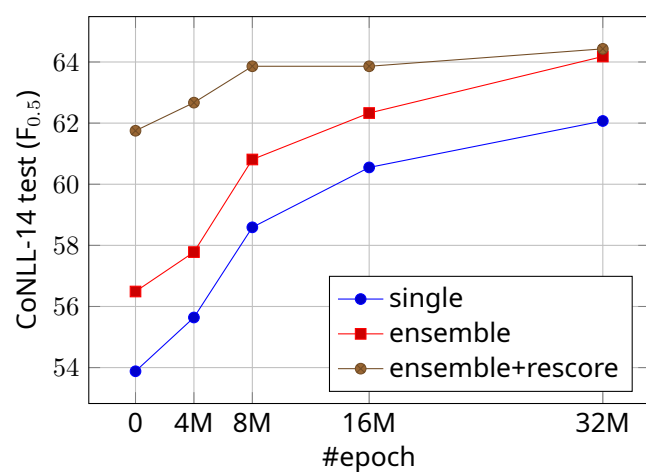
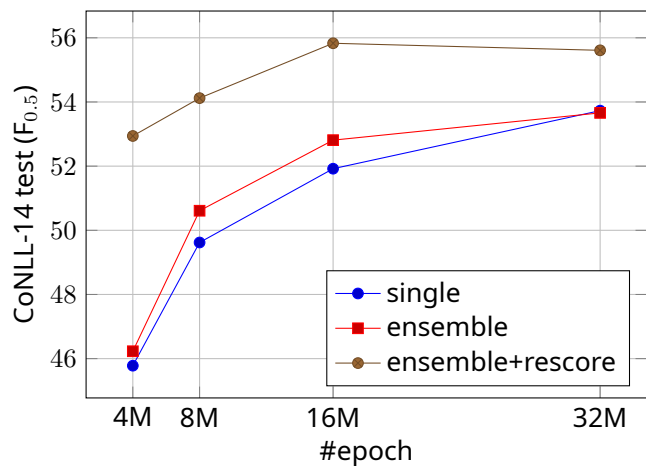
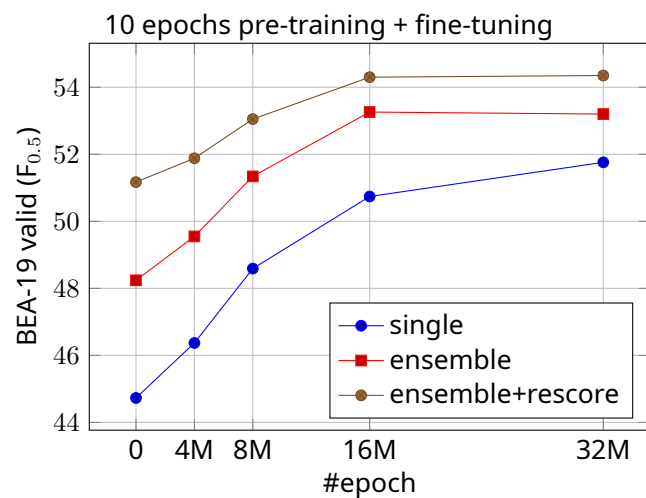
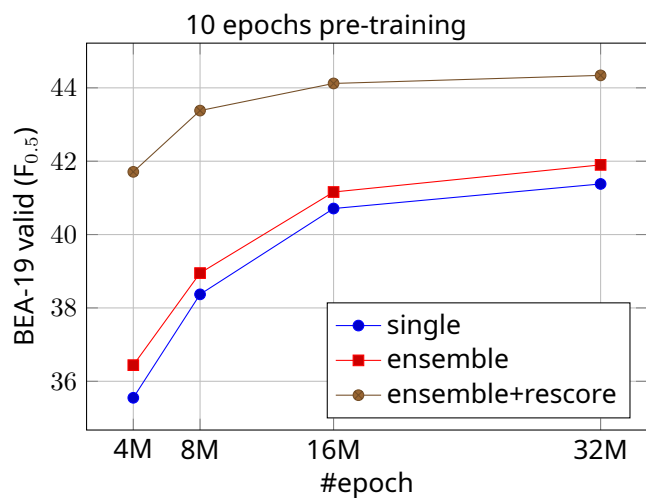
	main result															
	BEA-19						CoNLL-13			CoNLL-14			FCE			
	valid			test			(valid)			(test)			valid		test	
	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	valid	test
	baseline															
single	52.13	28.54	44.73	-	-	-	37.18	66.35	30.82	53.88	51.47	58.56	31.46	49.91	52.57	57.37
ensemble	58.37	28.47	48.24	70.07	46.91	63.98	37.58	72.24	30.17	56.49	55.20	63.68	31.61	52.94	52.77	58.02
+ rescore	59.15	33.23	51.17	69.97	53.45	65.90	43.37	70.27	41.59	61.75	55.64	63.78	34.38	54.47	55.55	61.04
	pre-training															
single	52.58	22.96	41.79	-	-	-	37.48	66.43	32.43	54.92	43.61	56.53	26.77	46.24	52.97	58.35
ensemble	53.51	22.77	42.14	65.93	44.89	60.28	37.28	67.23	32.26	55.25	44.58	57.51	26.78	46.77	53.05	58.55
+ rescore	53.35	27.19	44.74	66.01	51.10	62.37	40.14	63.51	39.15	56.48	46.95	56.77	31.15	48.75	54.79	60.10
	pre-training + fine-tuning															
single	60.61	35.03	52.83	-	-	-	46.74	70.82	44.44	63.25	56.12	64.25	38.02	56.40	56.45	62.23
ensemble	62.38	34.94	53.91	76.61	58.67	72.19	48.18	73.57	44.64	65.13	57.33	66.32	37.96	57.70	56.79	62.72
ensemble + rescore	61.75	37.50	54.68	75.79	61.81	72.51	49.74	70.82	50.16	65.43	57.48	65.50	39.48	57.87	58.17	63.69
	pre-training + fine-tuning + domain adaptation															
single	58.33	42.36	54.24	-	-	-	-	-	-	-	56.69	64.21	40.83	57.59	-	-
ensemble	61.11	43.09	56.39	74.89	64.54	72.57	-	-	-	-	58.38	66.63	40.95	59.21	-	-
ensemble + rescore	60.93	43.75	56.49	74.84	65.51	72.76	-	-	-	-	58.51	65.39	42.54	59.05	-	-

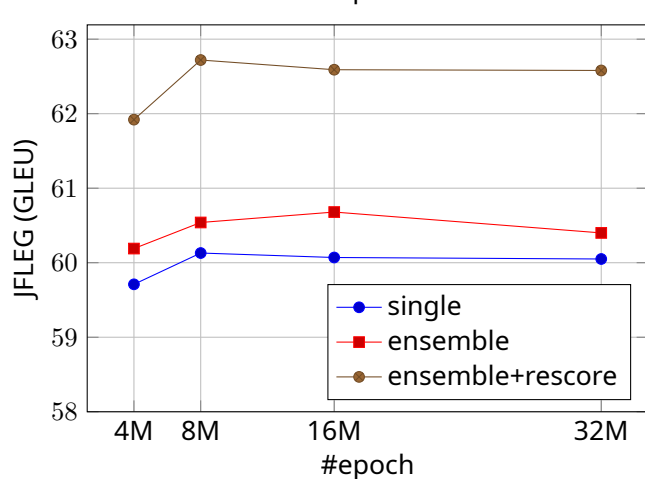
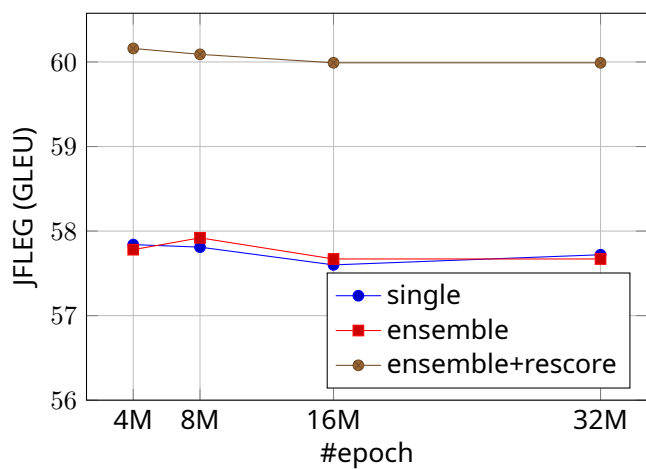
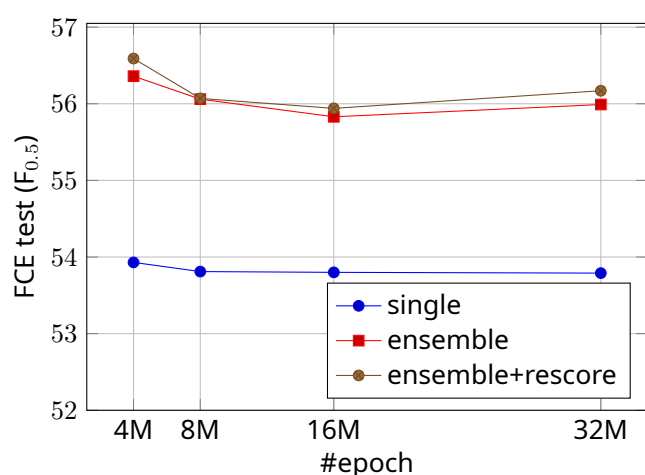
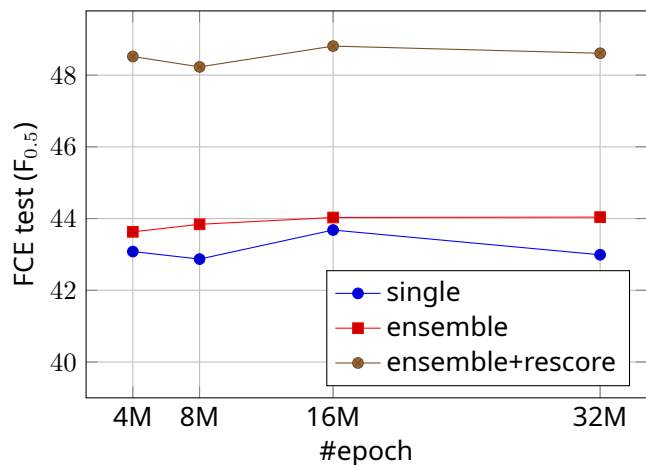
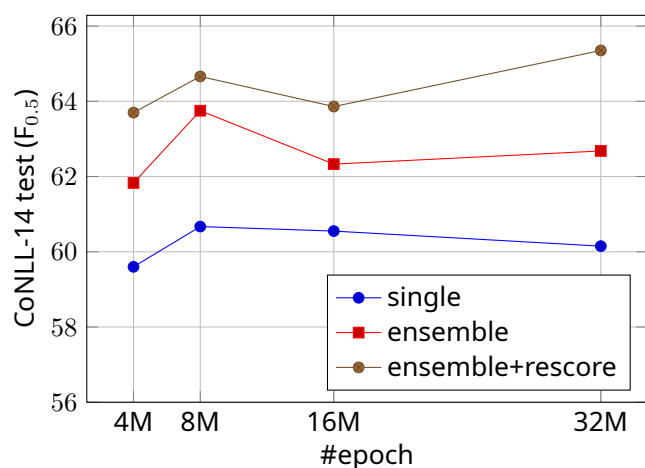
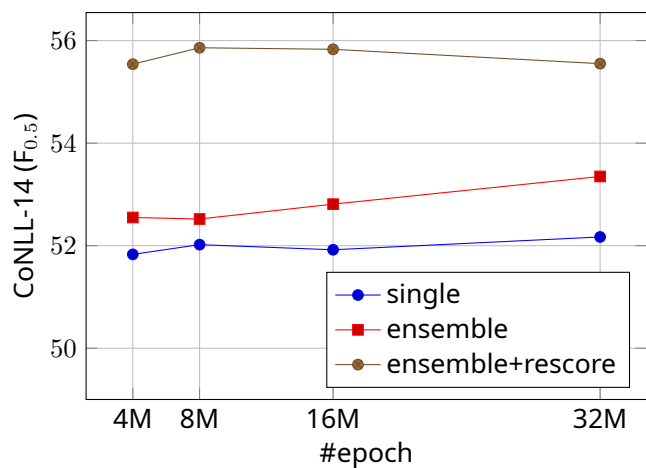
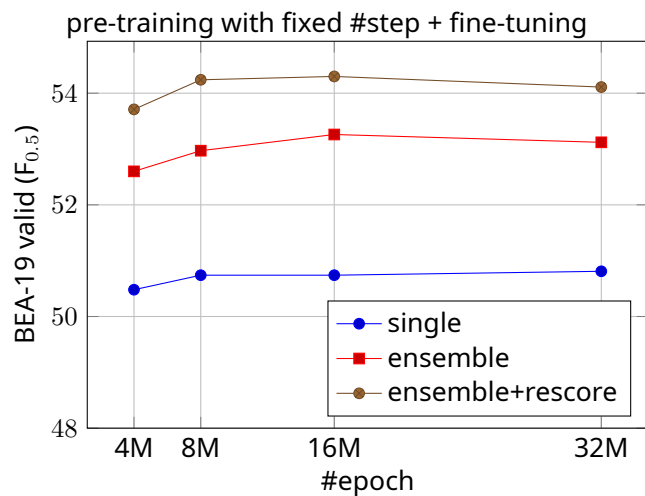
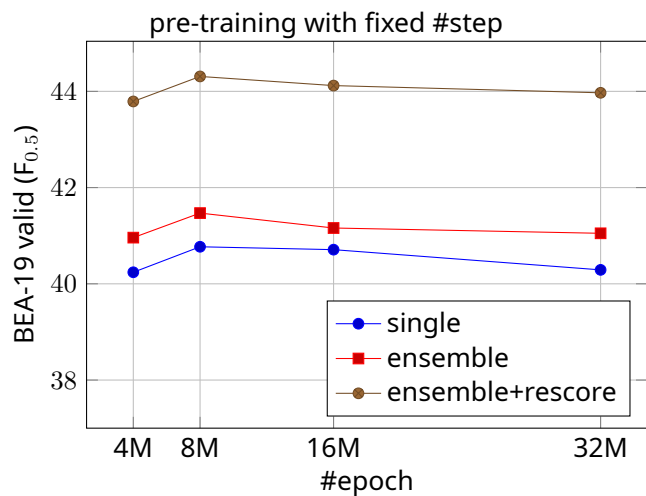
	pre-trained model											
	BEA-19 valid			CoNLL-14			FCE test			JFLEG test		
	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>
	4M x 10 epochs											
single	47.11	17.96	35.55	59.77	23.66	45.78	49.11	20.55	38.42	55.57		
ensemble	49.21	17.88	36.44	61.42	23.24	46.23	50.82	20.55	39.26	55.39		
ensemble + rescore	50.18	24.90	41.71	62.18	33.20	52.94	55.50	26.38	45.46	58.59		
	4M x 20 epochs											
single	48.92	20.38	38.21	62.44	26.87	49.36	50.72	23.06	40.91	56.72		
ensemble	50.97	20.44	39.25	64.25	26.62	50.09	52.11	23.04	41.61	56.75		
ensemble + rescore	51.94	26.16	43.39	61.17	36.49	53.88	56.10	30.03	47.80	59.39		
	4M x 30 epochs											
single	50.06	21.79	39.75	63.35	28.63	50.98	52.09	24.59	42.56	57.43		
ensemble	51.56	21.69	40.42	64.60	28.48	51.52	53.77	24.62	43.47	57.45		
ensemble + rescore	52.45	26.42	43.81	61.85	37.28	54.65	56.89	30.03	48.26	59.79		
	4M x 40 epochs											
single	50.38	22.31	40.24	63.76	29.66	51.83	52.62	24.98	43.08	57.84		
ensemble	51.72	22.32	40.94	65.13	29.65	52.55	53.63	24.99	43.63	57.78		
ensemble + rescore	51.36	27.54	43.79	63.45	37.07	55.54	56.52	30.97	48.52	60.16		
	8M x 10 epochs											
single	49.54	20.18	38.37	62.76	27.01	49.62	51.41	22.78	41.07	56.58		
ensemble	51.10	19.97	38.95	65.16	26.73	50.61	52.16	22.60	41.34	56.58		
ensemble + rescore	51.80	26.30	43.38	61.20	36.99	54.12	58.48	27.90	47.96	59.35		
	8M x 20 epochs											
single	51.07	22.58	40.77	63.95	29.79	52.02	52.32	24.89	42.87	57.81		
ensemble	52.66	22.42	41.47	65.11	29.61	52.52	53.92	25.08	43.84	57.92		
ensemble + rescore	52.15	27.66	44.31	64.32	36.61	55.86	56.58	30.34	48.23	60.09		
	16M x 10 epochs											
single	51.26	22.33	40.71	64.04	29.56	51.92	53.68	25.01	43.68	57.60		
ensemble	52.41	22.14	41.16	65.78	29.52	52.81	54.41	24.97	44.03	57.67		
ensemble + rescore	51.83	27.66	44.12	63.17	38.12	55.83	56.89	31.13	48.81	59.99		
	32M x 5 epochs											
single	50.84	22.02	40.29	64.33	29.71	52.17	52.79	24.67	42.99	57.72		
ensemble	52.43	21.97	41.05	66.85	29.51	53.35	54.53	24.88	44.04	57.67		
ensemble + rescore	51.88	27.32	43.97	62.52	38.42	55.55	56.88	30.73	48.61	59.99		
	32M x 10 epochs											
single	51.73	22.99	41.38	65.41	31.35	53.73	54.04	25.78	44.32	58.21		
ensemble	52.68	23.04	41.90	65.39	31.23	53.66	55.46	26.01	45.22	58.31		
ensemble + rescore	52.74	27.07	44.34	63.11	37.69	55.61	56.85	31.48	48.96	60.30		

	pre-train x 10 epochs + fine-tuning											
	BEA-19 valid			CoNLL-14			FCE test			JFLEG test		
	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>
	4M x 10 epochs pre-training + fine-tuning											
single	55.50	28.11	46.37	67.98	32.29	55.64	59.17	32.21	50.61	57.69		
ensemble	61.53	27.86	49.55	73.36	31.25	57.78	64.30	32.47	53.76	58.10		
ensemble + rescore	60.45	33.11	51.88	71.49	41.95	62.67	63.72	33.59	54.03	61.03		
	8M x 10 epochs pre-training + fine-tuning											
single	57.85	29.76	48.59	69.89	35.59	58.59	60.33	33.94	52.17	58.78		
ensemble	62.78	29.70	51.34	74.31	35.23	60.81	64.34	34.18	54.69	59.28		
ensemble + rescore	60.03	36.21	53.05	70.92	45.66	63.86	62.80	36.40	54.85	61.81		
	16M x 10 epochs pre-training + fine-tuning											
single	58.25	33.53	50.74	70.42	38.85	60.55	60.30	37.61	53.80	60.07		
ensemble	62.02	34.03	53.26	73.44	38.84	62.33	63.18	38.10	55.83	60.68		
ensemble + rescore	60.82	38.00	54.30	70.11	47.07	63.86	62.64	39.17	55.94	62.59		
	32M x 10 epochs pre-training + fine-tuning											
single	59.58	34.05	51.76	72.66	39.28	62.07	62.47	36.82	54.81	60.46		
ensemble	62.06	33.87	53.20	76.14	39.41	64.18	64.13	36.91	55.89	60.85		
ensemble + rescore	61.32	37.37	54.35	70.18	48.55	64.43	62.74	40.01	56.34	62.99		

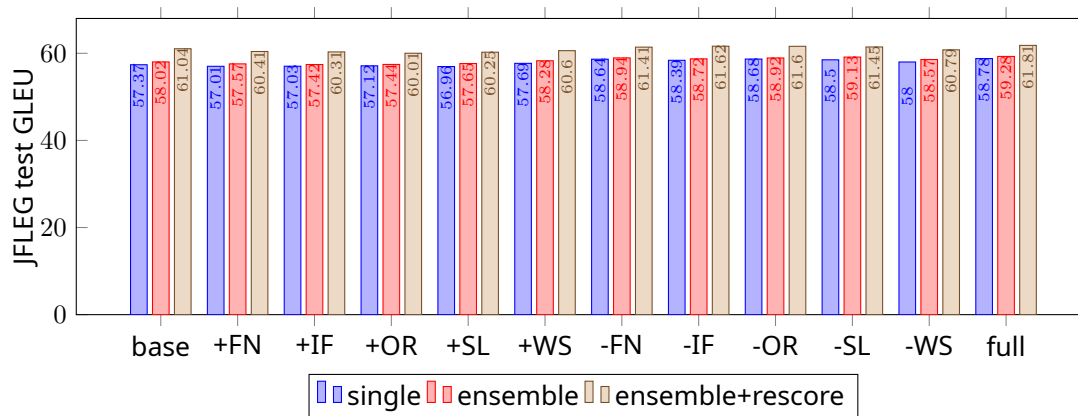
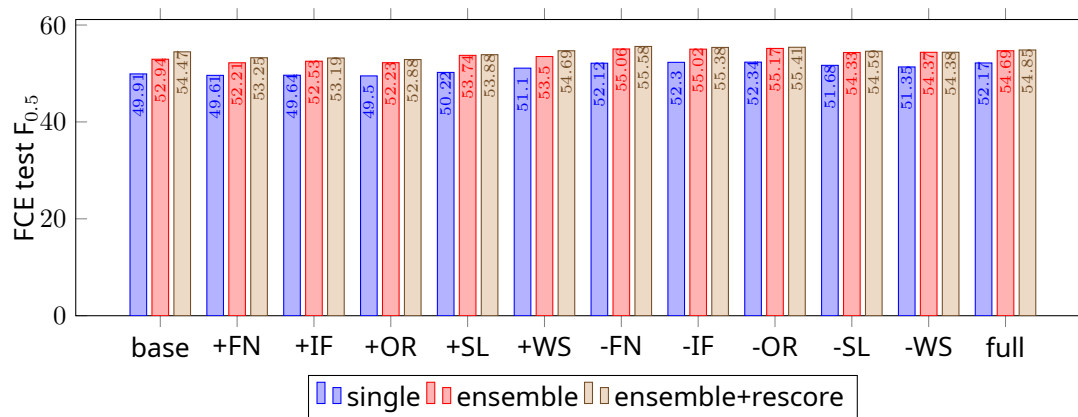
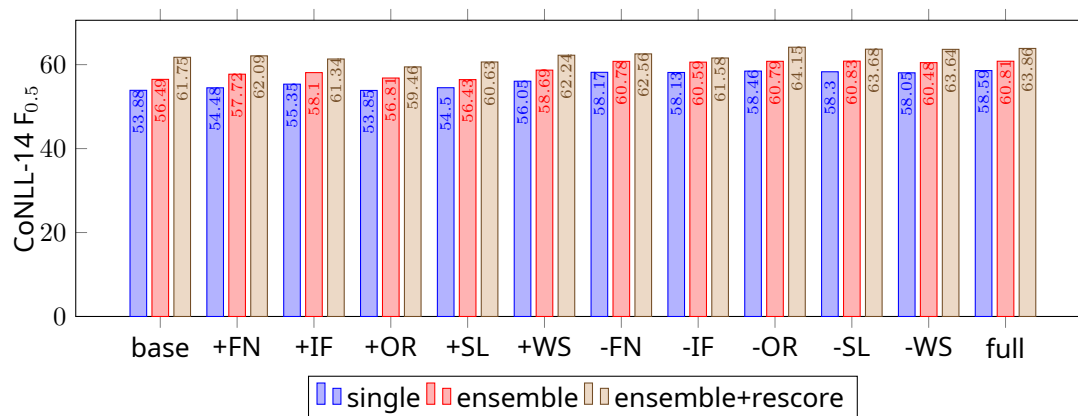
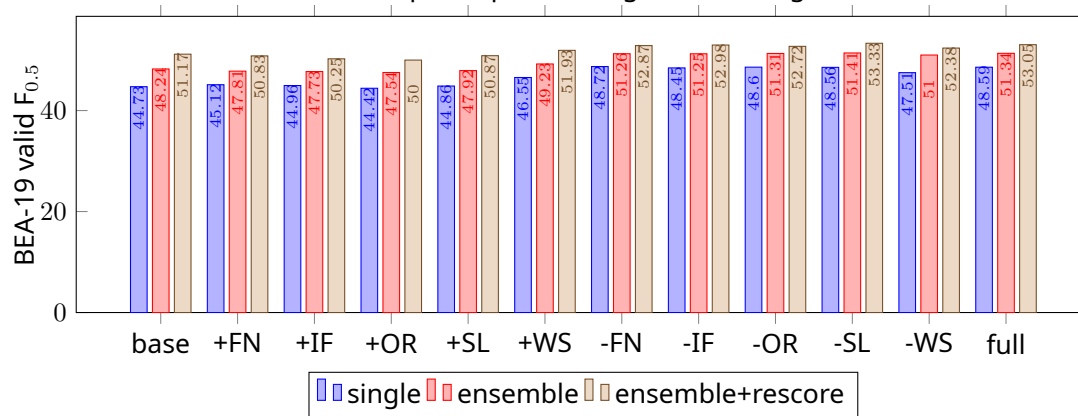
pre-training with fixed #step + fine-tuning												
BEA-19 valid				CoNLL-14			FCE test			JFLEG test		
	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>	P	R	F <sub>0.5</sub>
4M x 40 epochs pre-training + fine-tuning												
single	58.03	33.20	50.48	69.89	37.67	59.60	60.82	37.13	53.93	59.71		
ensemble	61.46	33.36	52.60	73.39	37.93	61.83	64.43	37.55	56.36	60.19		
ensemble + rescore	60.81	36.62	53.71	70.26	46.39	63.70	63.82	38.93	56.59	61.92		
8M x 20 epochs pre-training + fine-tuning												
single	58.17	33.64	50.74	71.35	38.02	60.67	61.31	36.16	53.81	60.13		
ensemble	61.73	33.79	52.97	76.15	38.60	63.75	64.59	36.69	56.06	60.54		
ensemble + rescore	60.69	38.05	54.24	71.22	47.25	64.66	63.12	38.76	56.07	62.72		
16M x 10 epochs pre-training + fine-tuning												
single	58.25	33.53	50.74	70.42	38.85	60.55	60.30	37.61	53.80	60.07		
ensemble	62.02	34.03	53.26	73.44	38.84	62.33	63.18	38.10	55.83	60.68		
ensemble + rescore	60.82	38.00	54.30	70.11	47.07	63.86	62.64	39.17	55.94	62.59		
32M x 5 epochs pre-training + fine-tuning												
single	58.90	32.80	50.81	70.02	38.50	60.15	60.81	36.81	53.79	60.05		
ensemble	62.66	33.01	53.12	73.86	39.04	62.68	63.95	37.39	55.99	60.40		
ensemble + rescore	61.41	36.68	54.11	73.48	45.30	65.35	63.56	38.34	56.17	62.58		



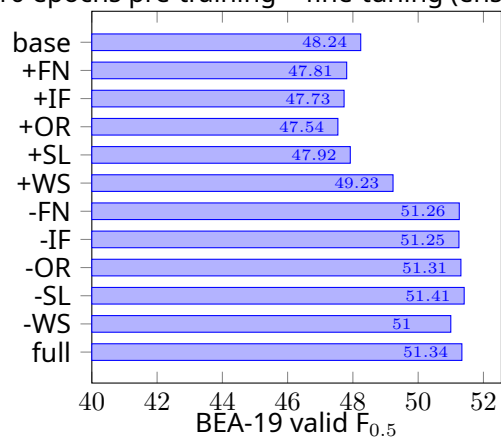




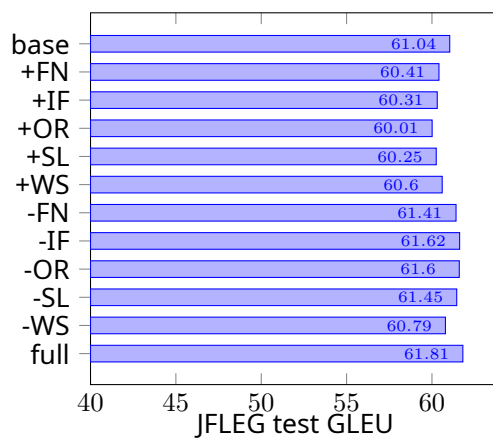
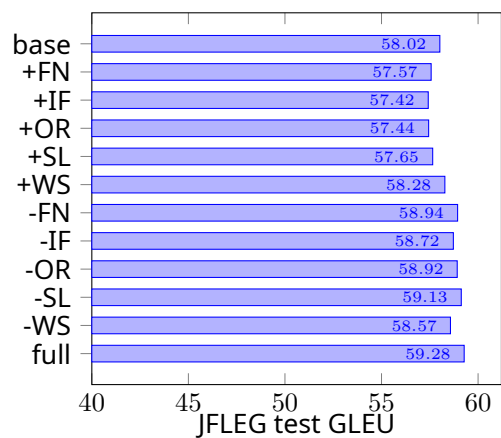
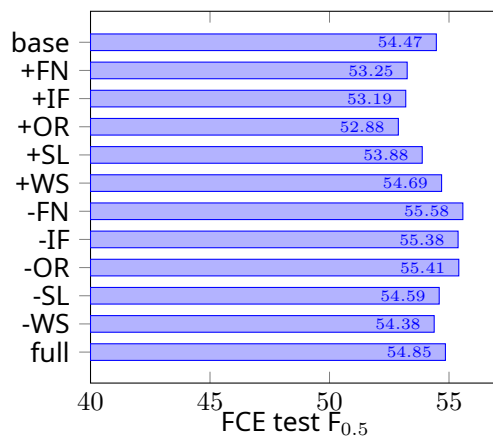
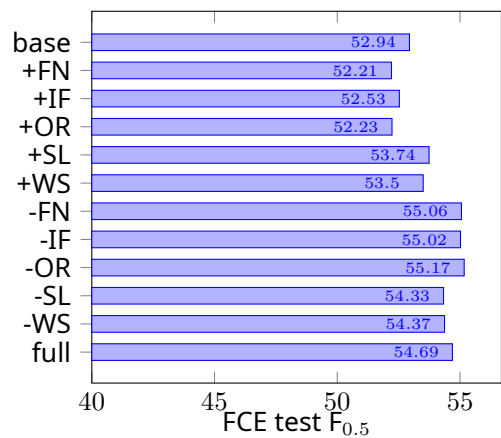
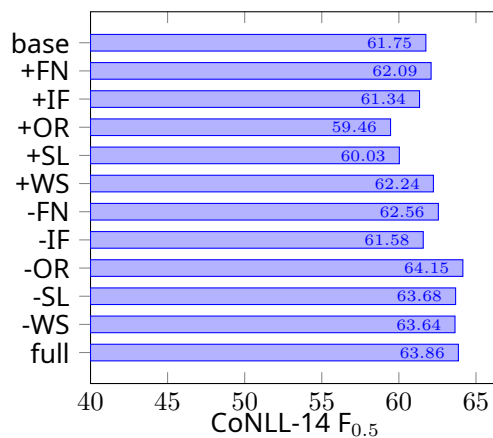
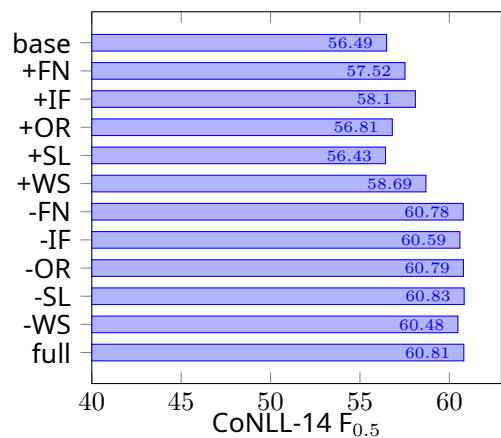
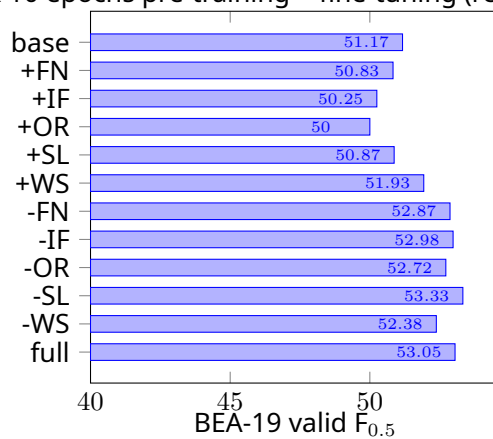
8M x 10 epochs pre-training + fine-tuning

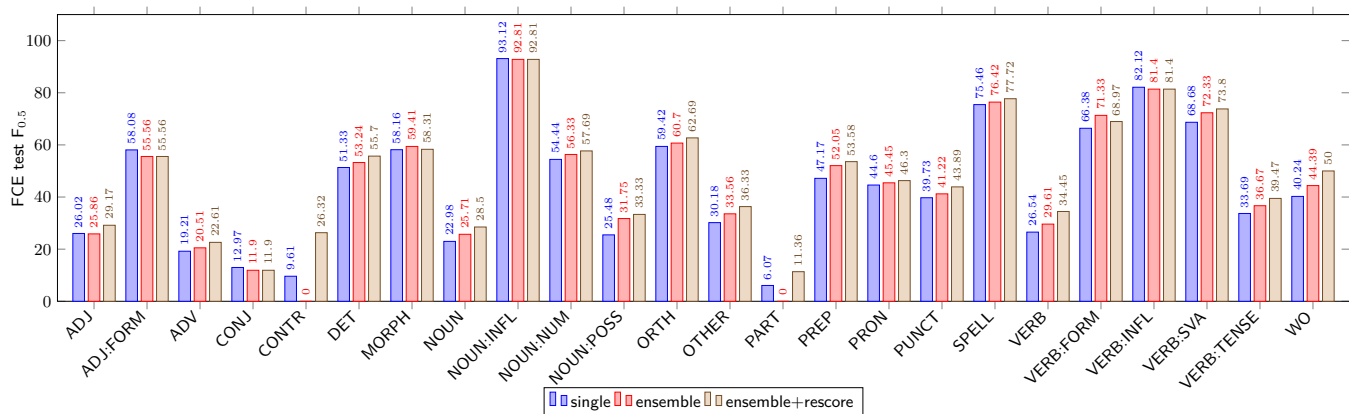
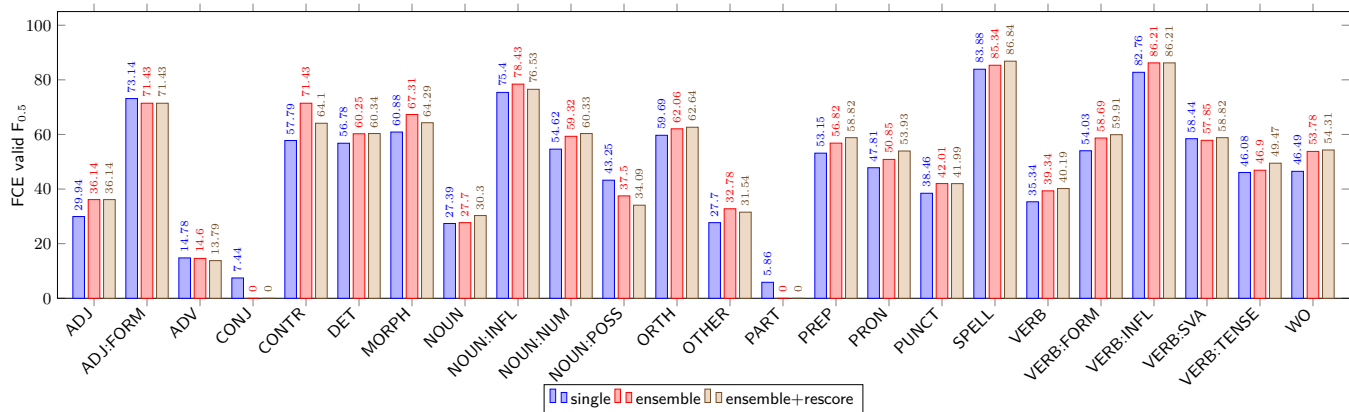
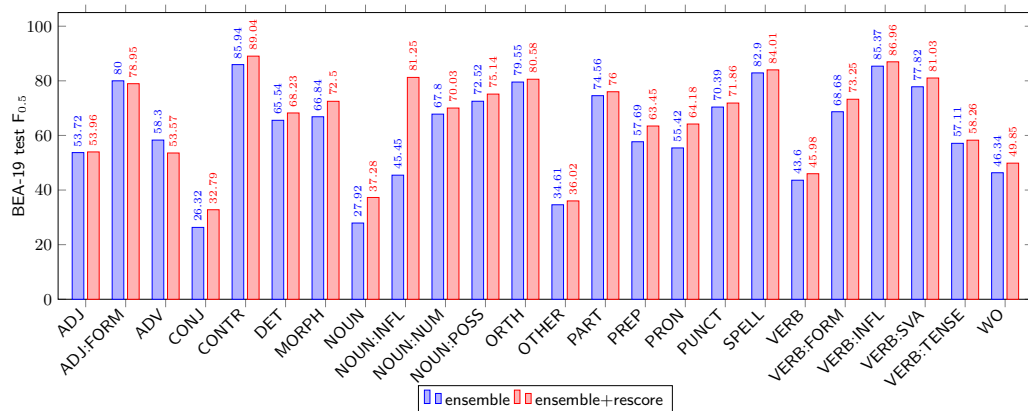
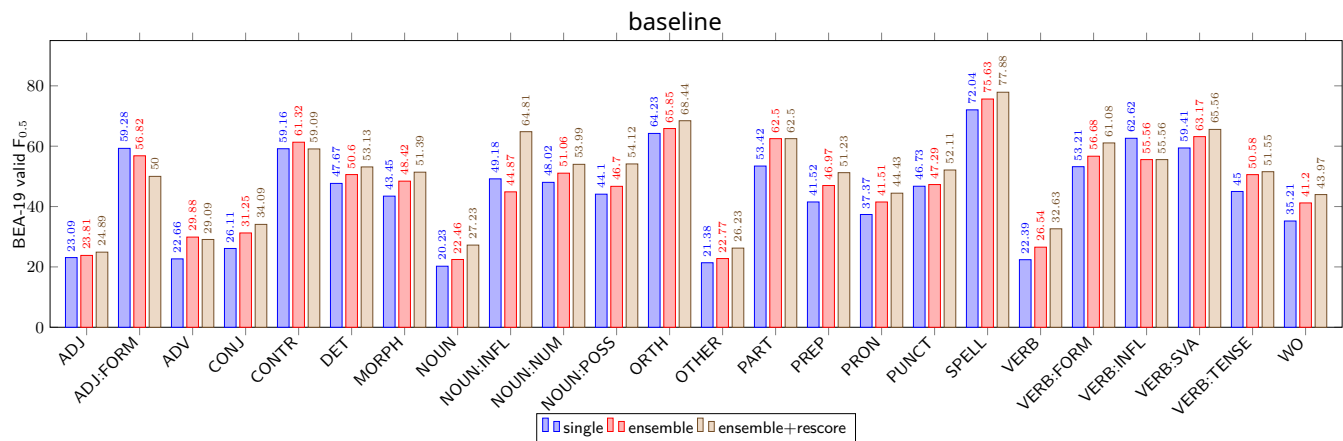


8M x 10 epochs pre-training + fine-tuning (ensemble)

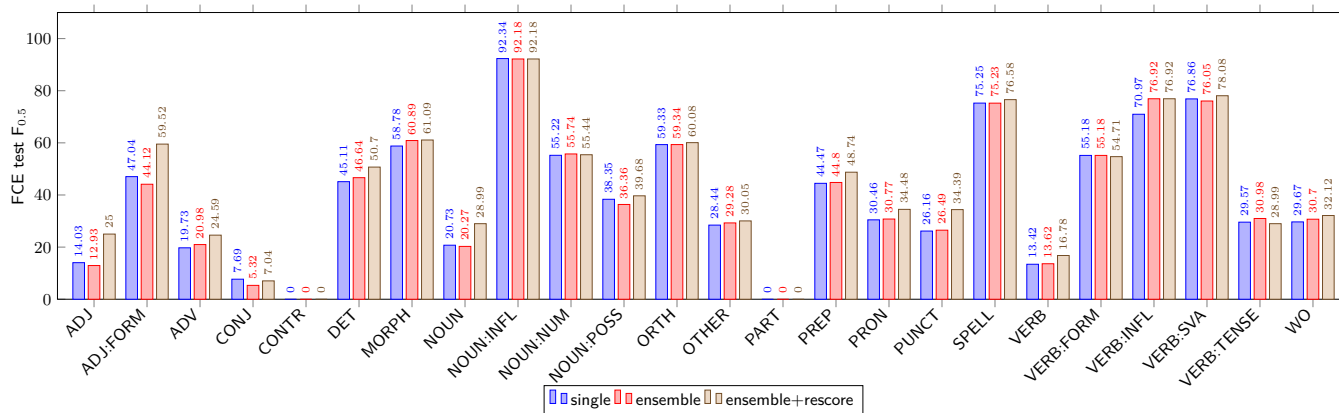
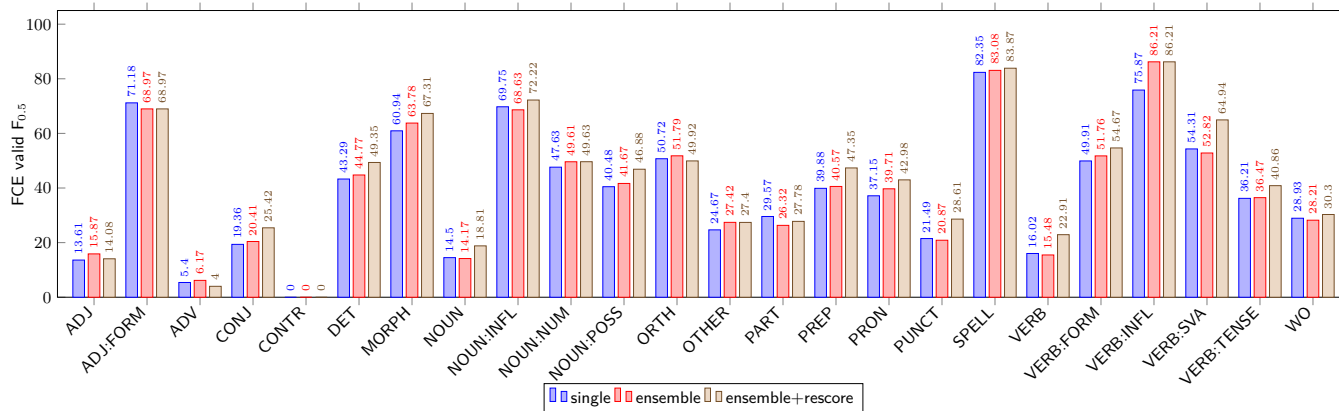
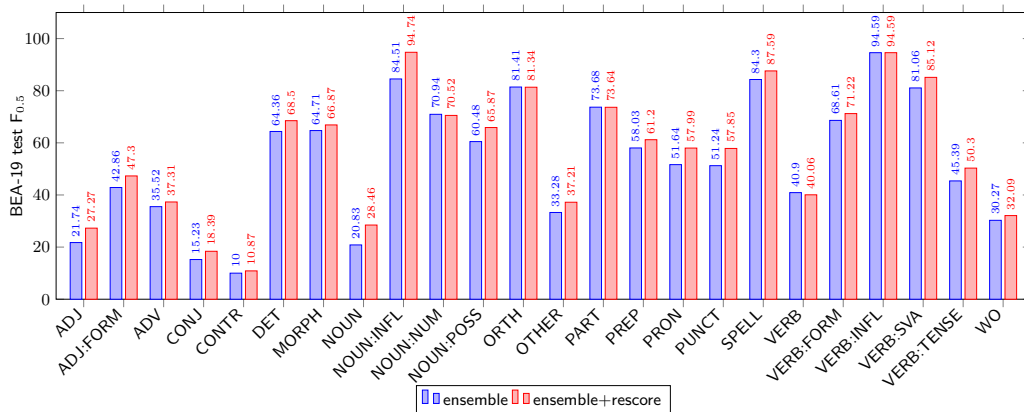
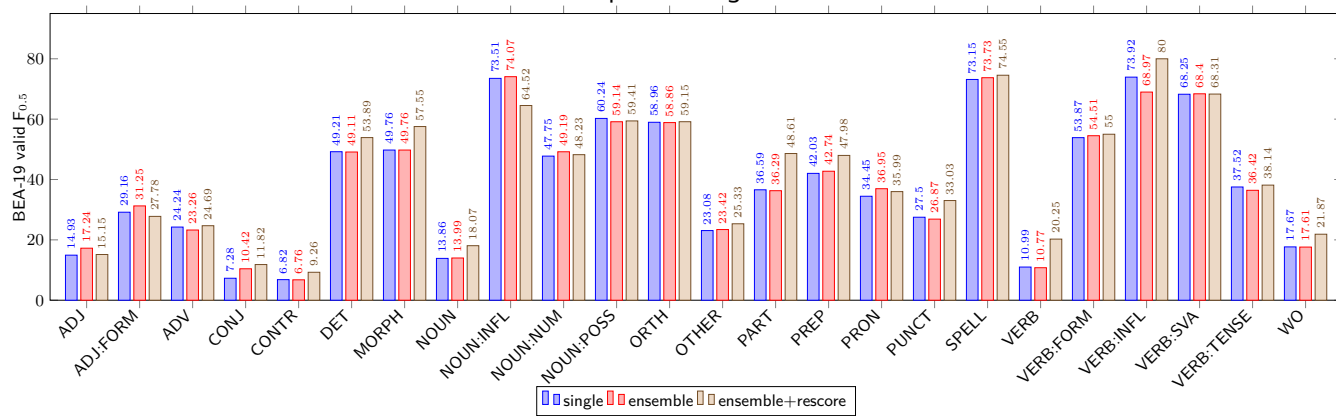


8M x 10 epochs pre-training + fine-tuning (rescore)



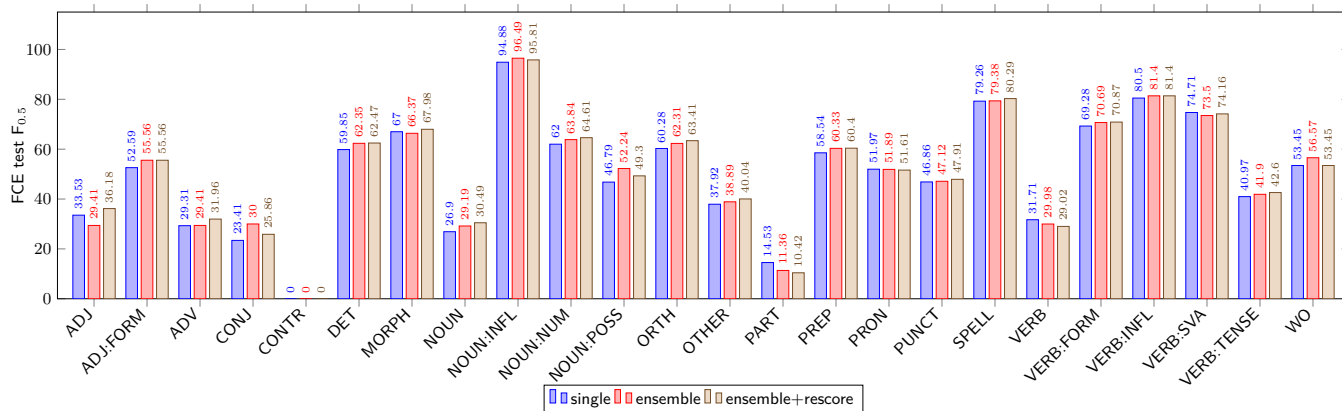
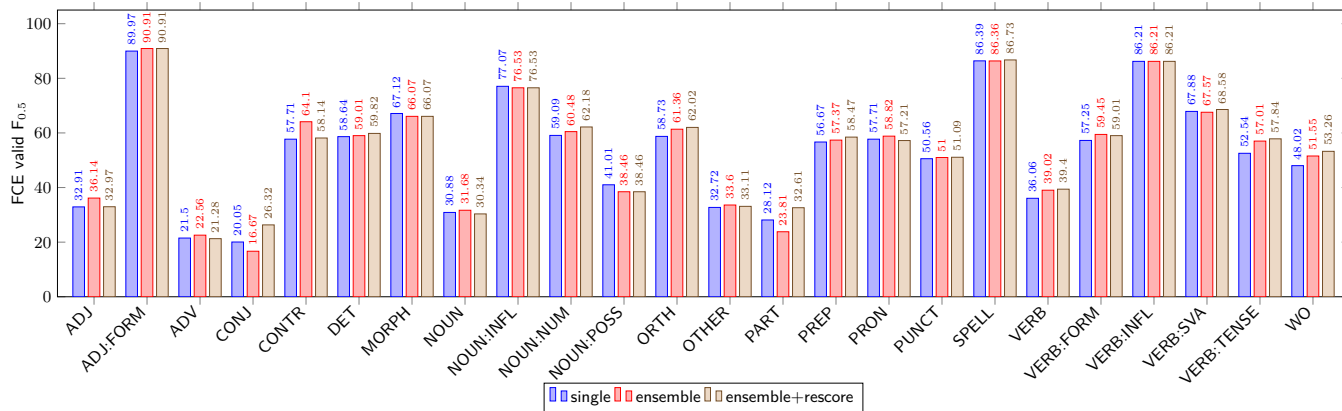
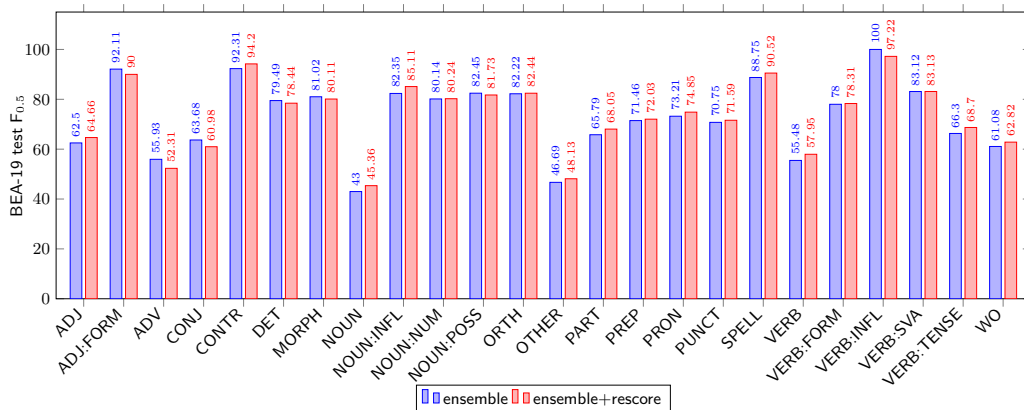
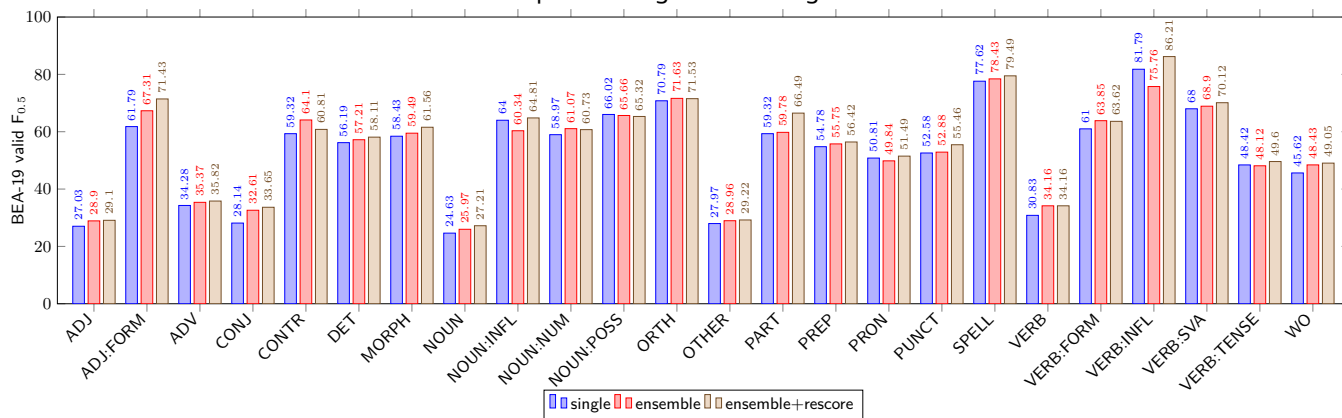


pre-training

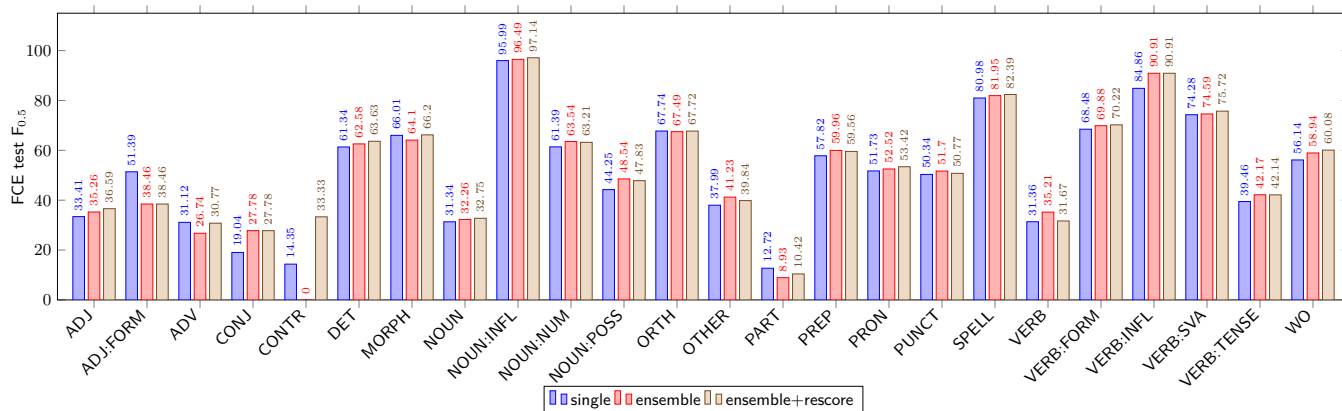
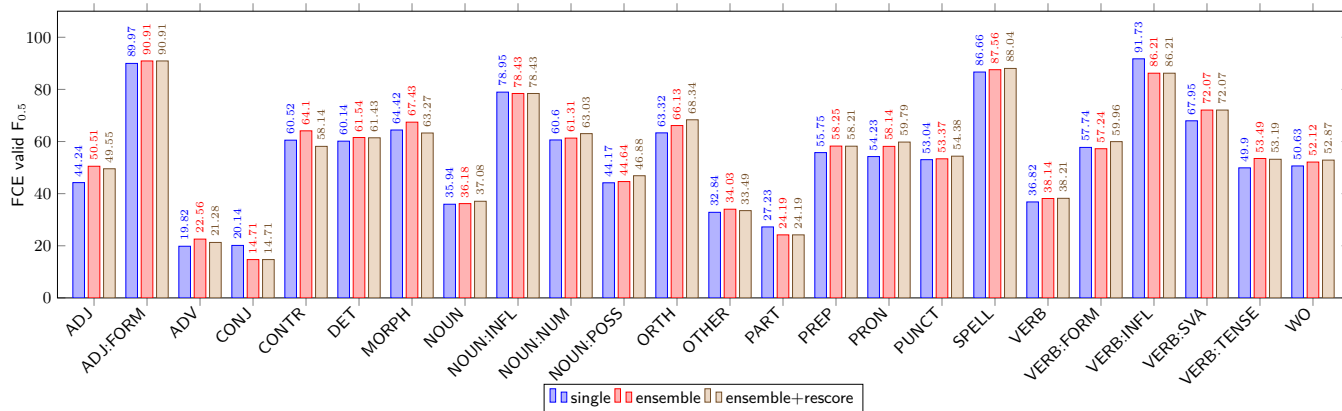
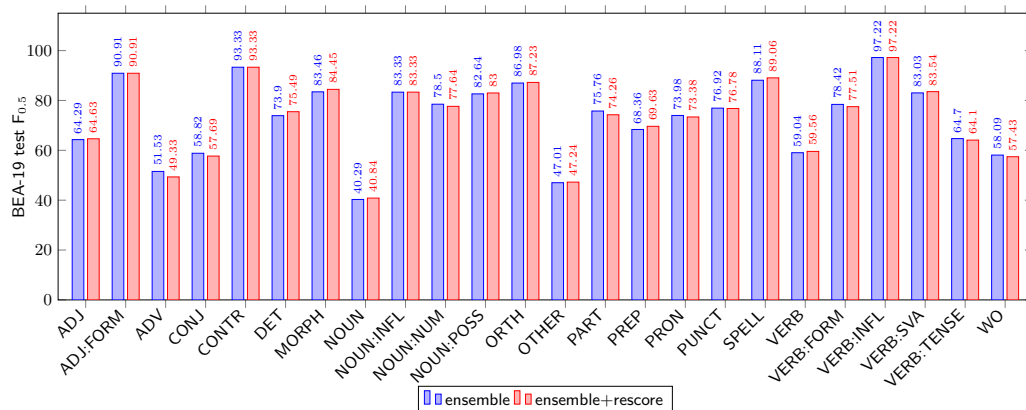
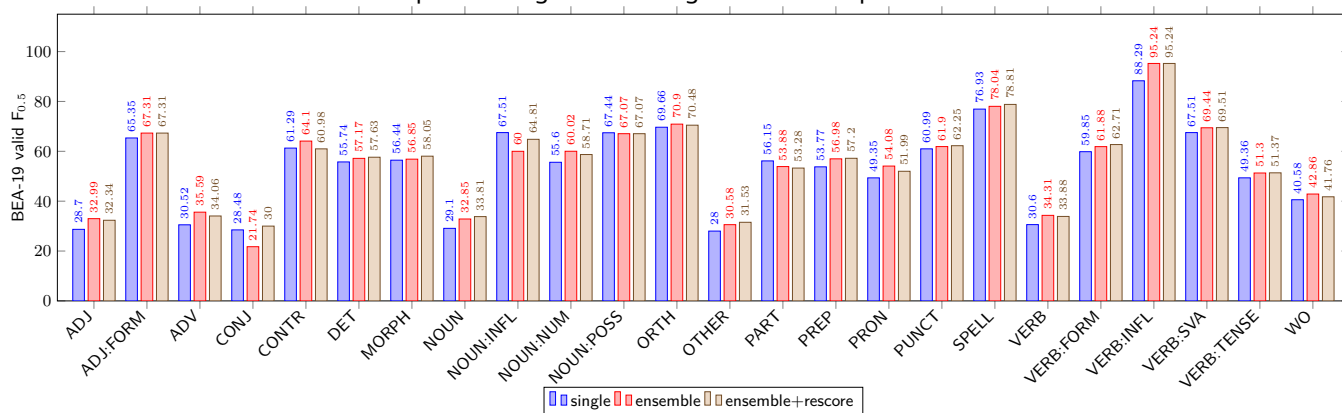




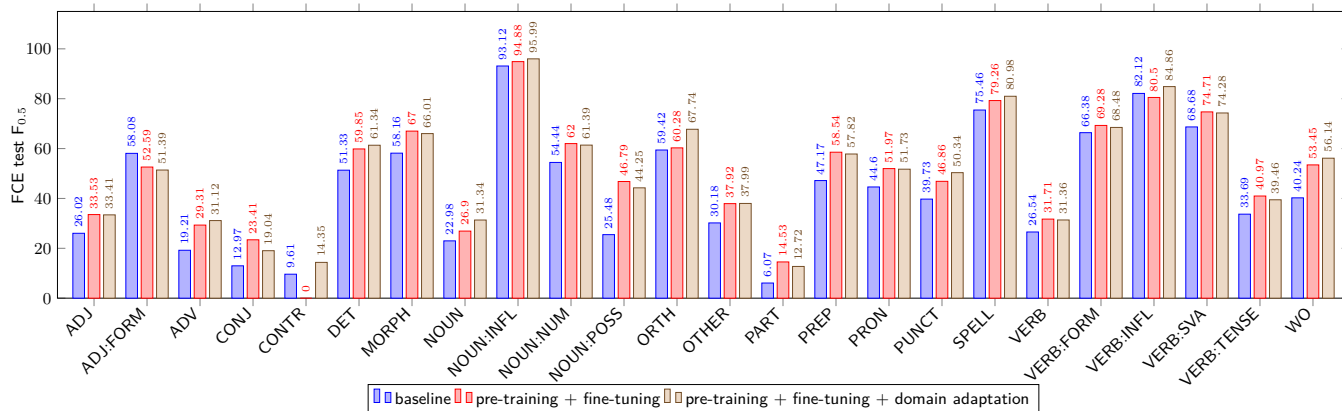
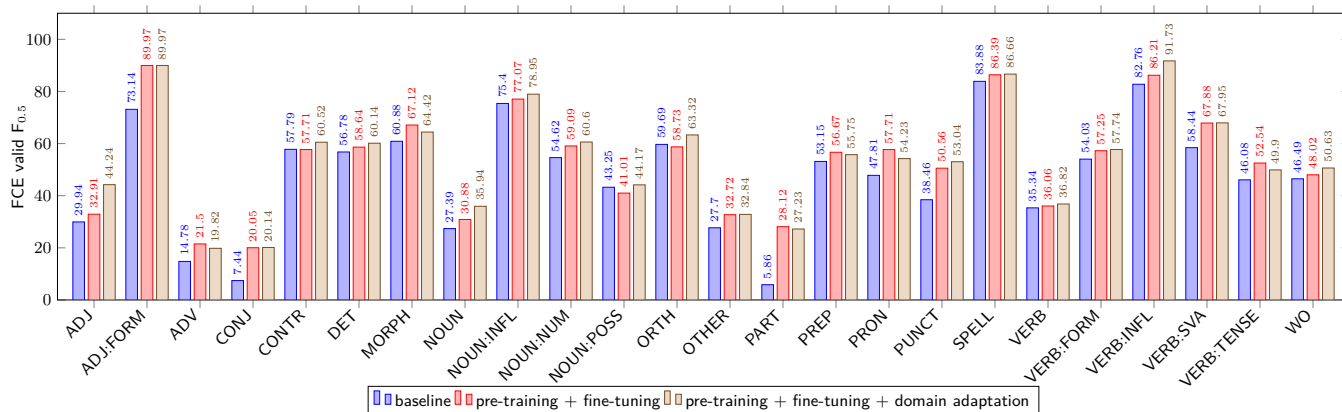
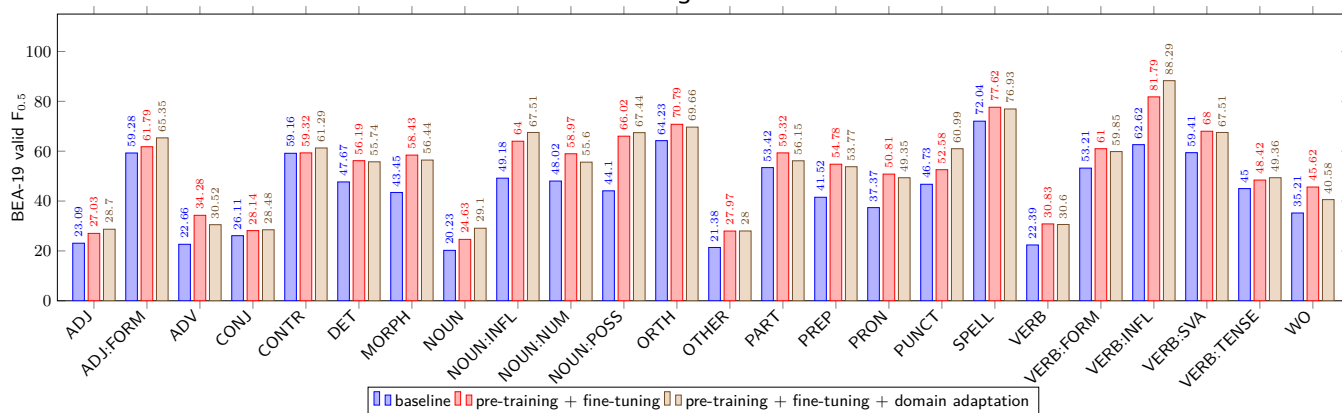
# pre-training + fine-tuning



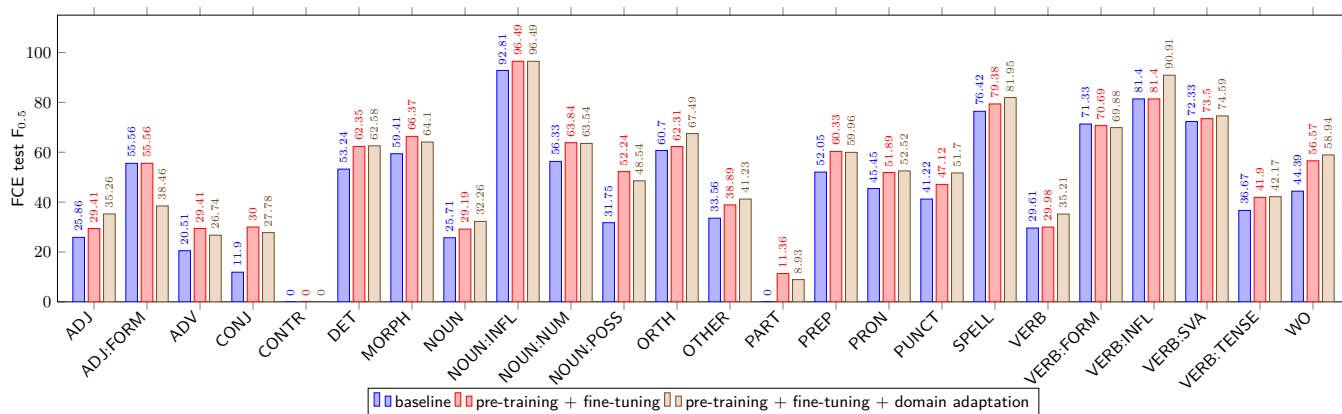
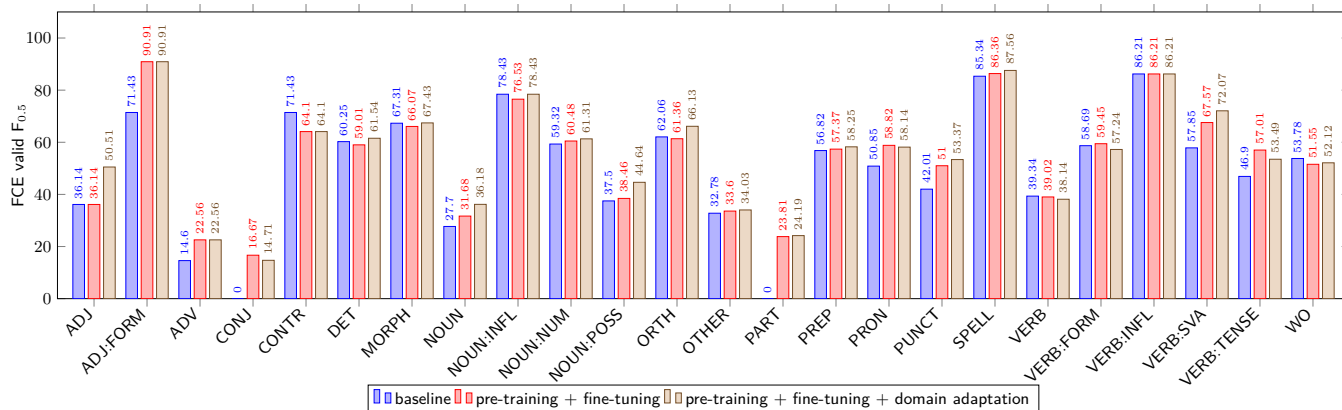
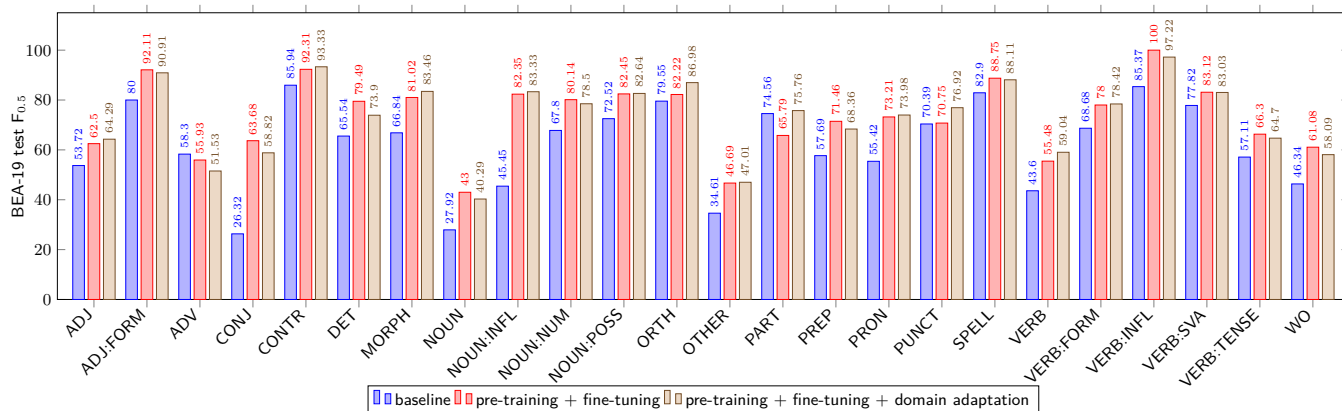
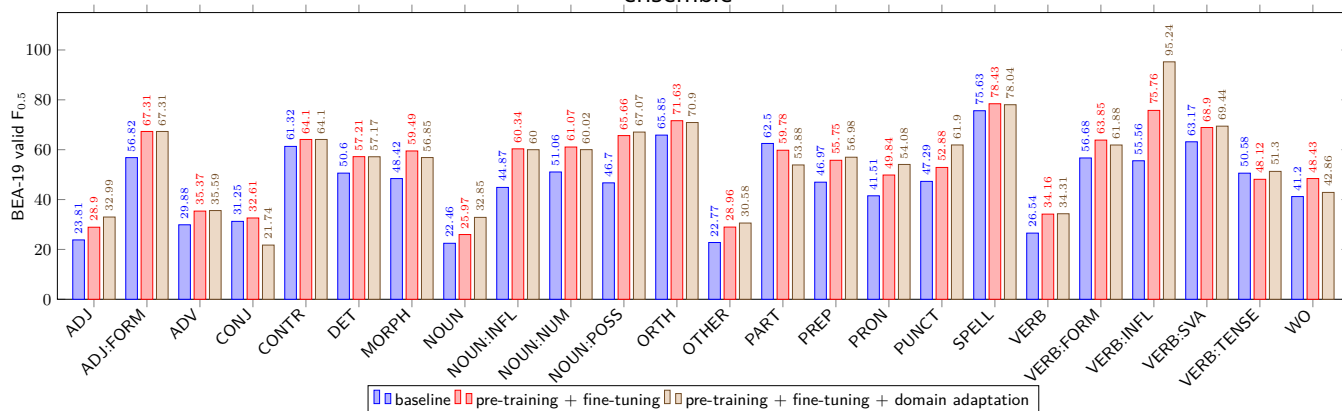
# pre-training + fine-tuning + domain adaptation

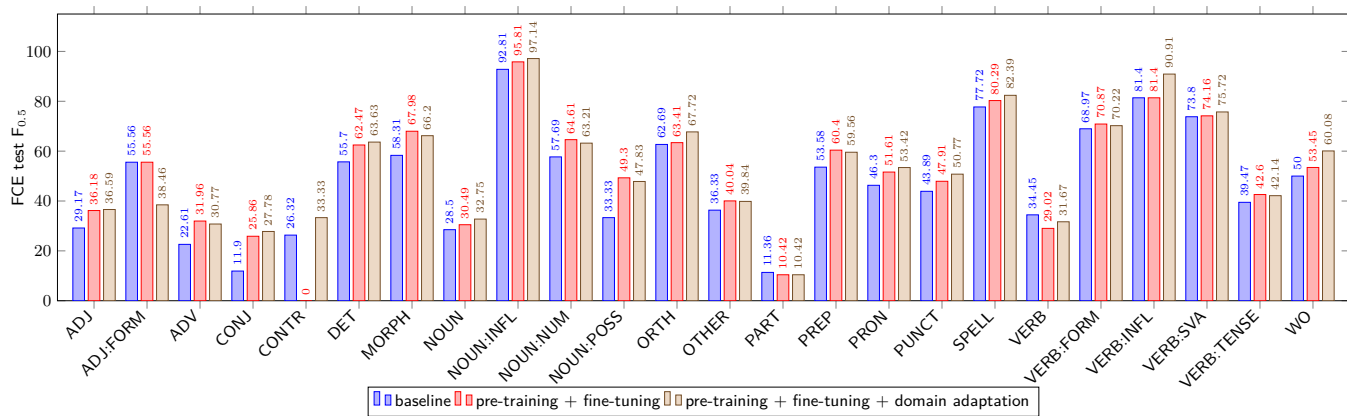
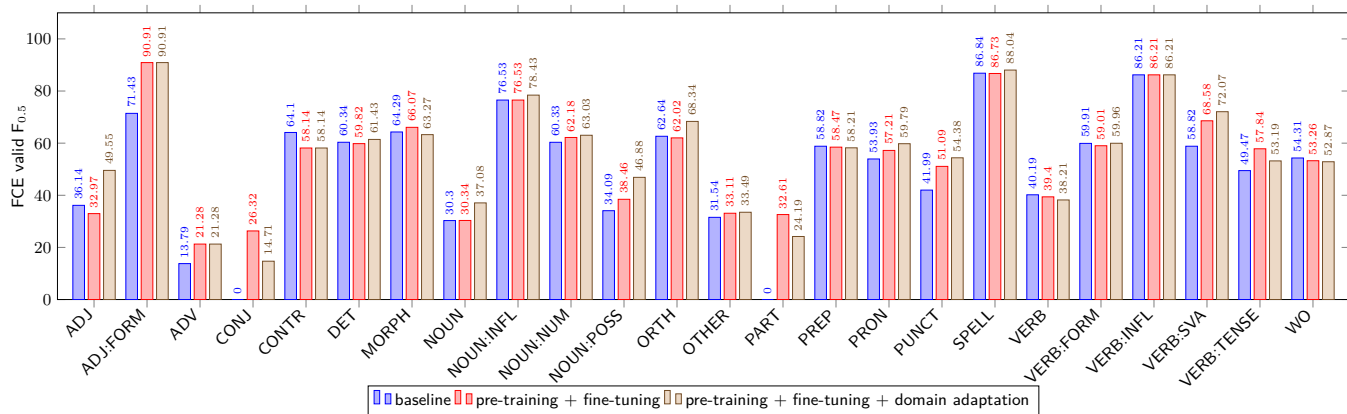
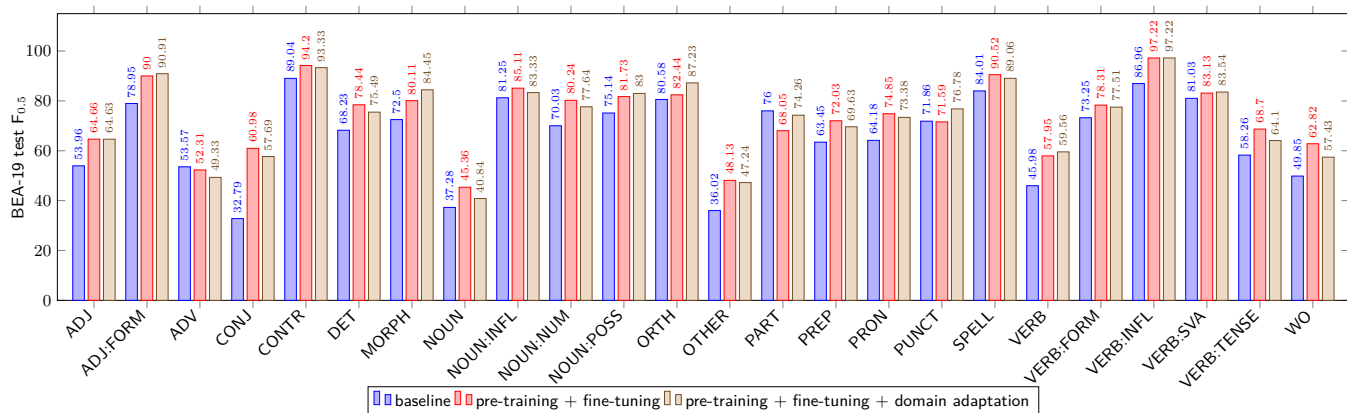
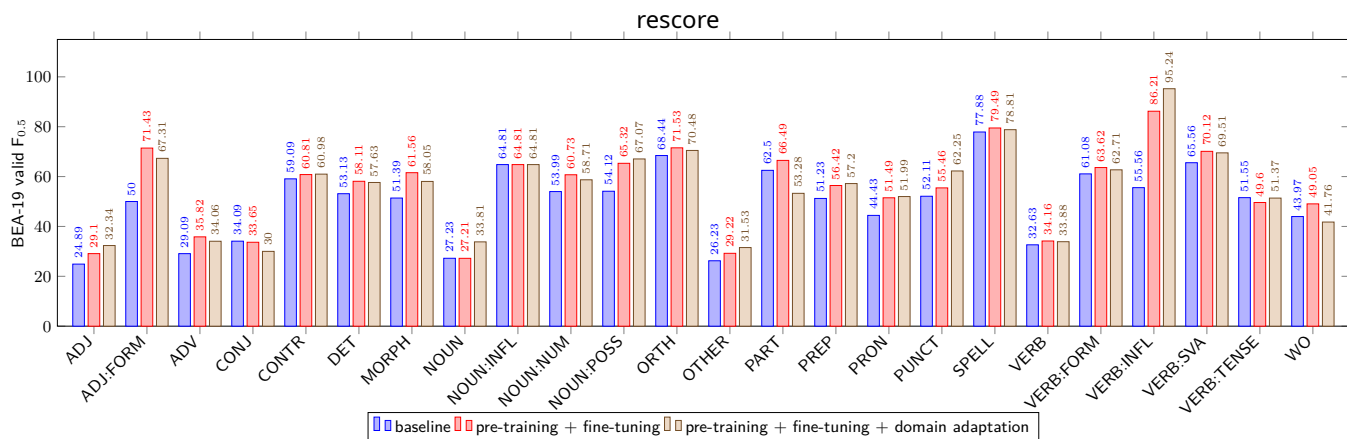


single

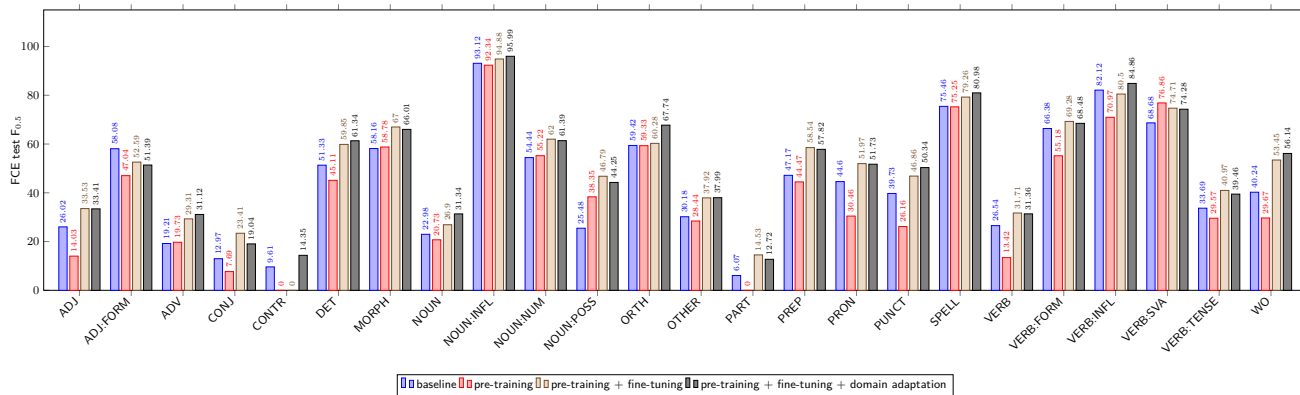
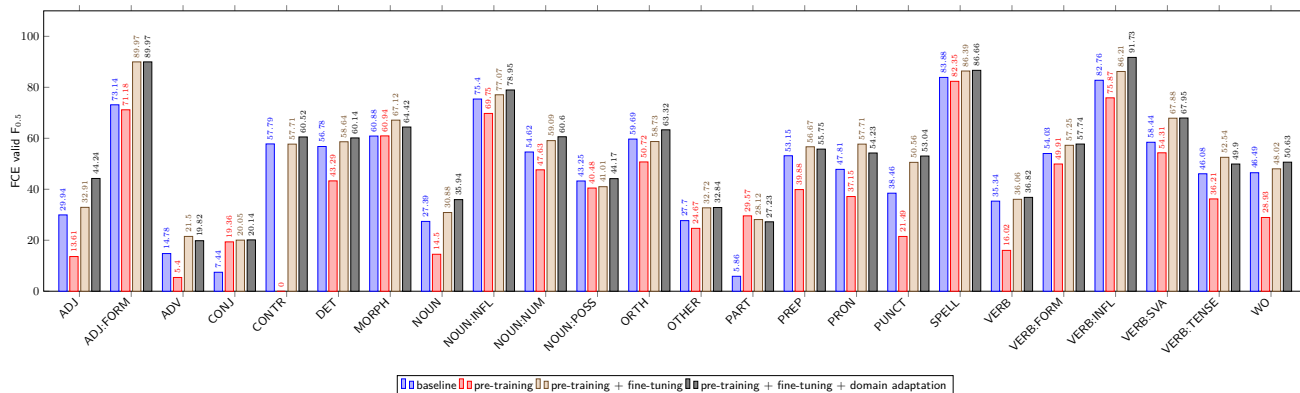
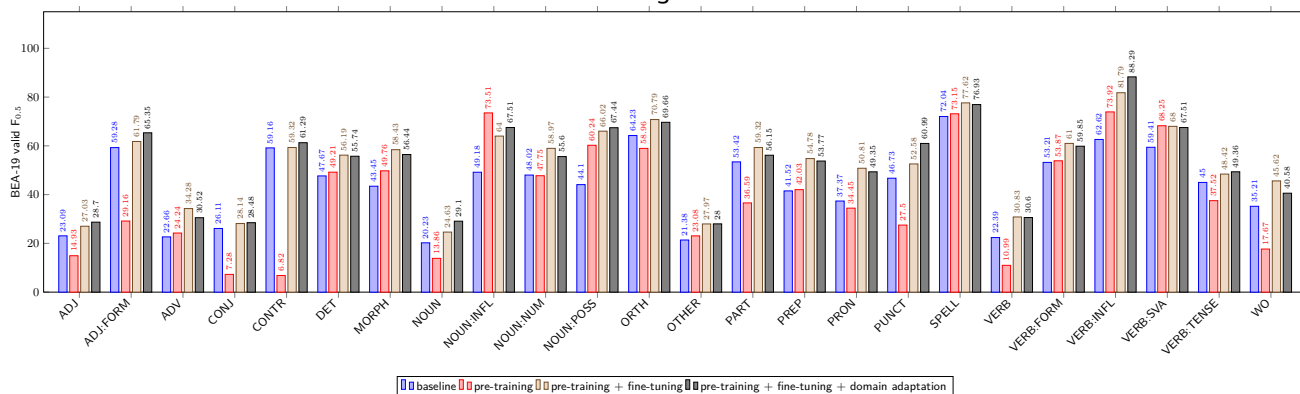


ensemble

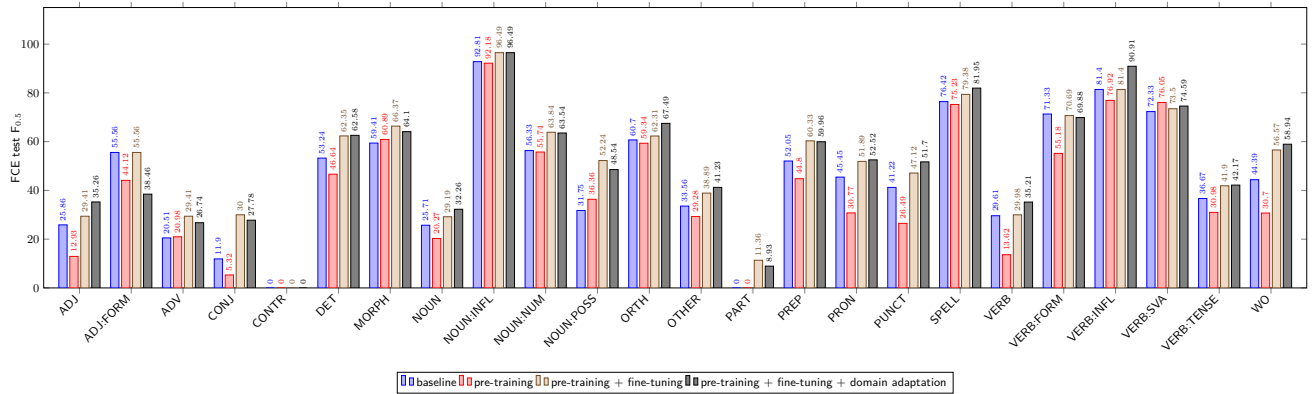
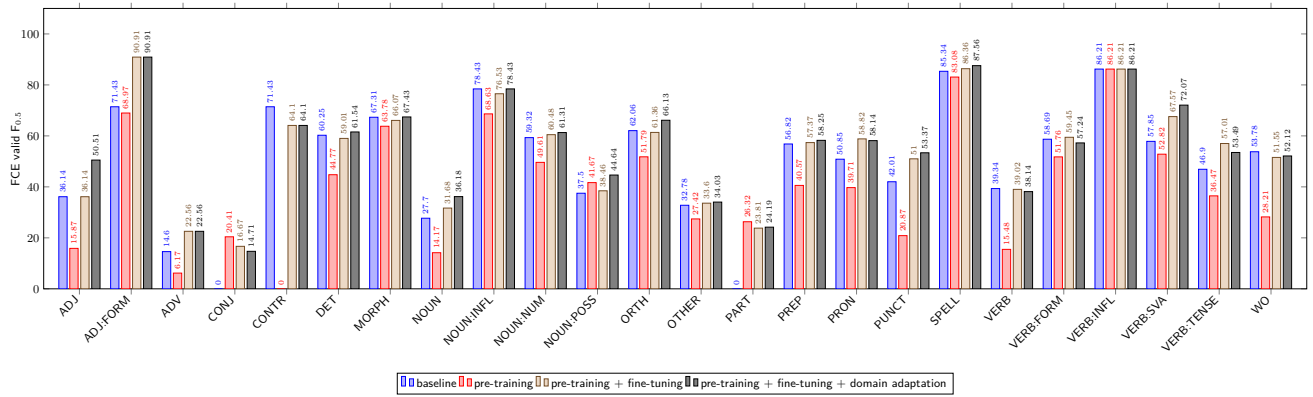
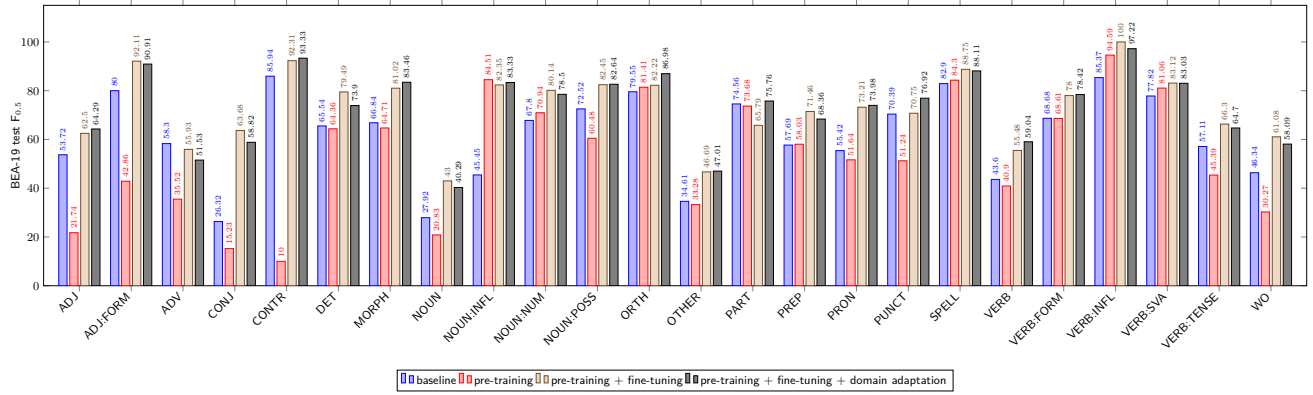
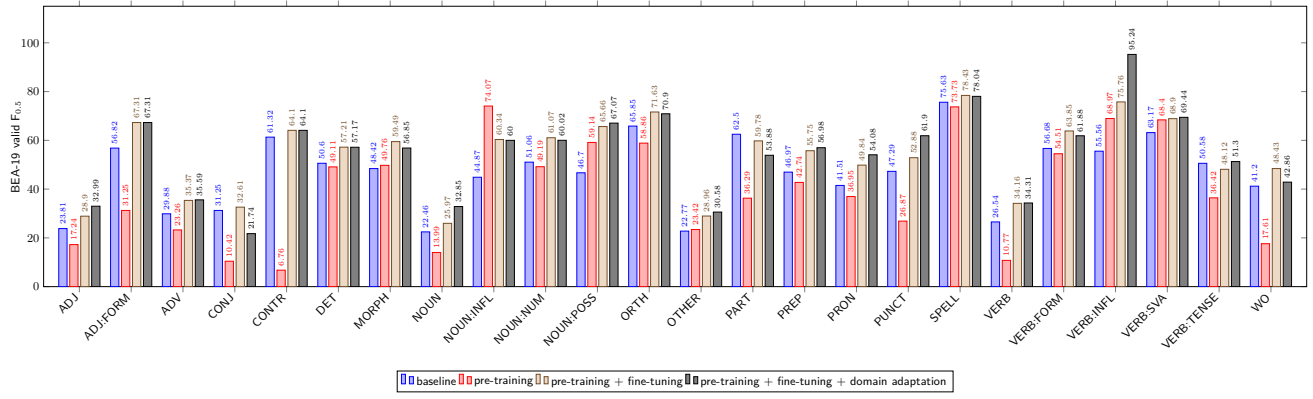




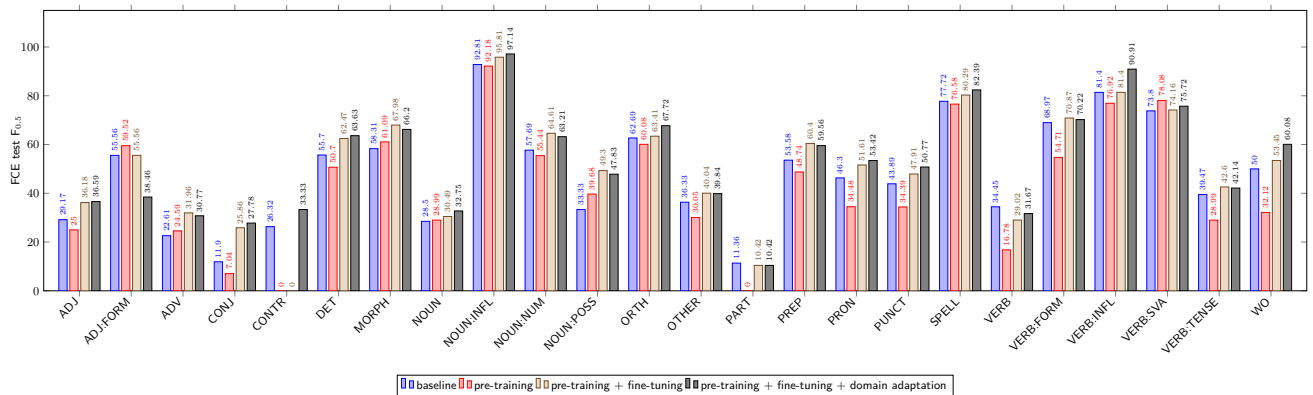
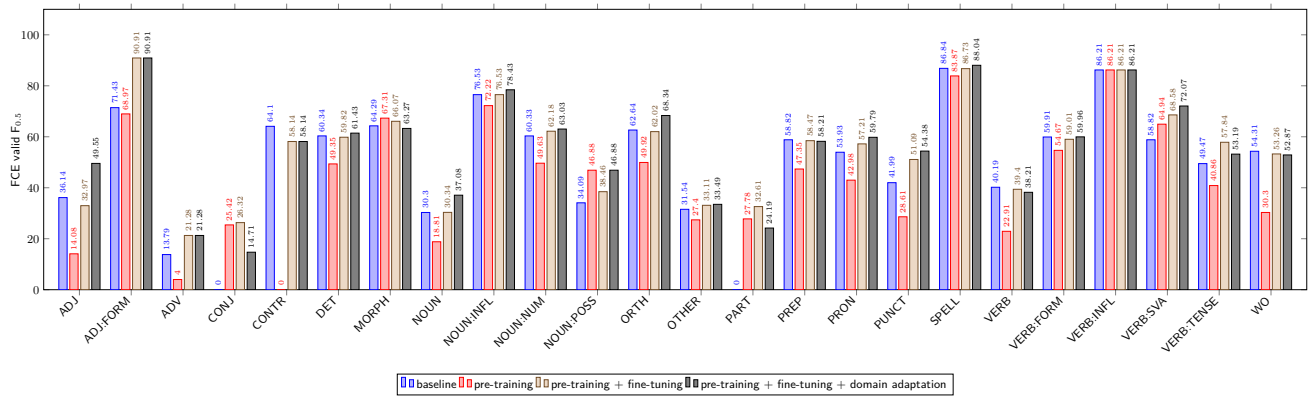
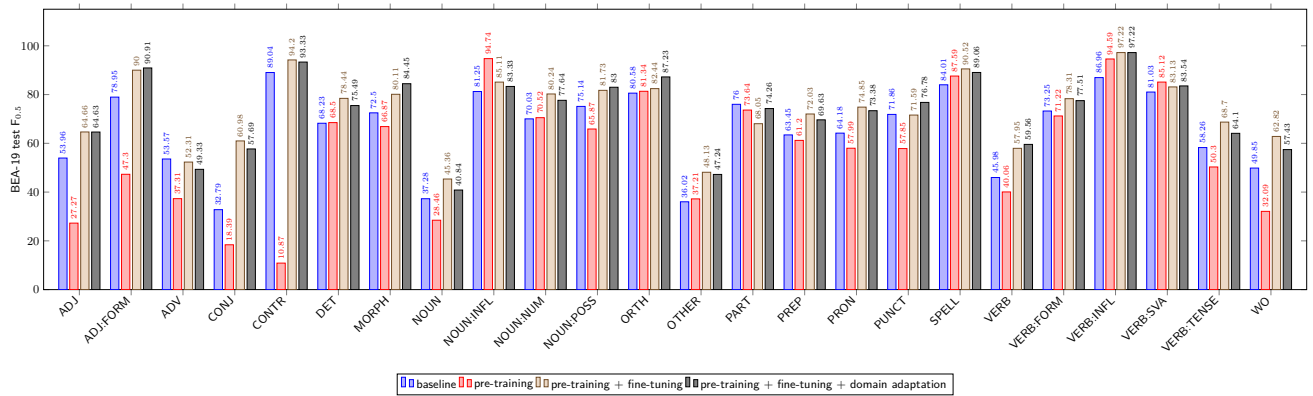
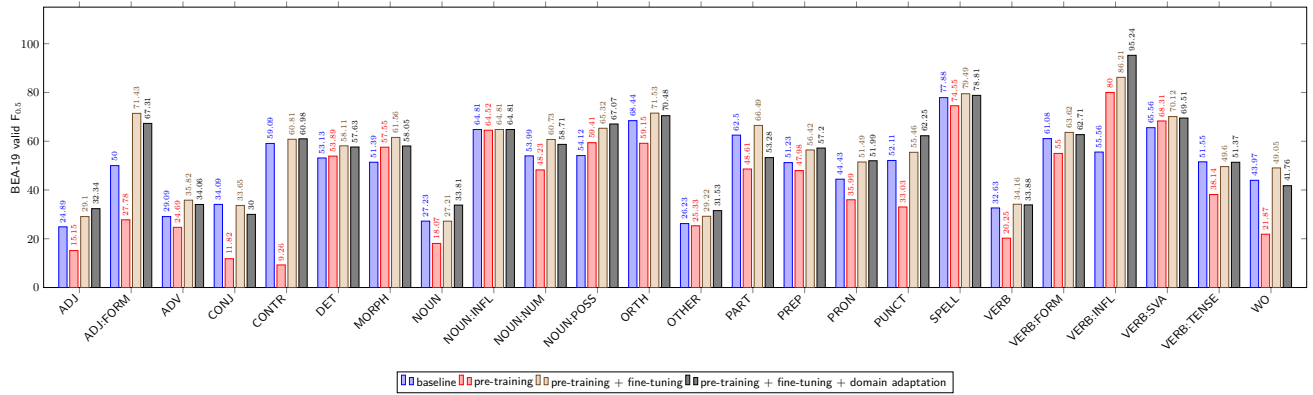
single



## ensemble

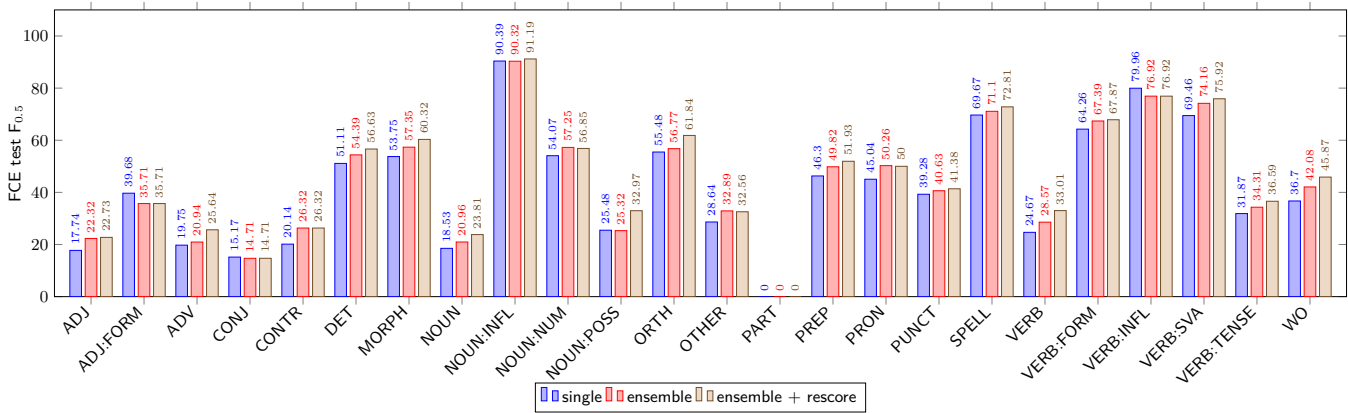
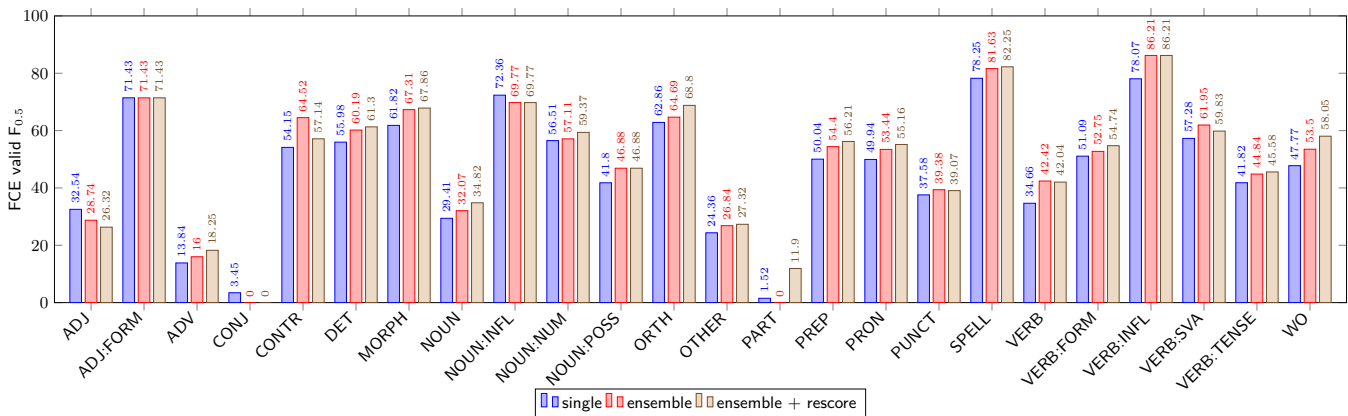
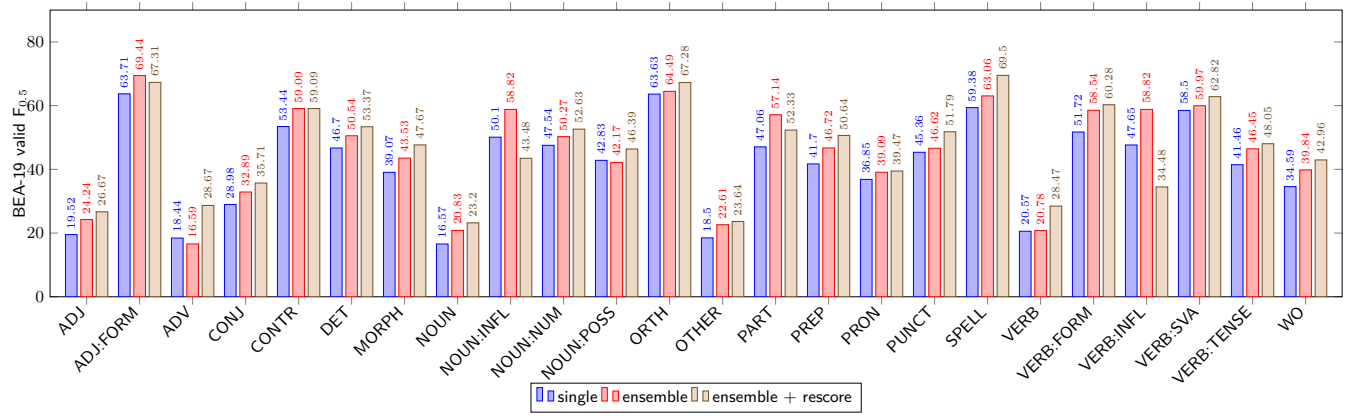


# rescore

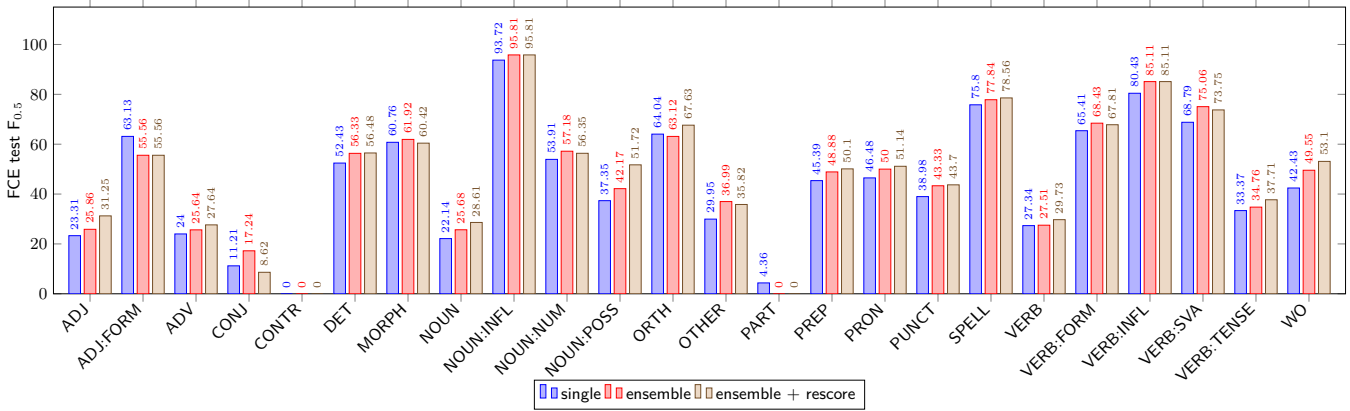
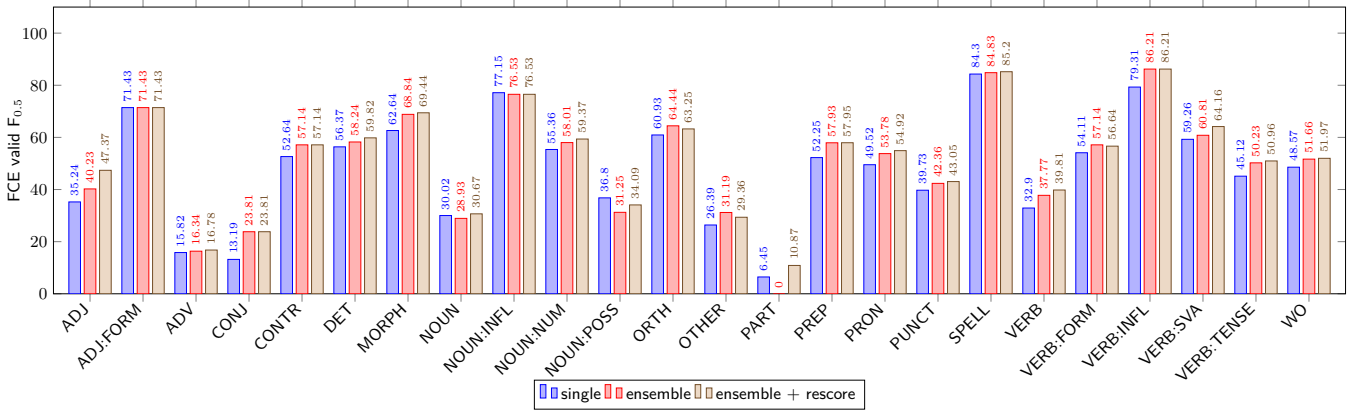
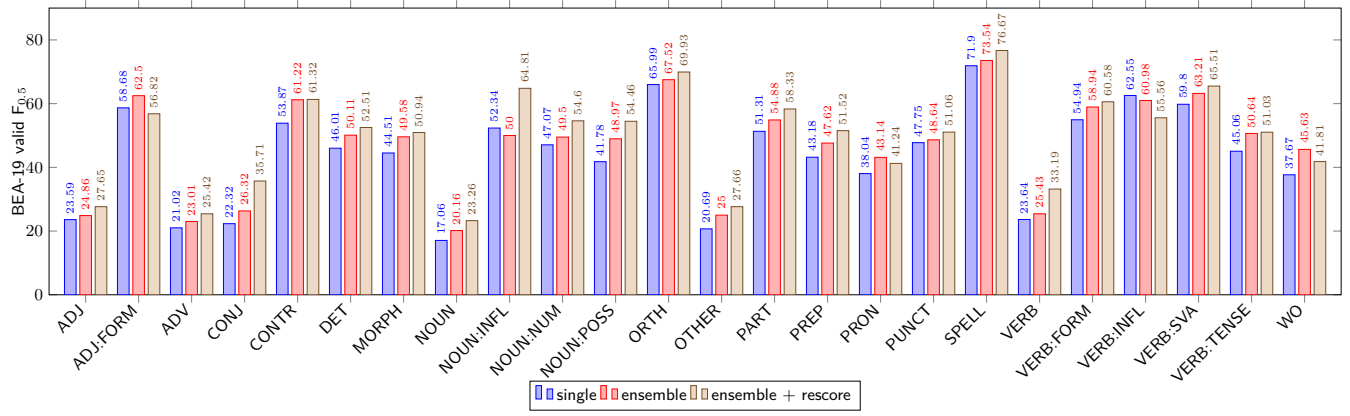




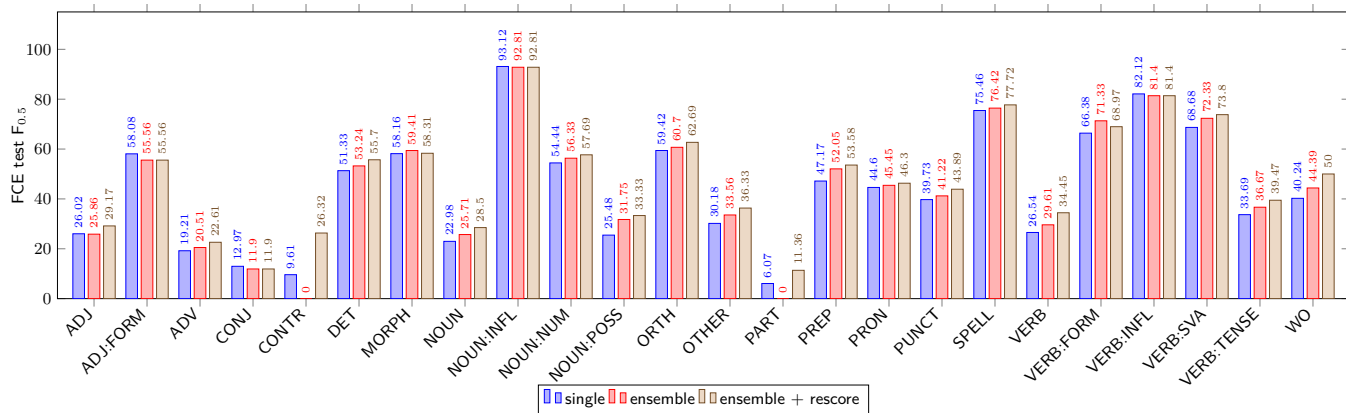
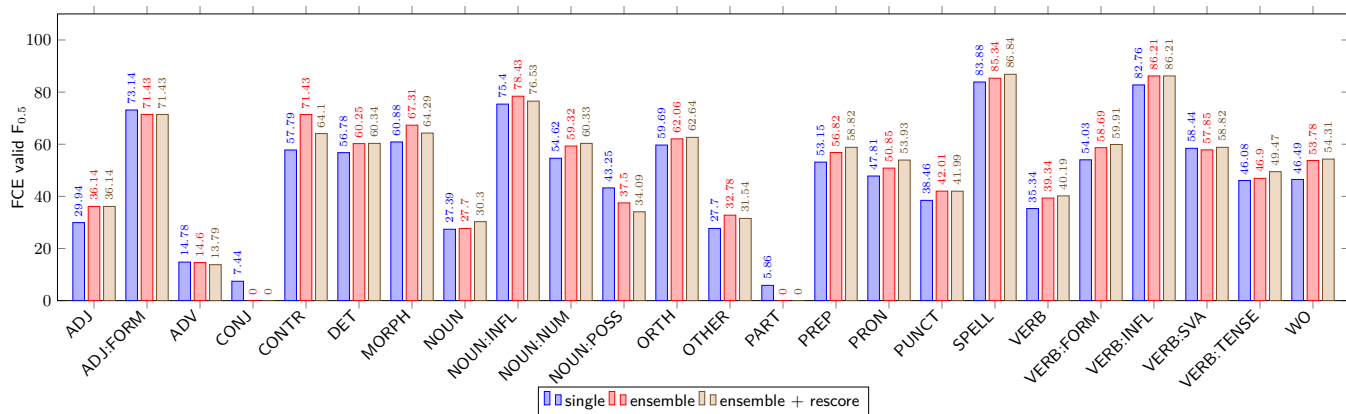
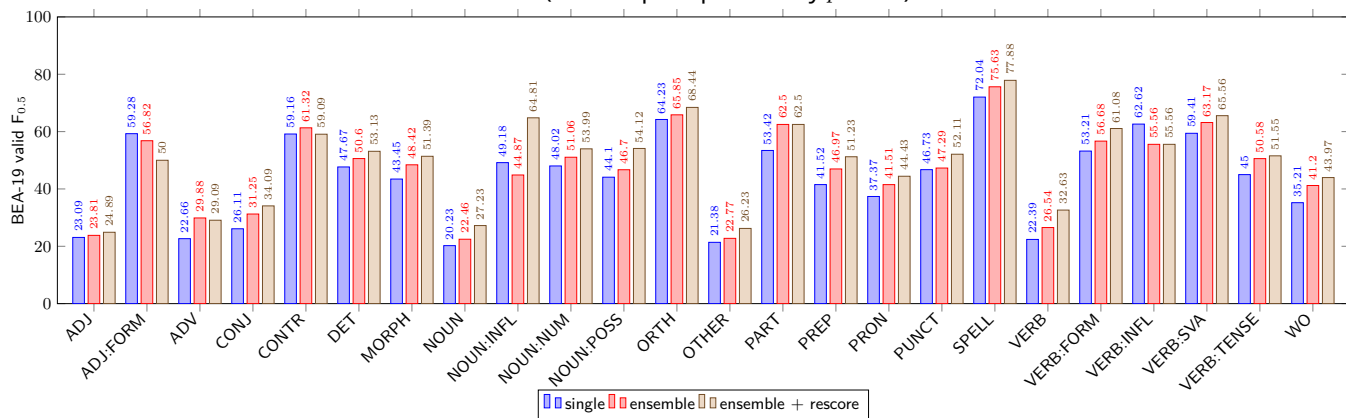
baseline (BPE-dropout probability  $p = 0$ )



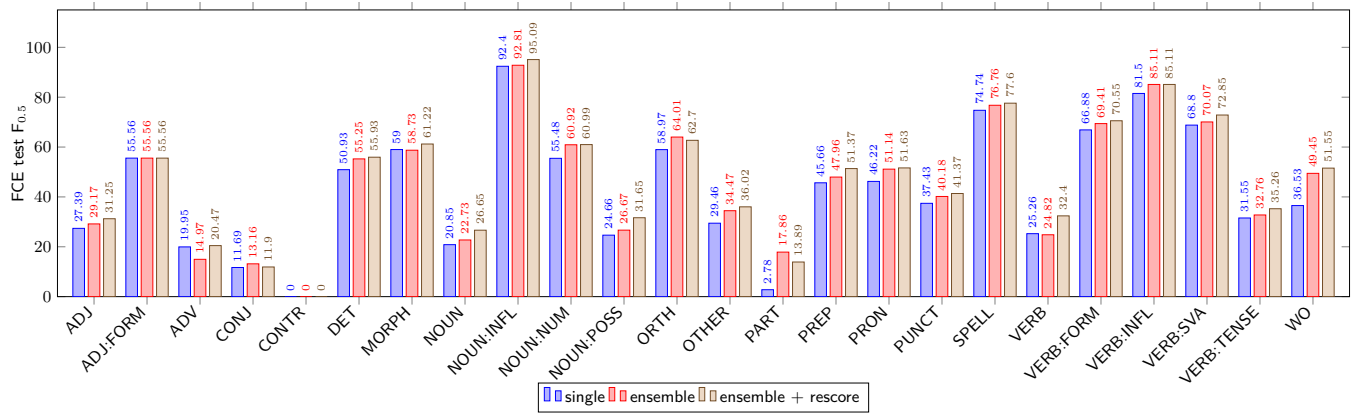
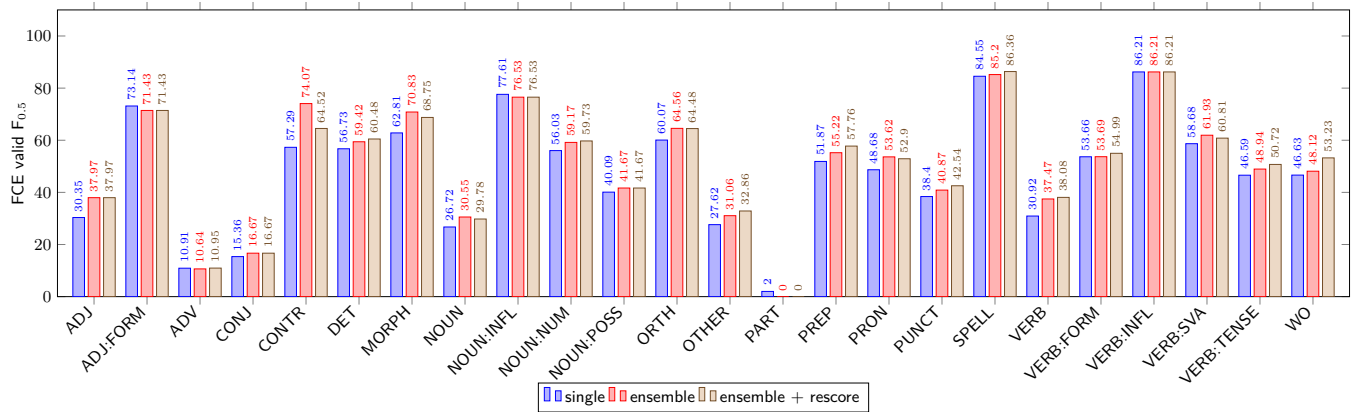
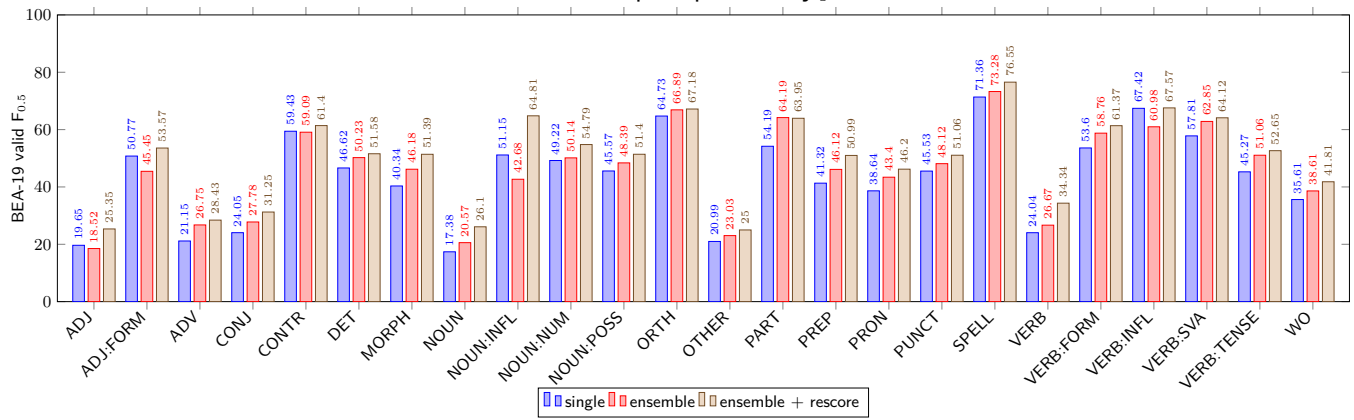
baseline (BPE-dropout probability  $p = 0.05$ )



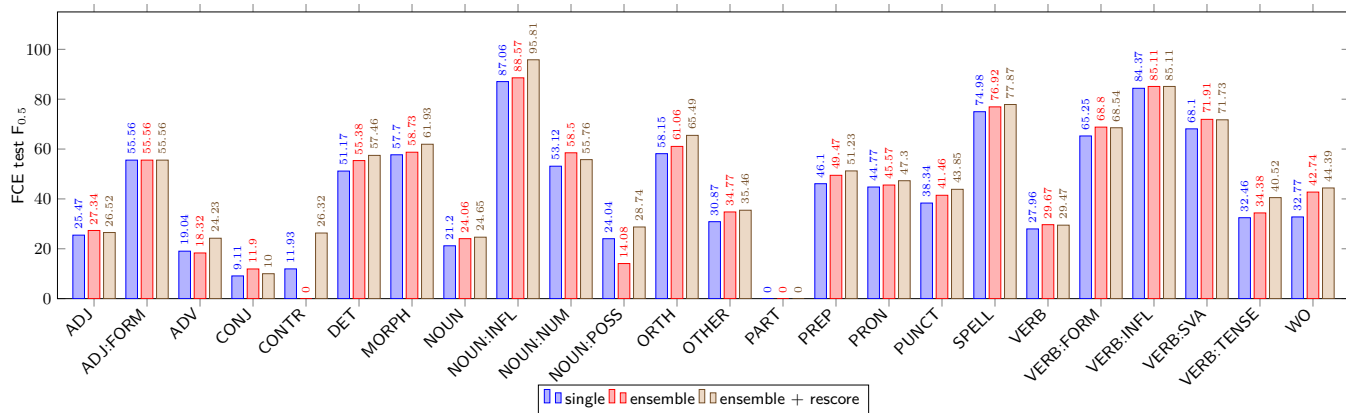
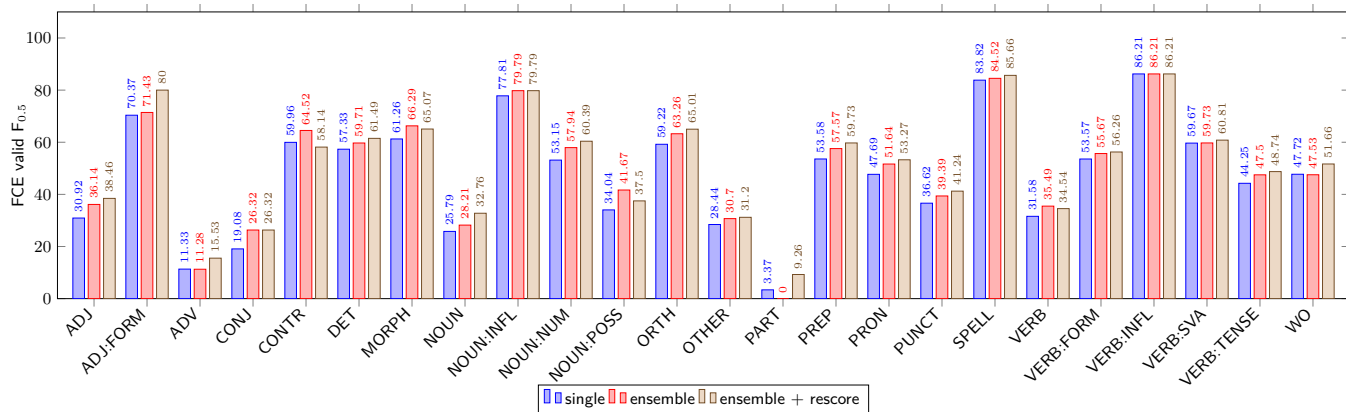
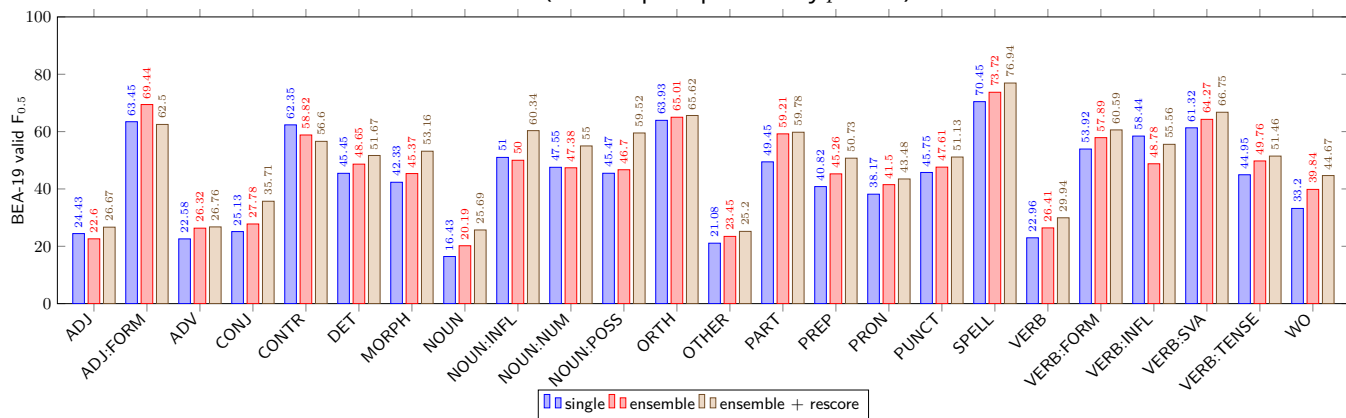
baseline (BPE-dropout probability  $p = 0.1$ )



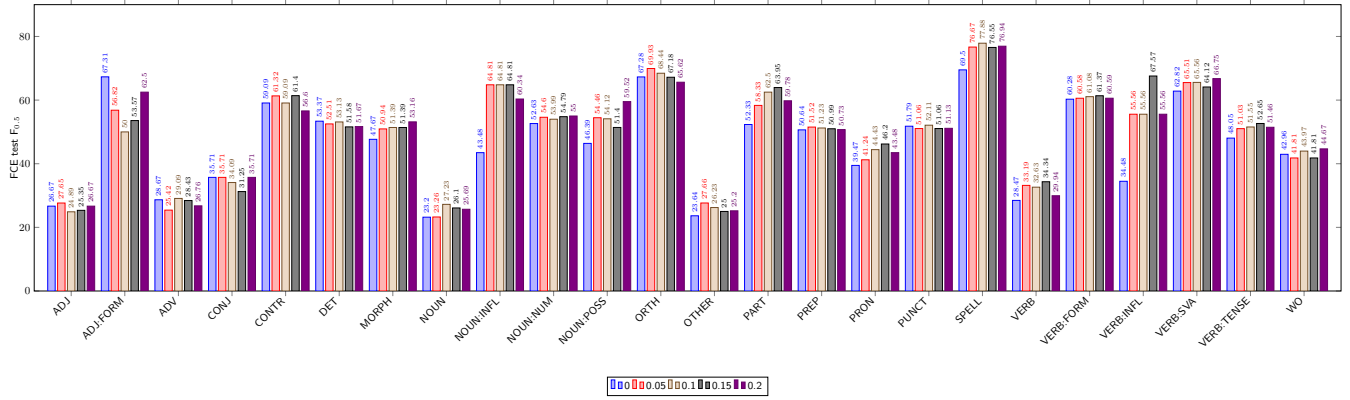
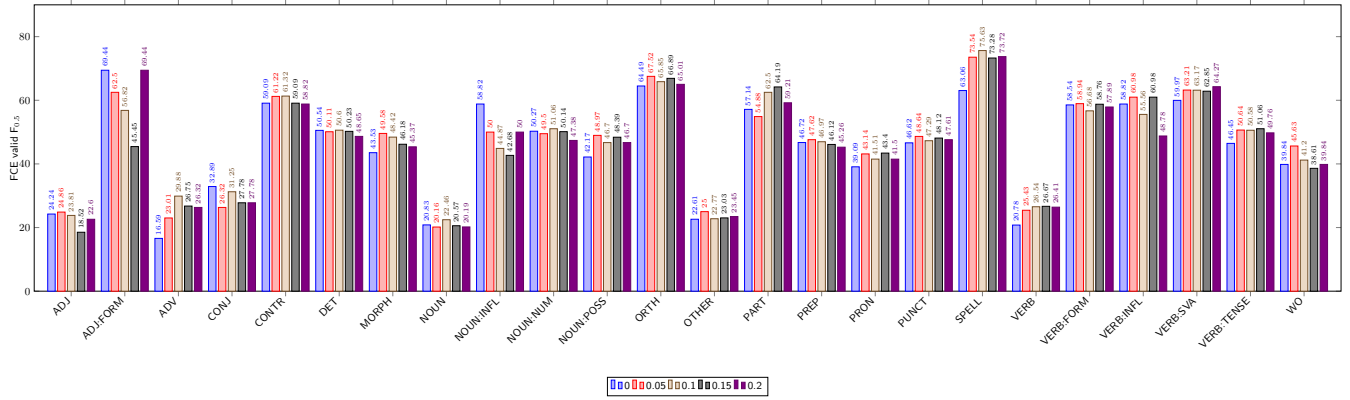
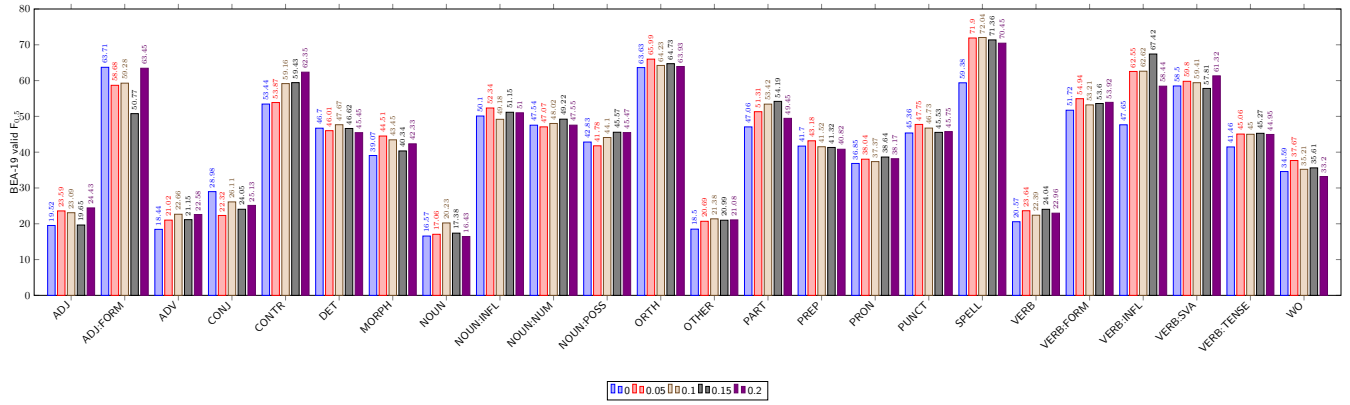
baseline (BPE-dropout probability  $p = 0.15$ )



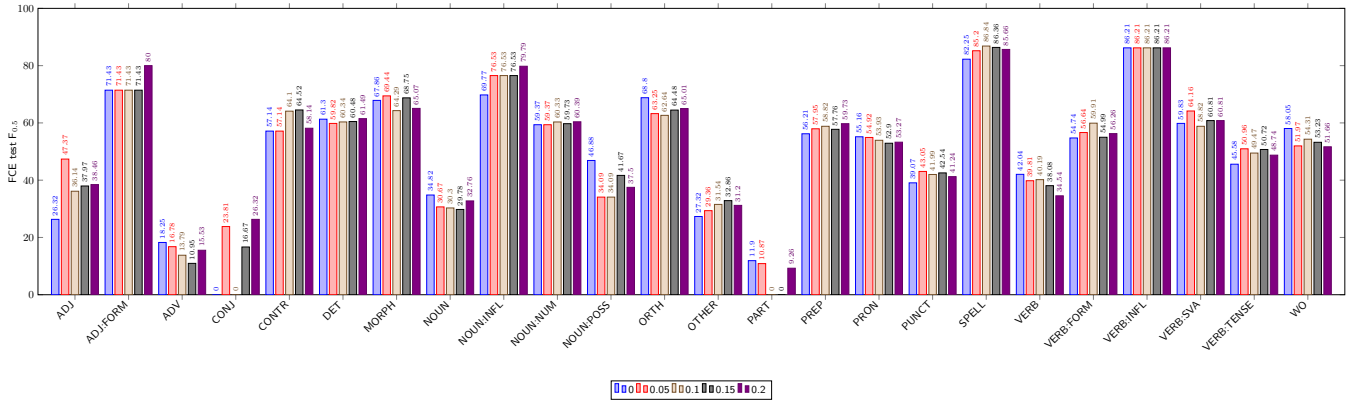
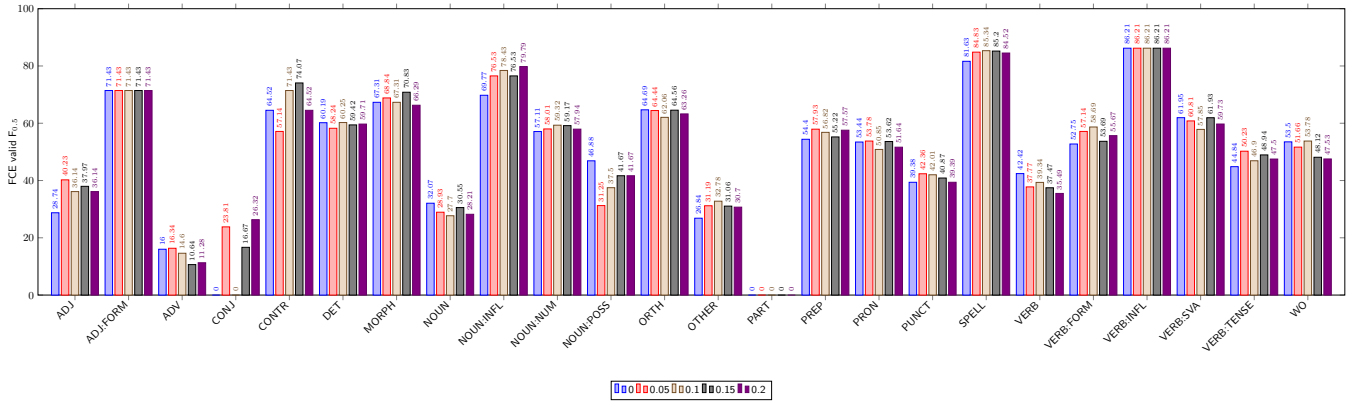
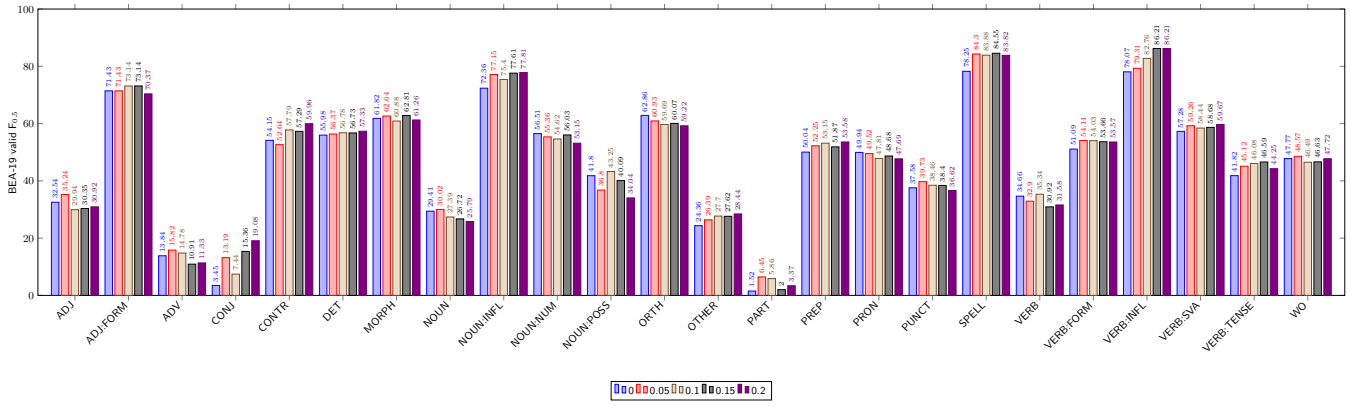
baseline (BPE-dropout probability  $p = 0.2$ )



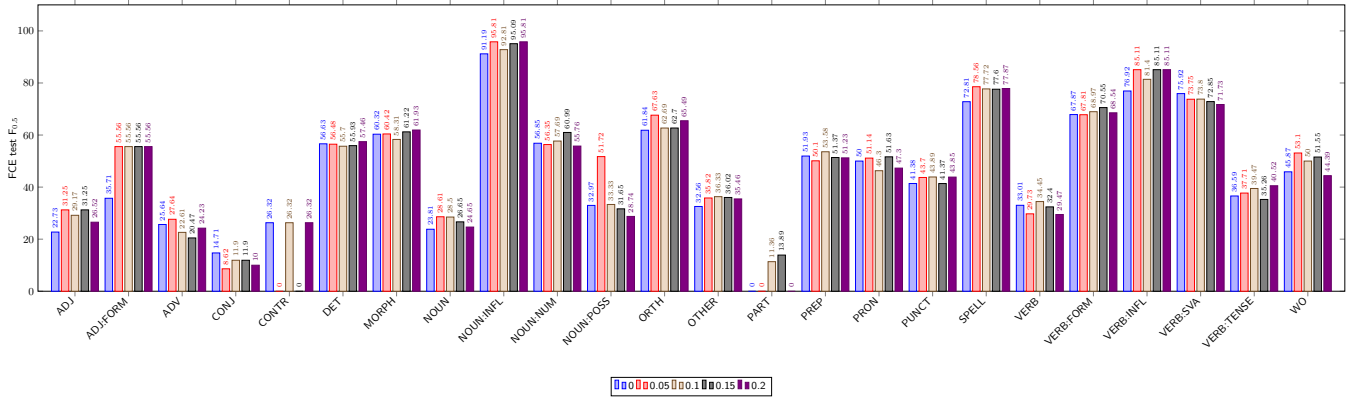
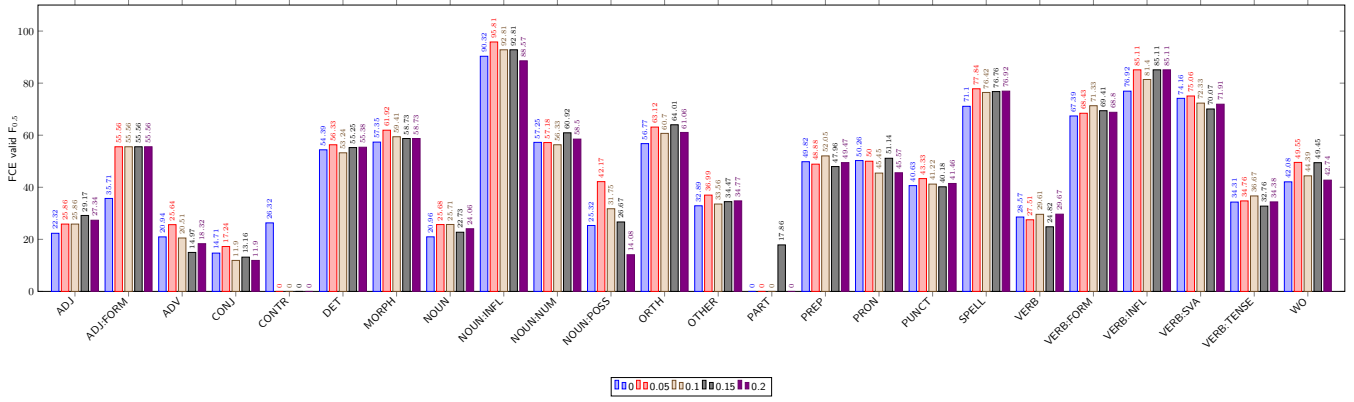
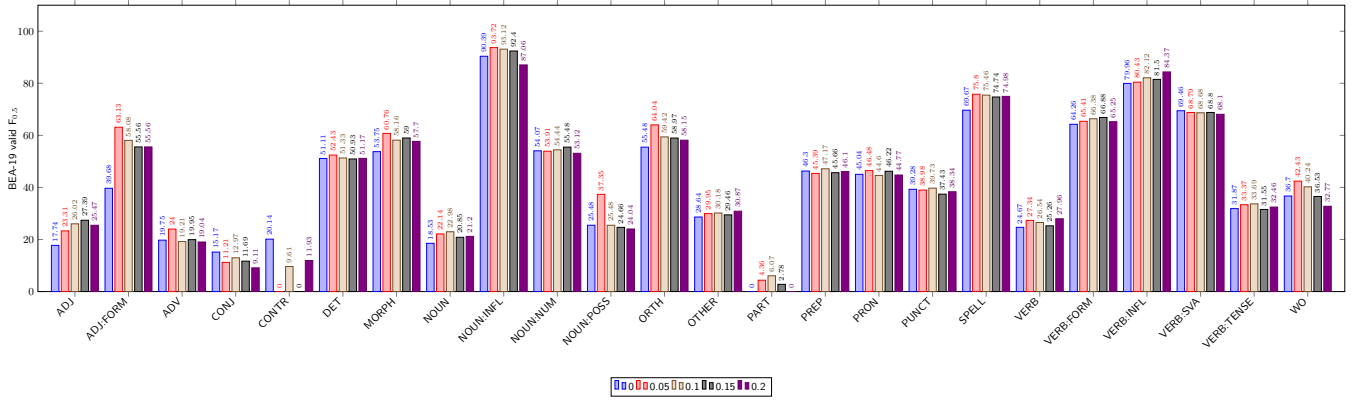
baseline (single)



baseline (ensemble)

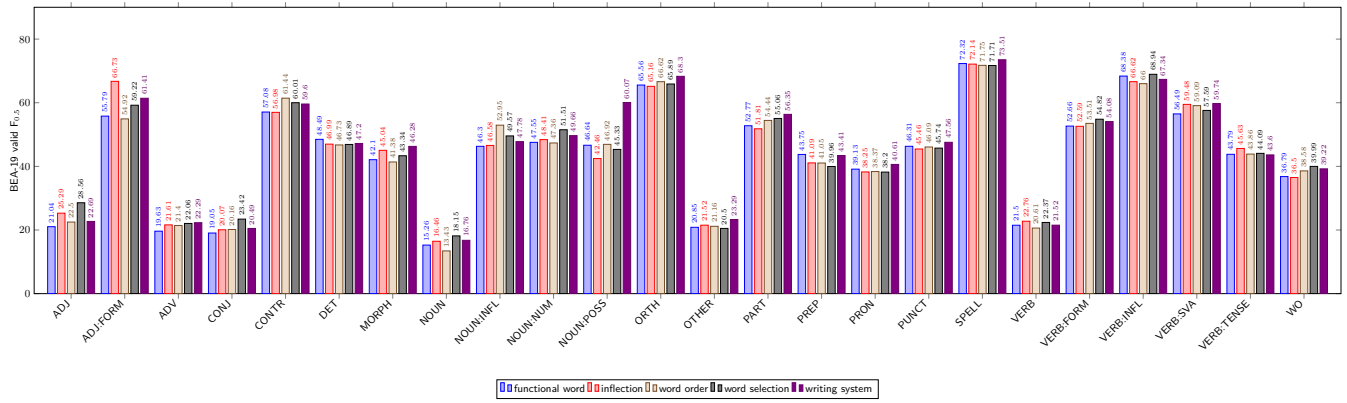


baseline (ensemble + rescore)

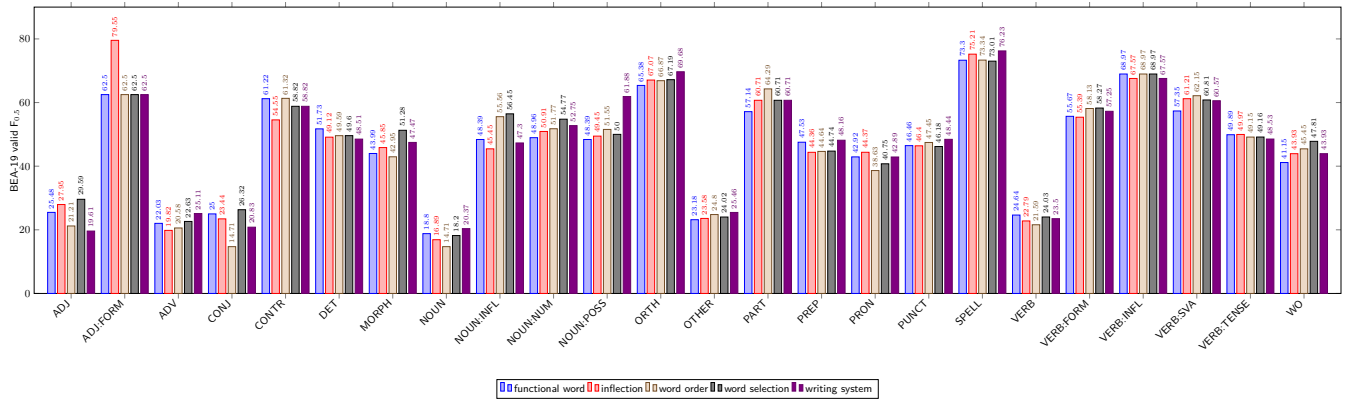




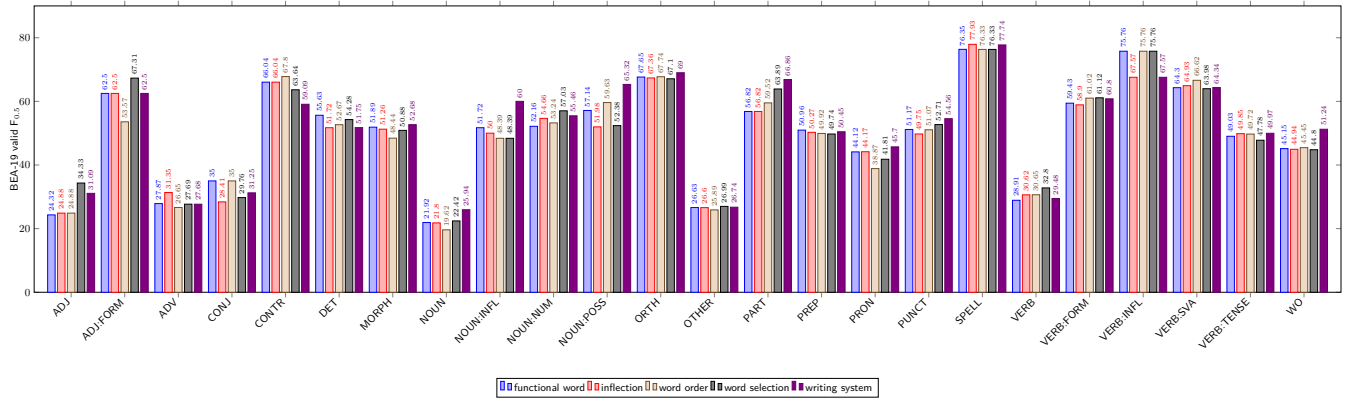
8M x 10 epochs pre-training (only 1 category) + fine-tuning (single)



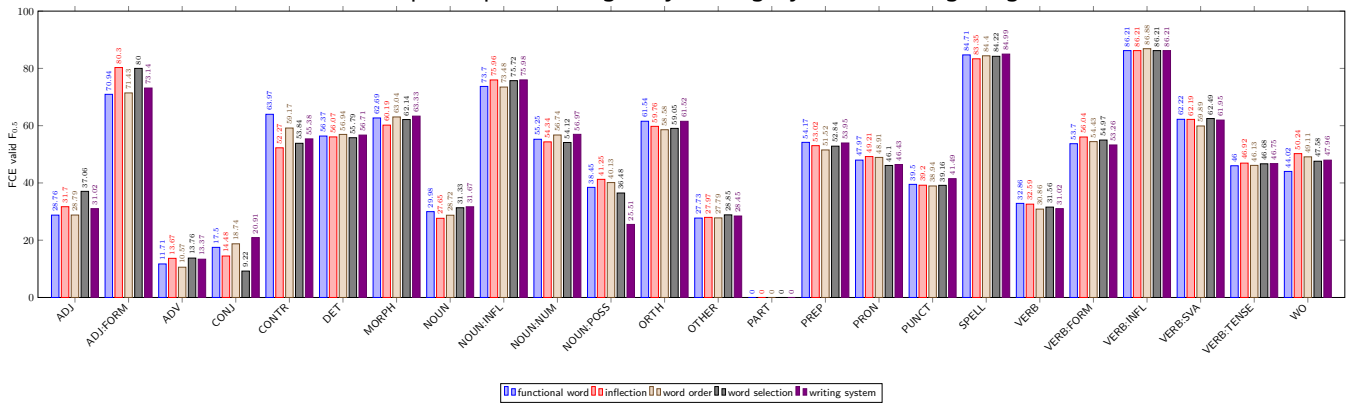
8M x 10 epochs pre-training (only 1 category) + fine-tuning (ensemble)



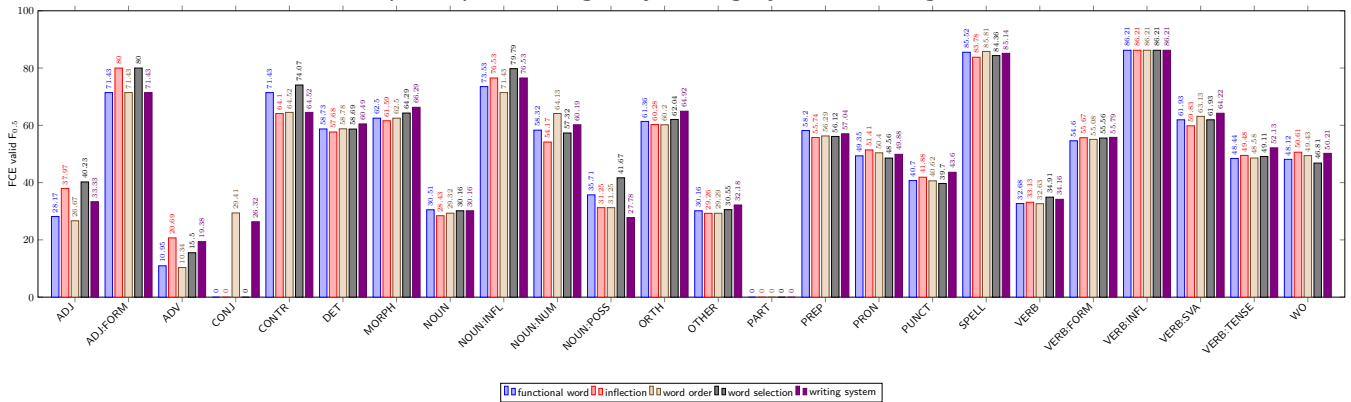
8M x 10 epochs pre-training (only 1 category) + fine-tuning (ensemble + rescore)



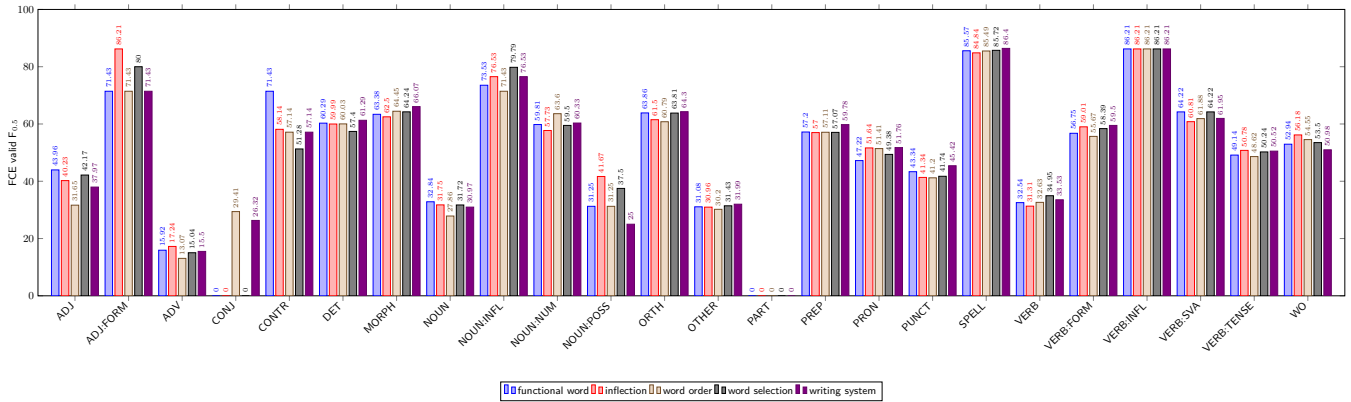
8M x 10 epochs pre-training (only 1 category) + fine-tuning (single)



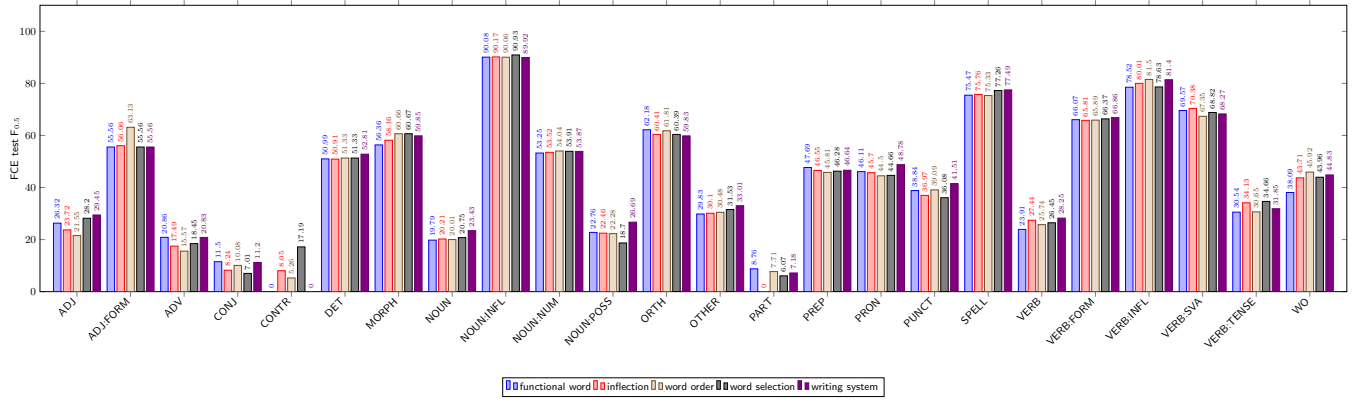
8M x 10 epochs pre-training (only 1 category) + fine-tuning (ensemble)



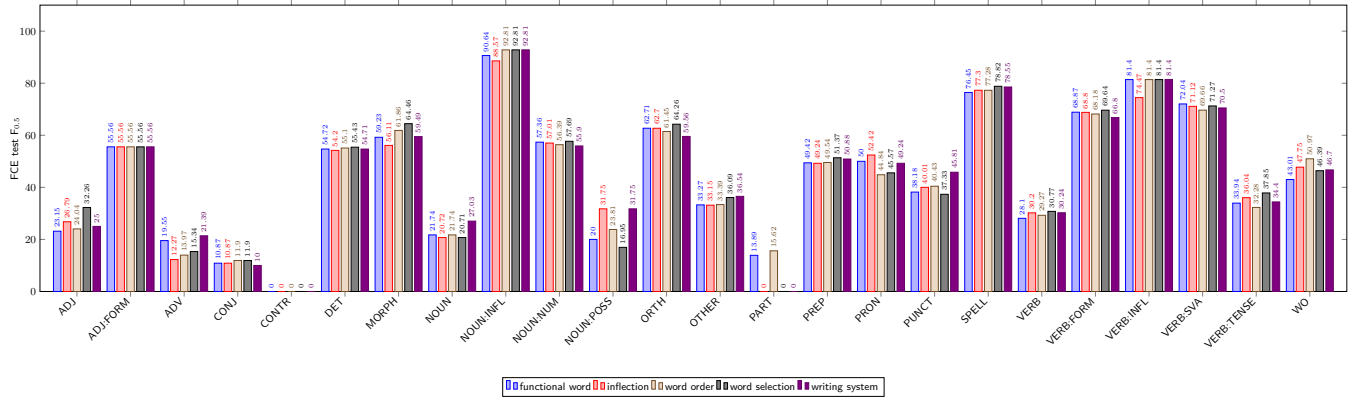
8M x 10 epochs pre-training (only 1 category) + fine-tuning (ensemble + rescore)



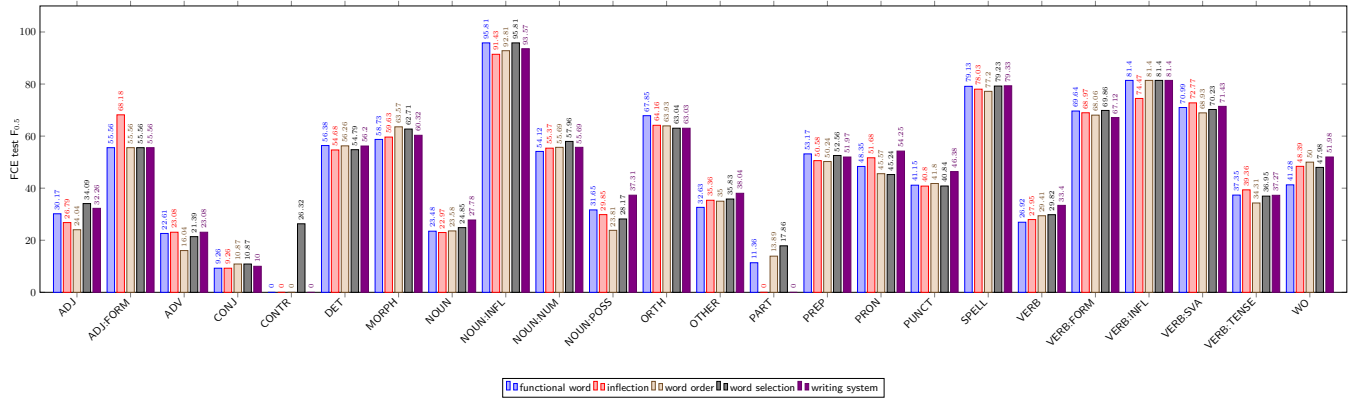
8M x 10 epochs pre-training (only 1 category) + fine-tuning (single)



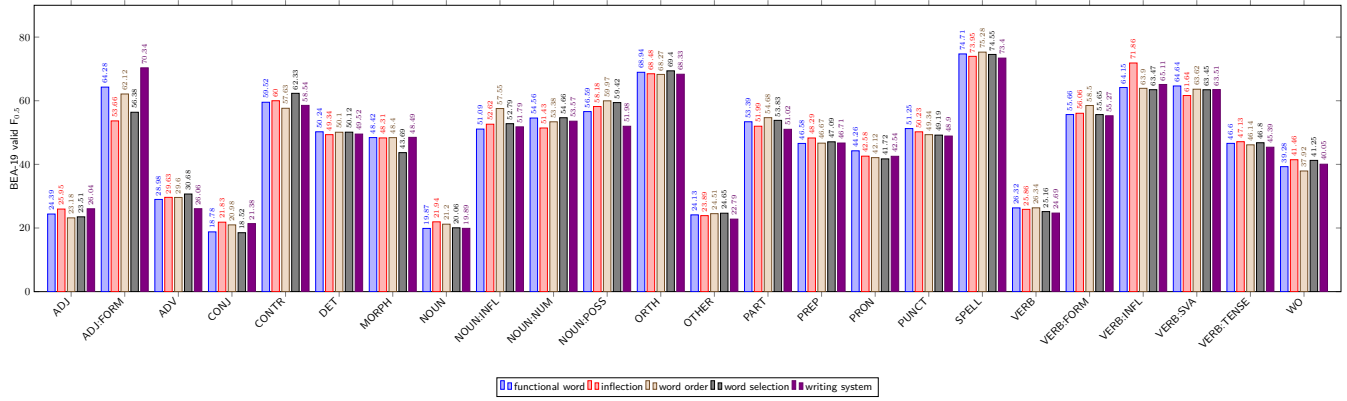
8M x 10 epochs pre-training (only 1 category) + fine-tuning (ensemble)



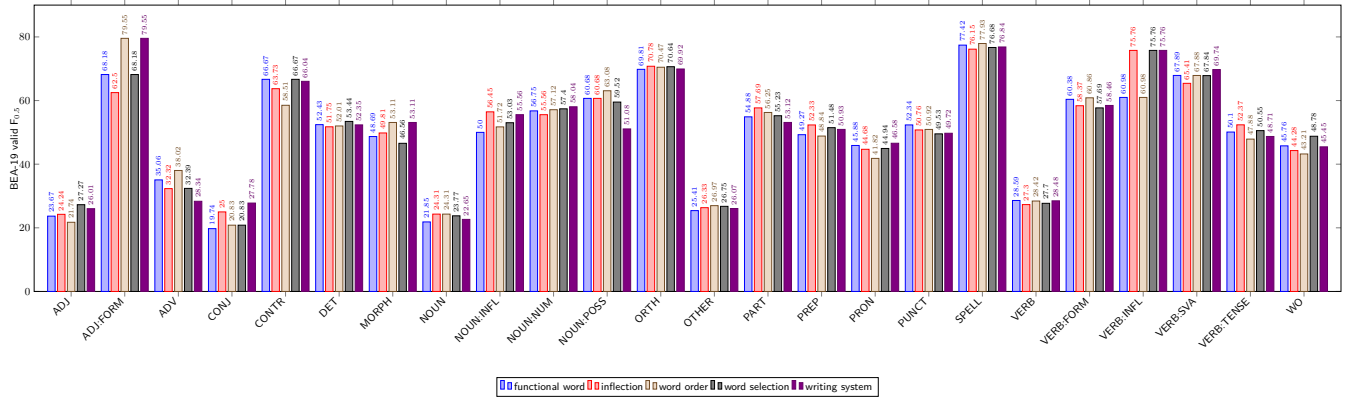
8M x 10 epochs pre-training (only 1 category) + fine-tuning (ensemble + rescore)



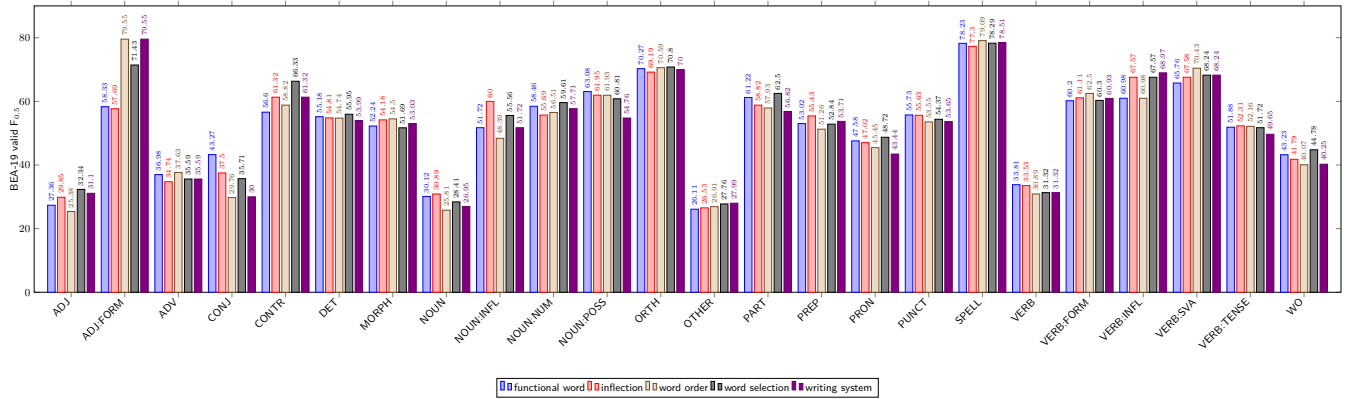
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (single)



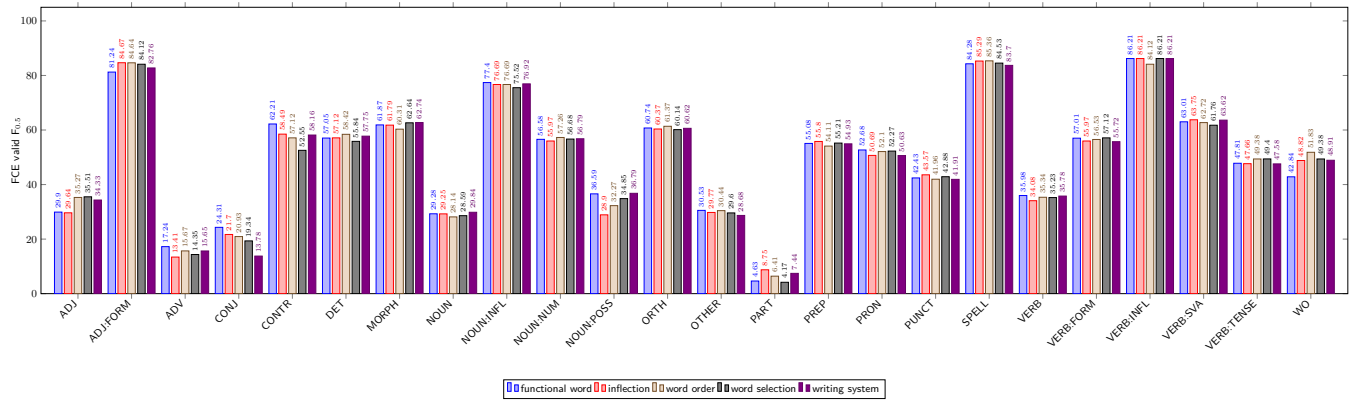
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (ensemble)



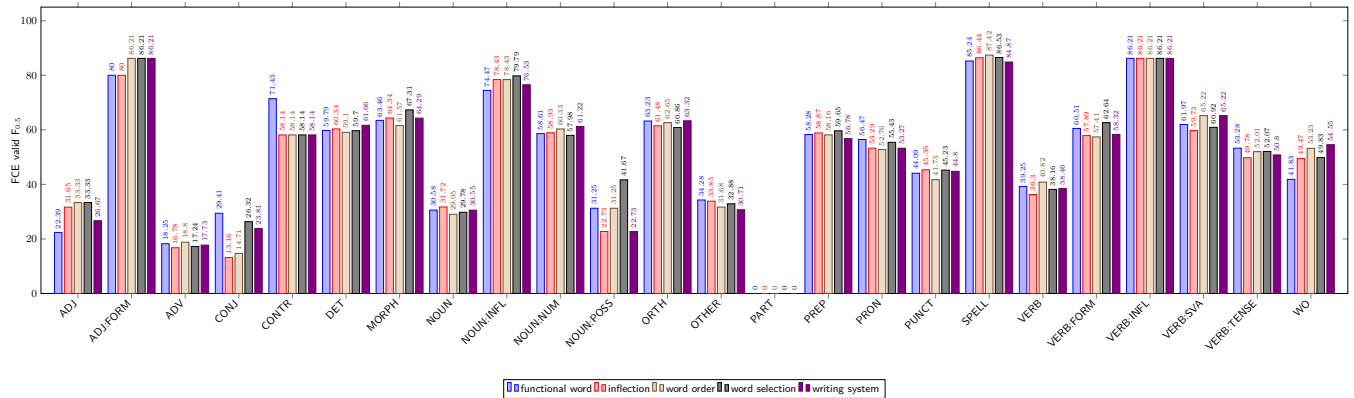
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (ensemble + rescore)



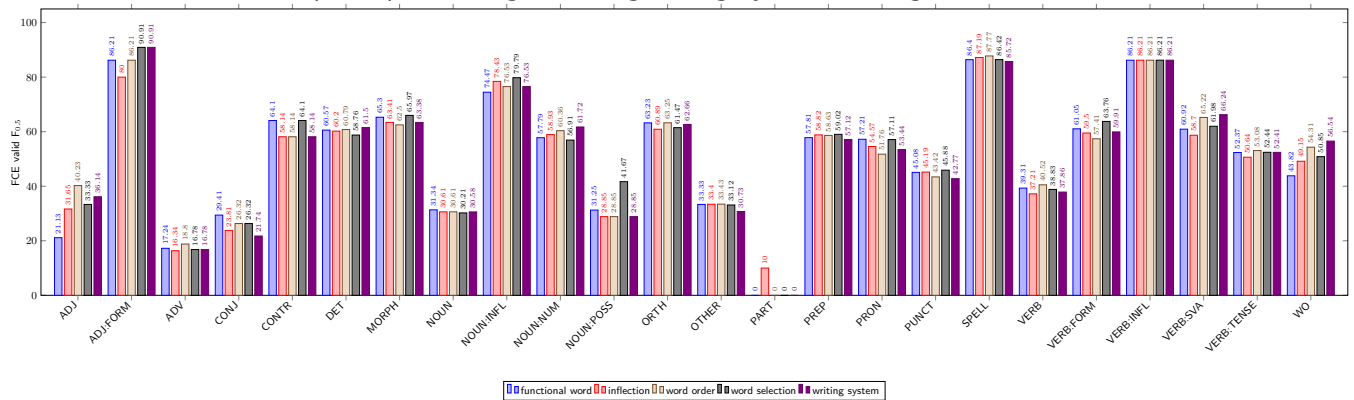
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (single)



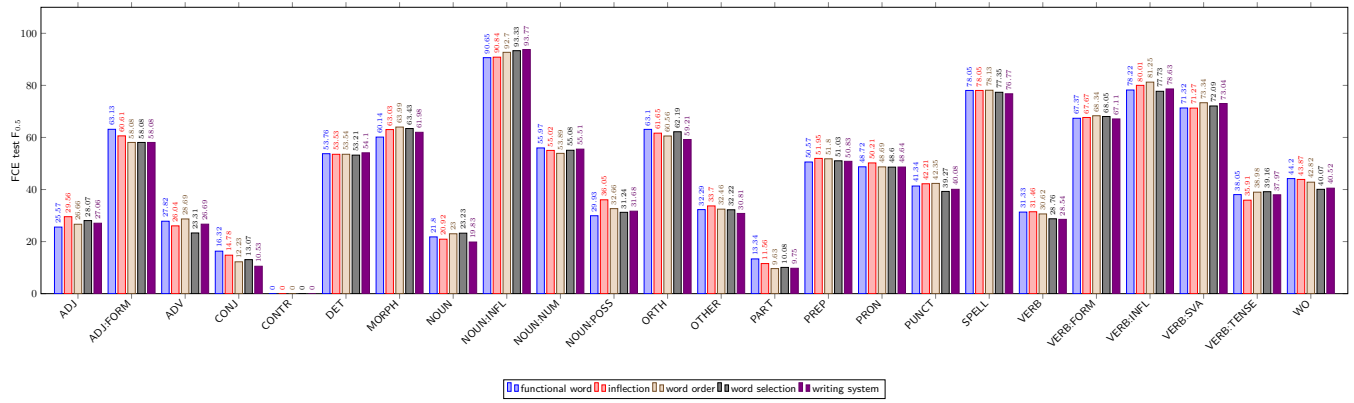
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (ensemble)



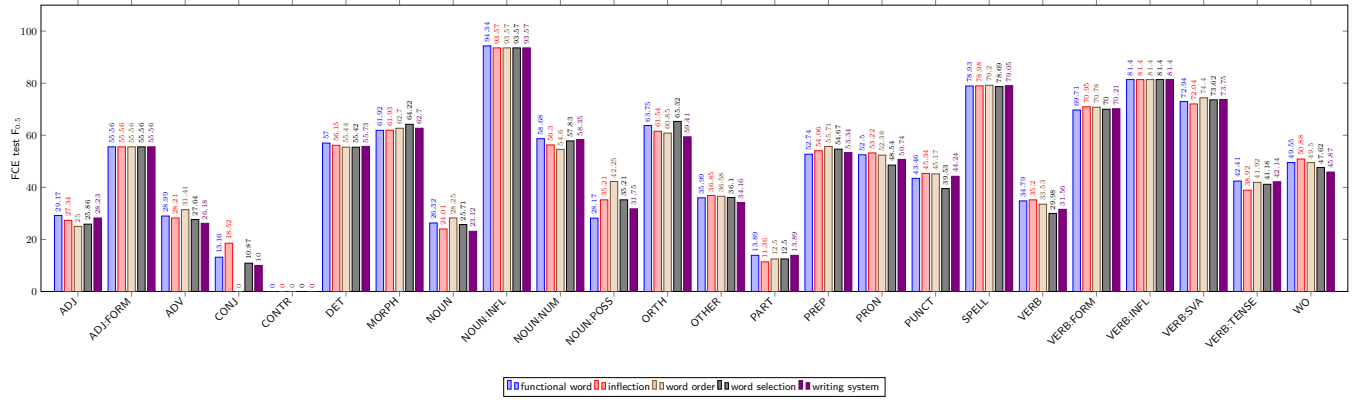
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (ensemble + rescore)



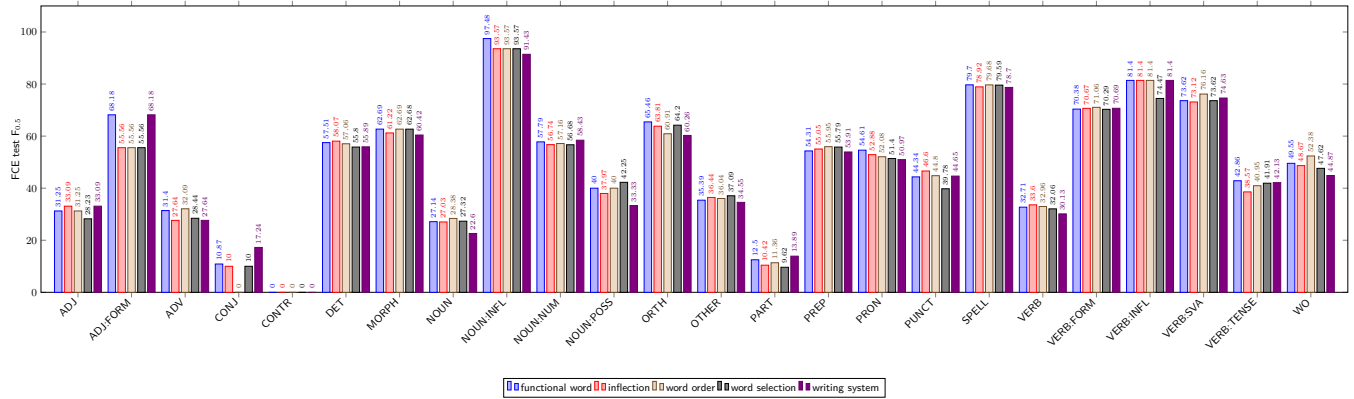
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (single)



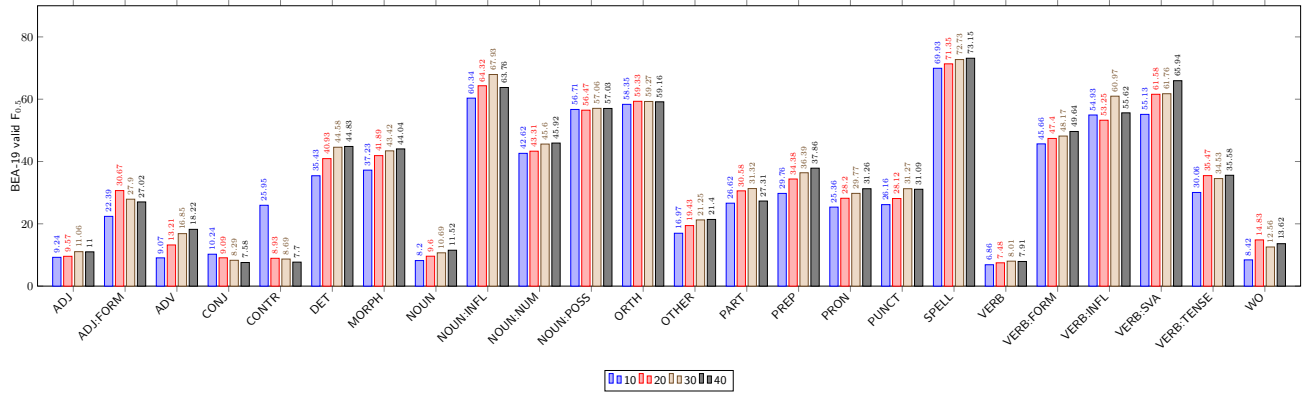
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (ensemble)



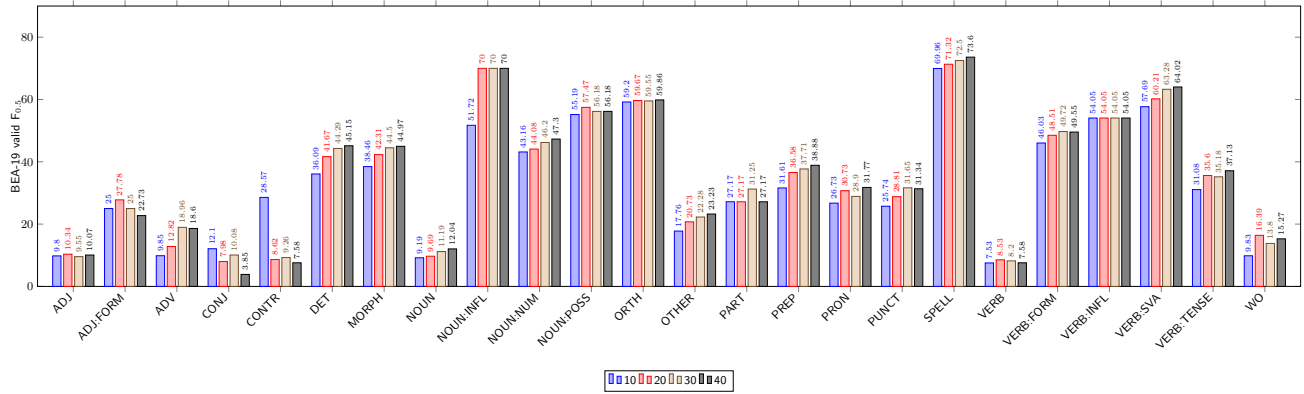
8M x 10 epochs pre-training (excluding 1 category) + fine-tuning (ensemble + rescore)



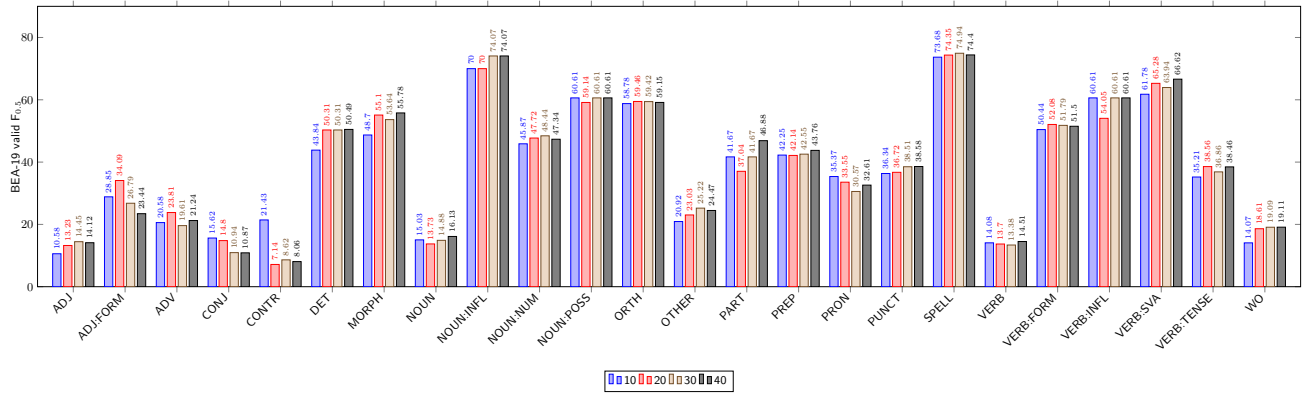
4M pre-training (single)



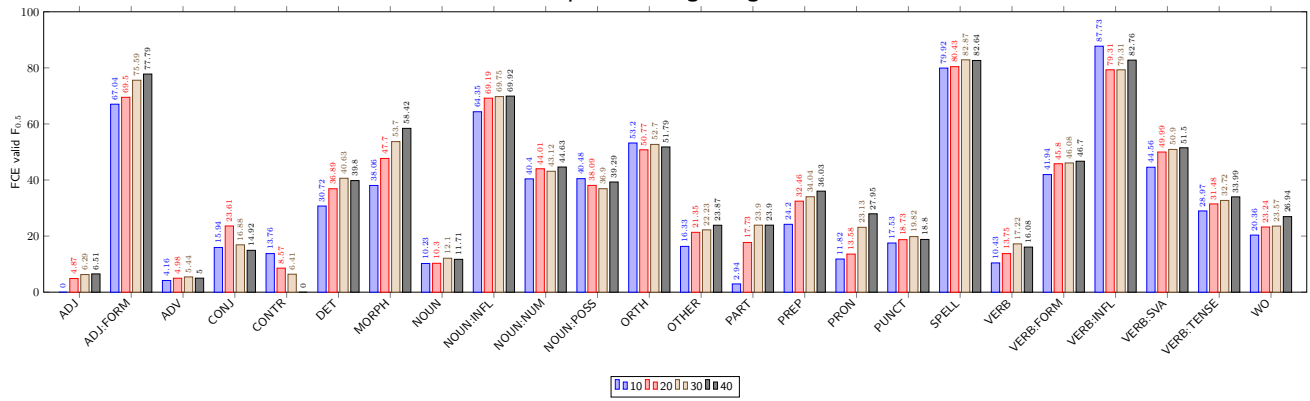
4M pre-training (ensemble)



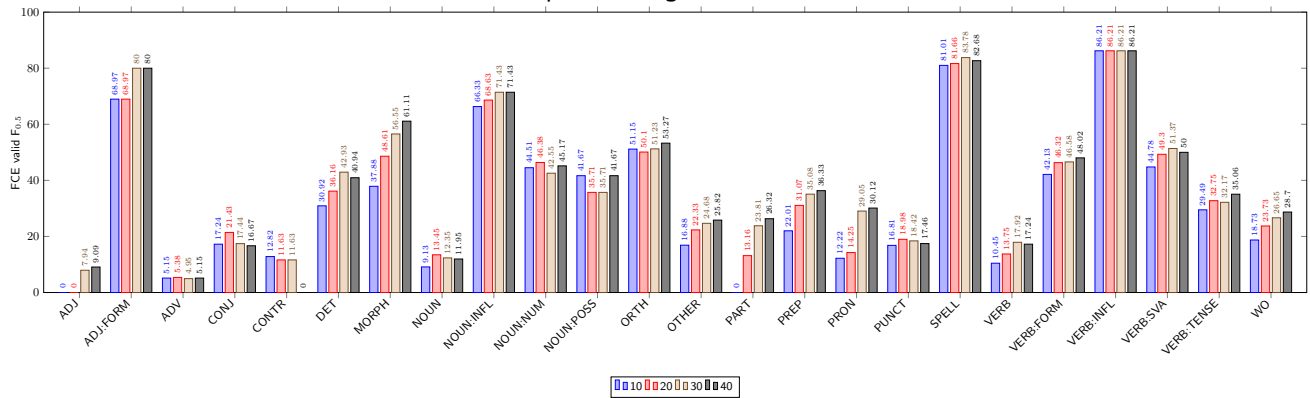
4M pre-training (ensemble + rescore)



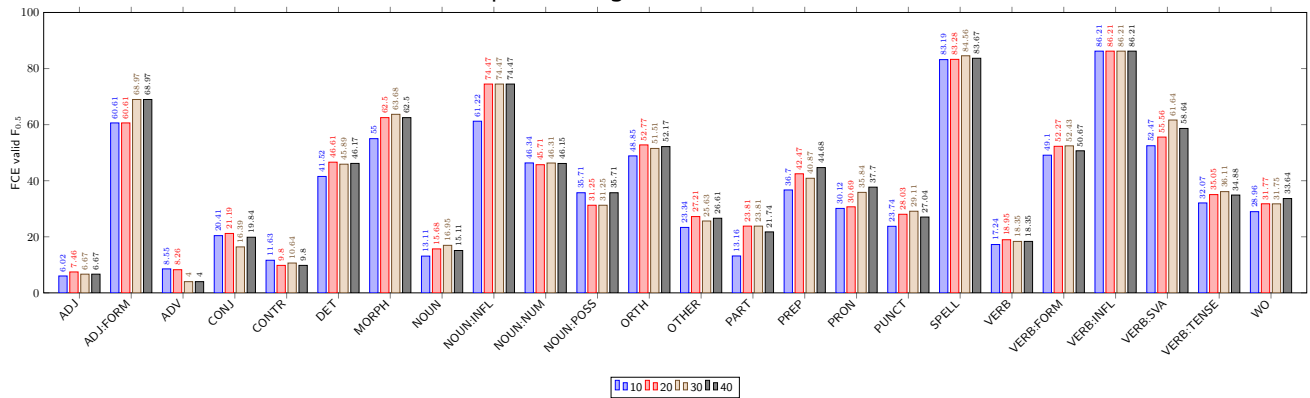
4M pre-training (single)



4M pre-training (ensemble)

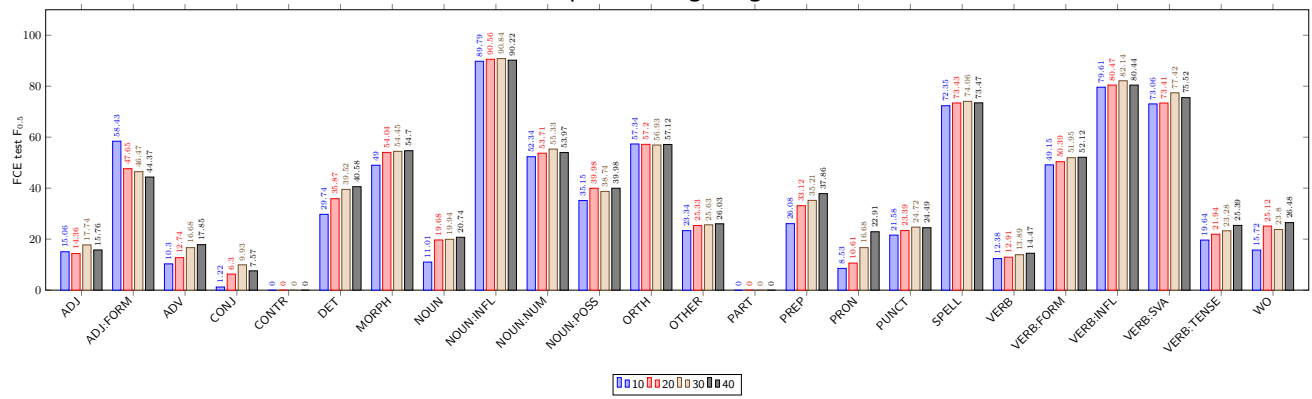


4M pre-training (ensemble + rescore)

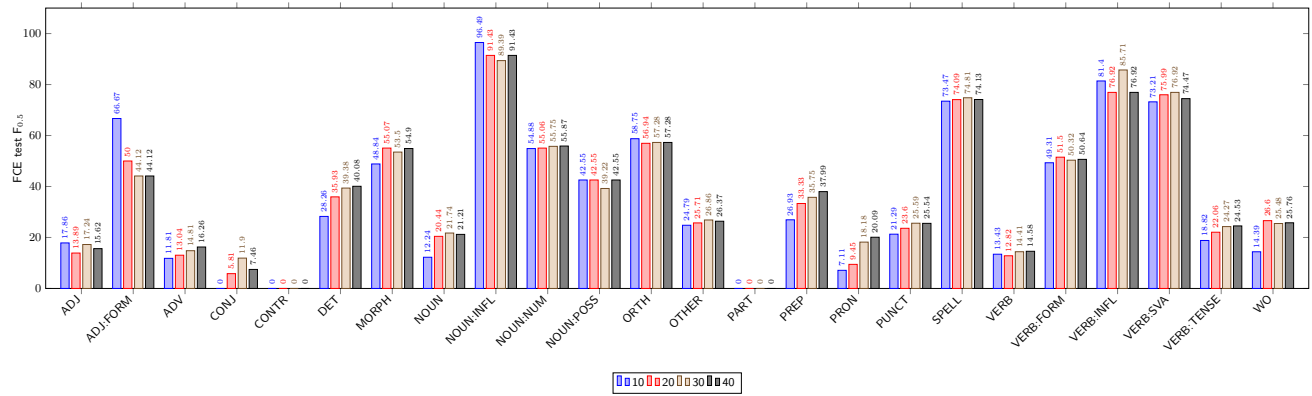




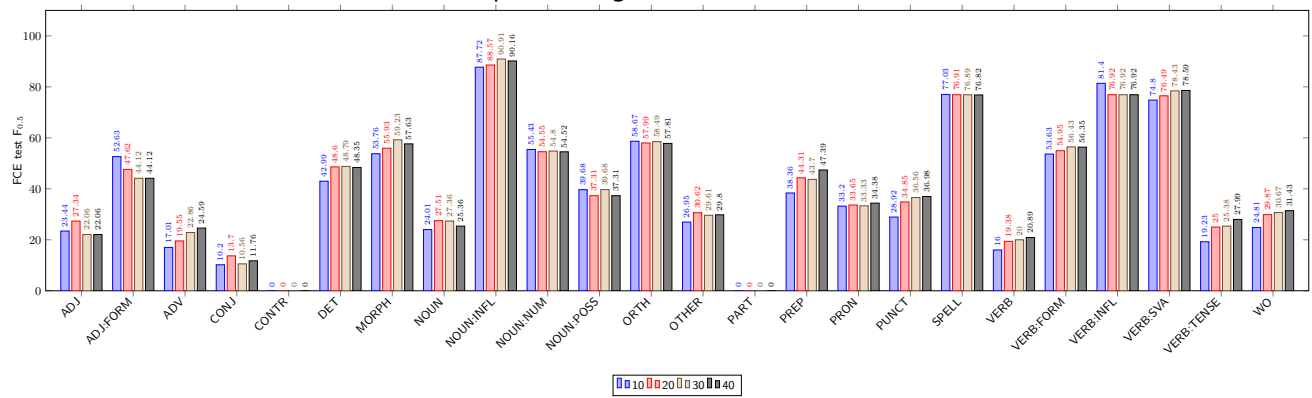
4M pre-training (single)



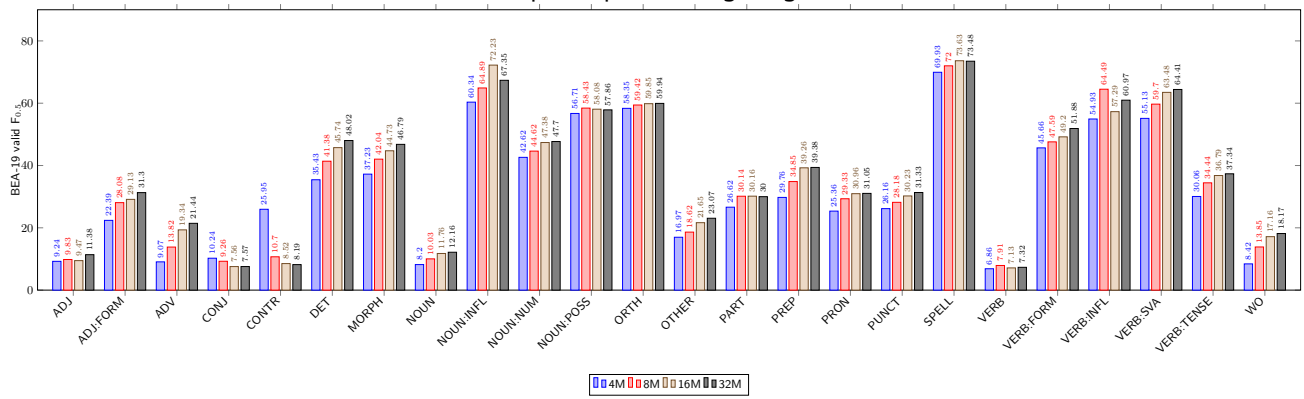
4M pre-training (ensemble)



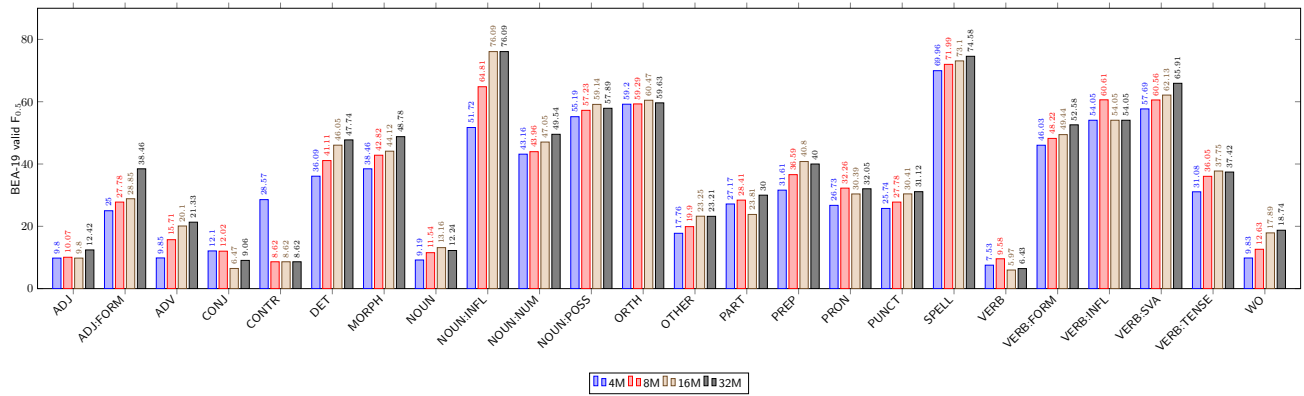
4M pre-training (ensemble + rescore)



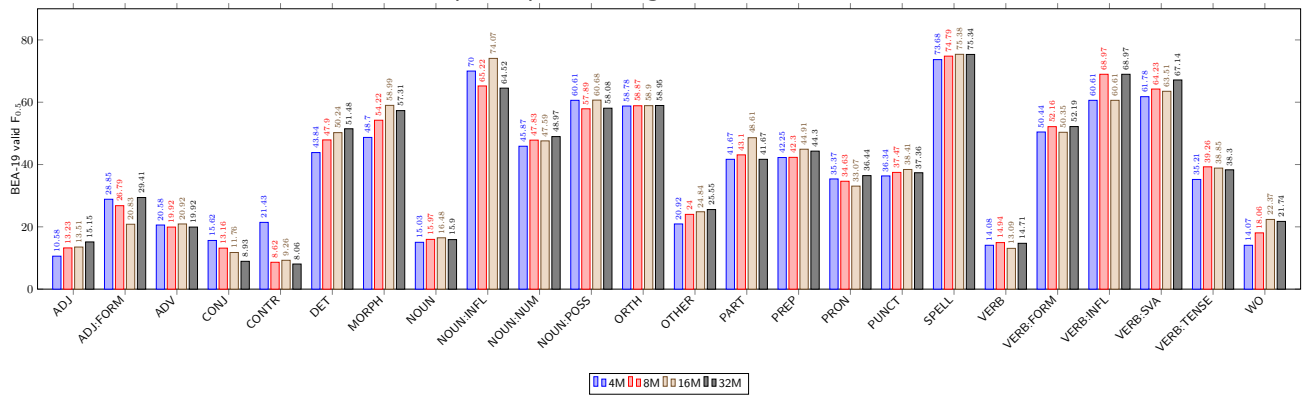
10 epochs pre-training (single)



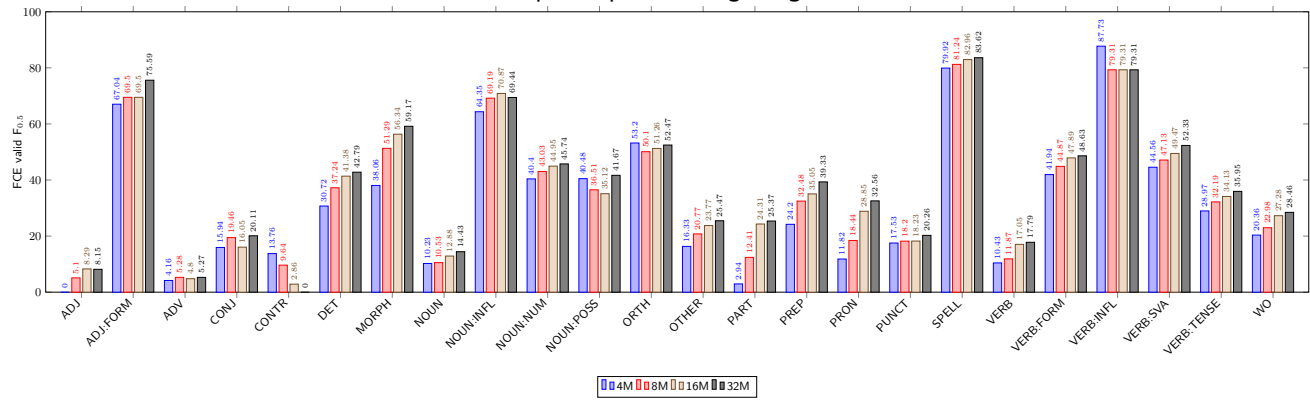
10 epochs pre-training (ensemble)



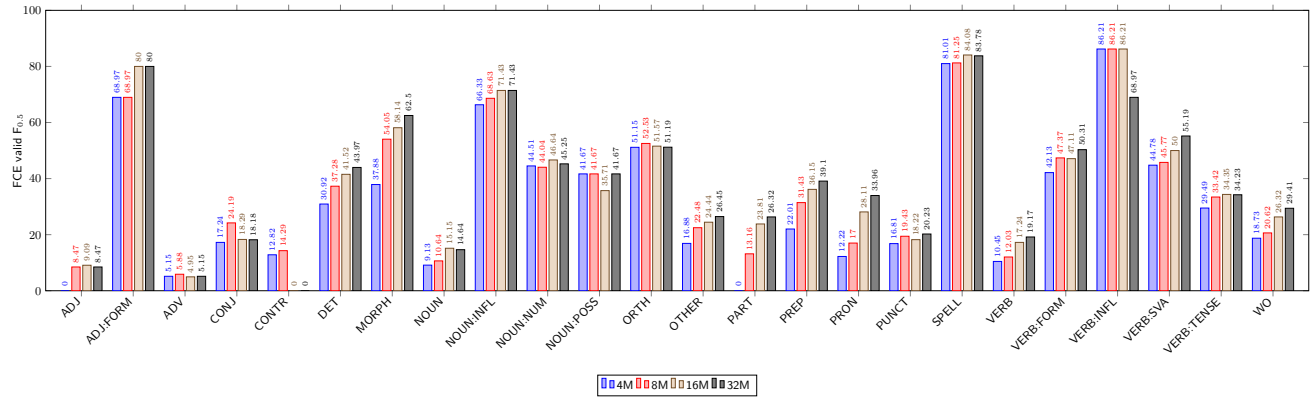
10 epochs pre-training (ensemble + rescore)



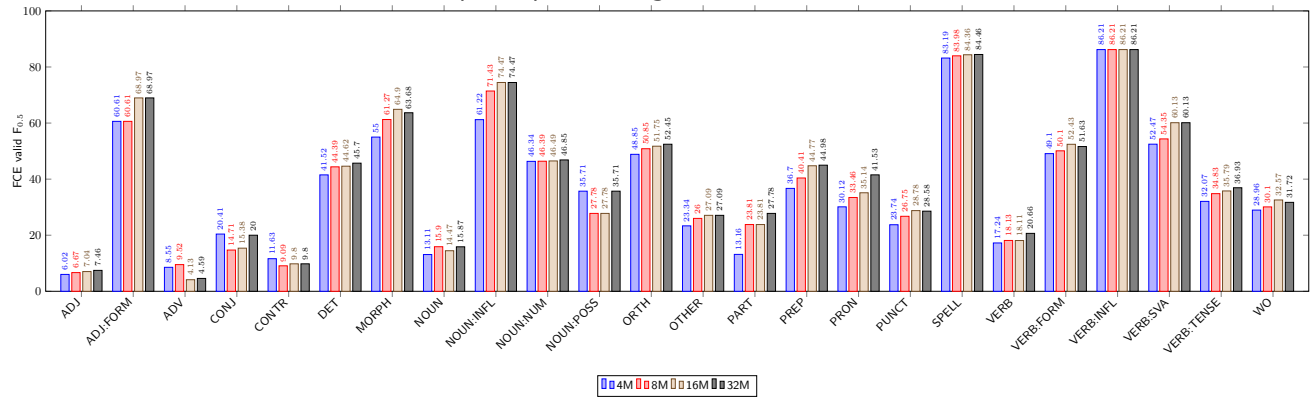
10 epochs pre-training (single)



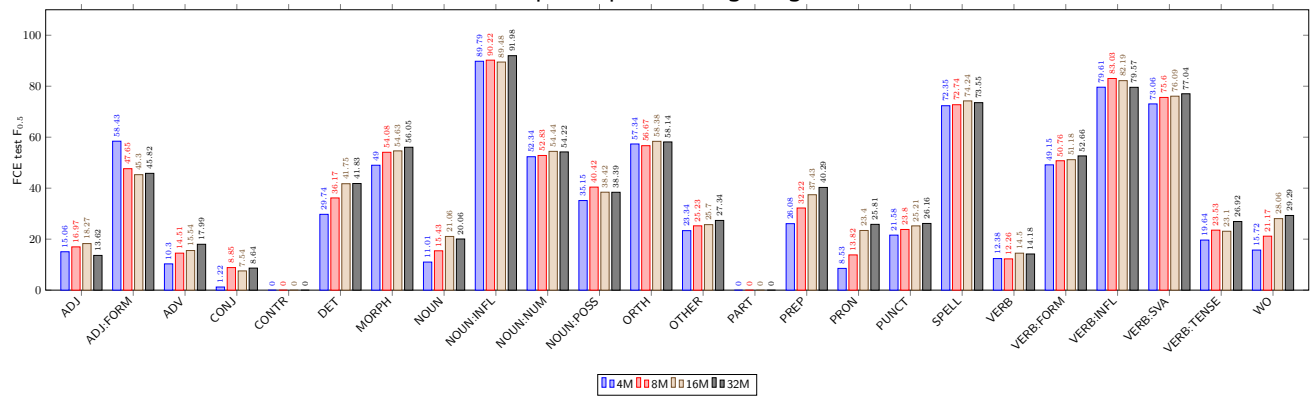
10 epochs pre-training (ensemble)



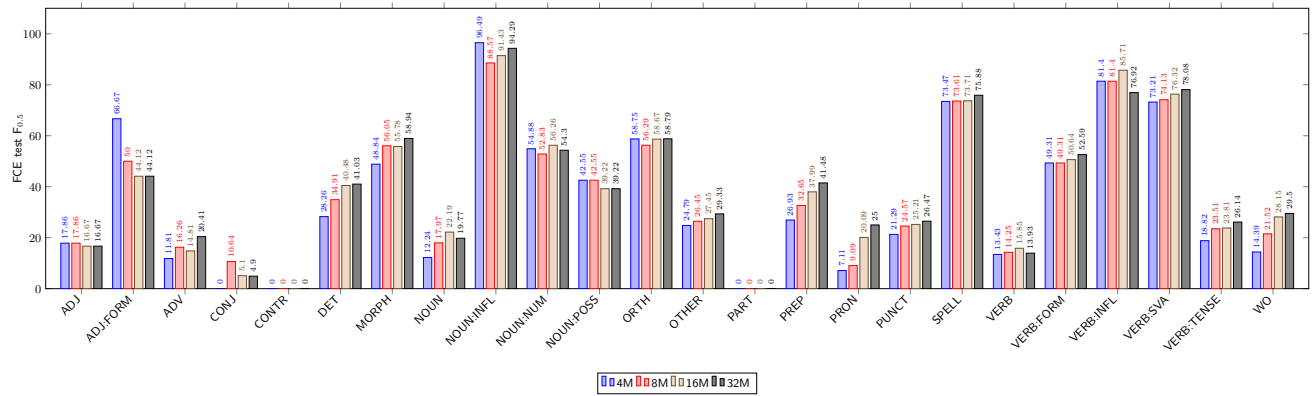
10 epochs pre-training (ensemble + rescore)



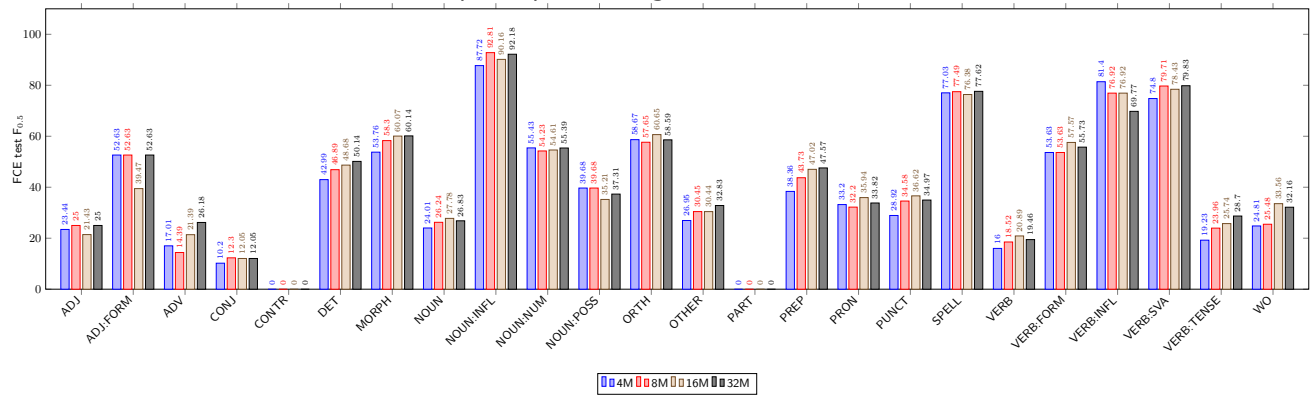
10 epochs pre-training (single)



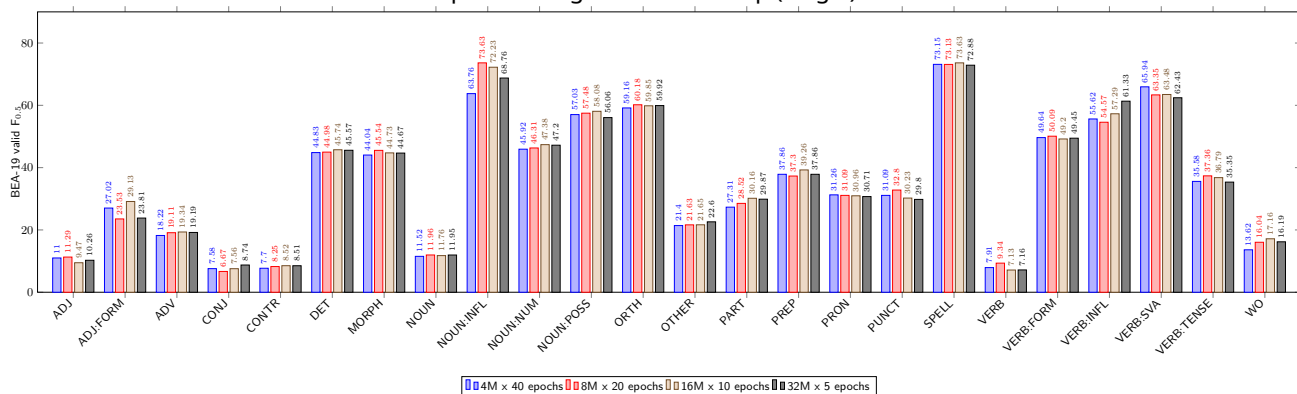
10 epochs pre-training (ensemble)



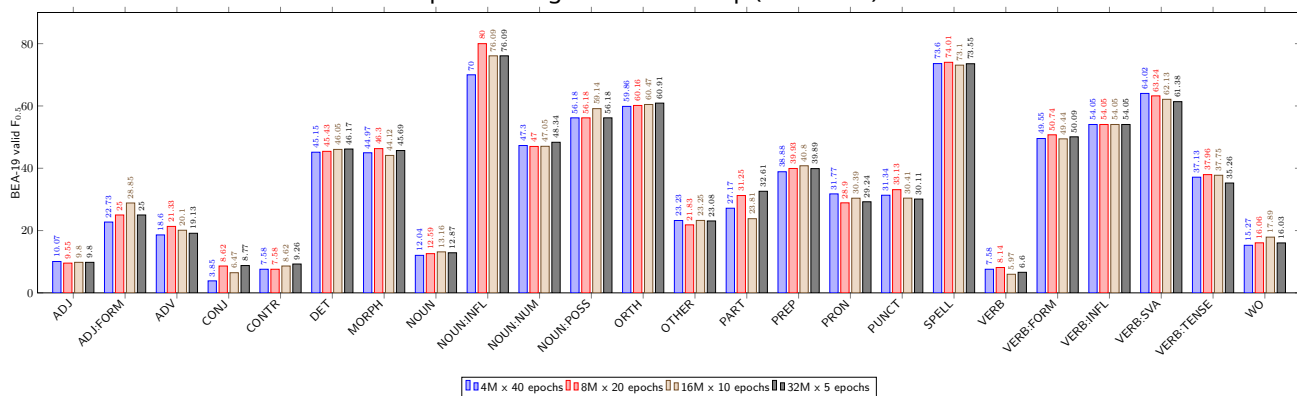
10 epochs pre-training (ensemble + rescore)



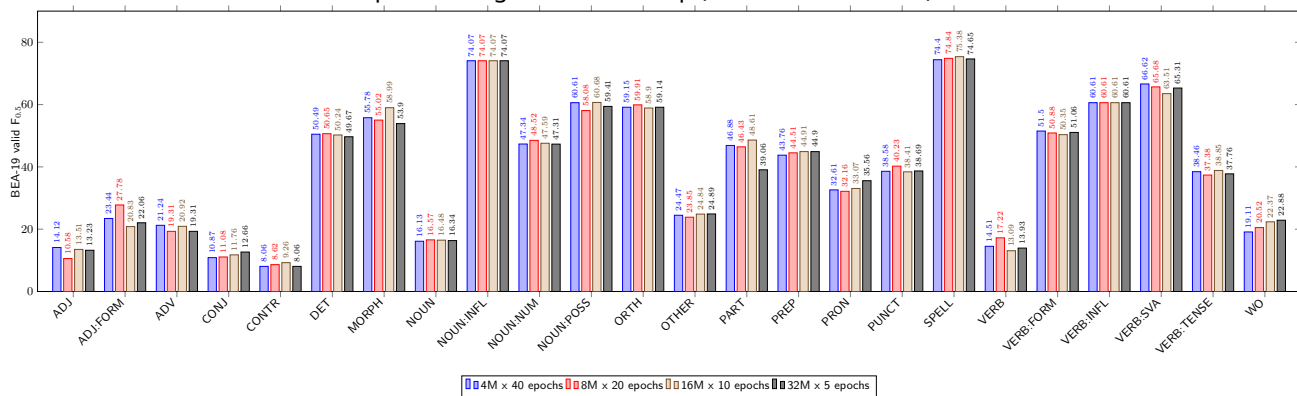
pre-training with fixed #step (single)



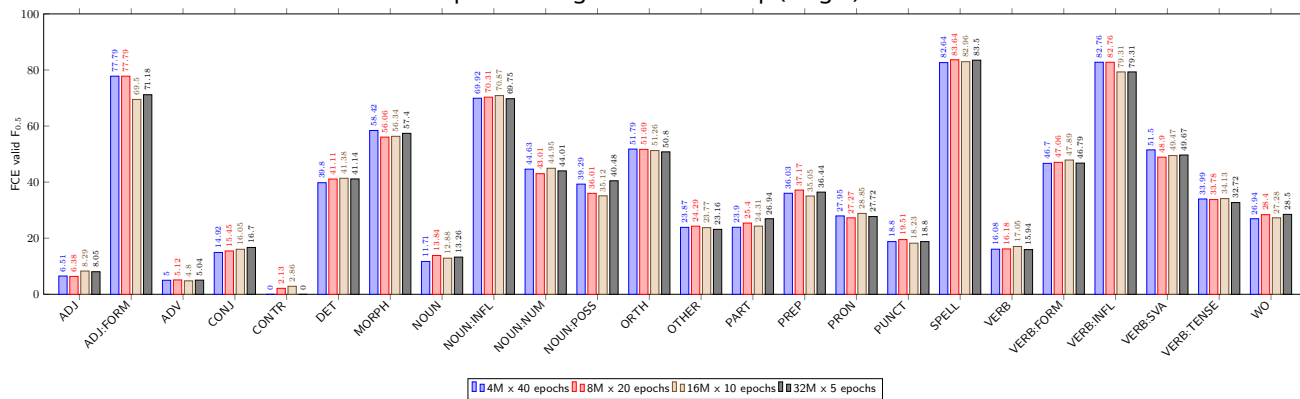
pre-training with fixed #step (ensemble)



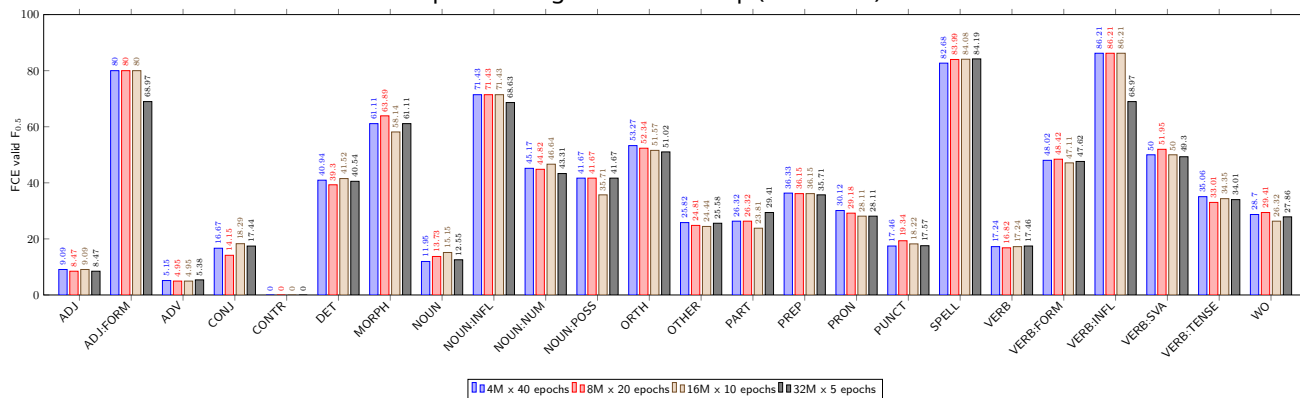
pre-training with fixed #step (ensemble + rescore)



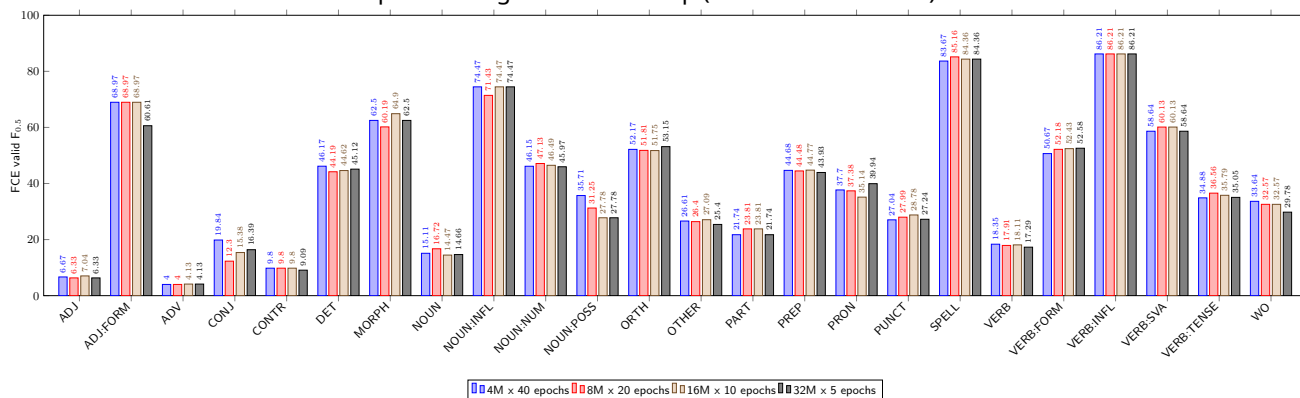
pre-training with fixed #step (single)



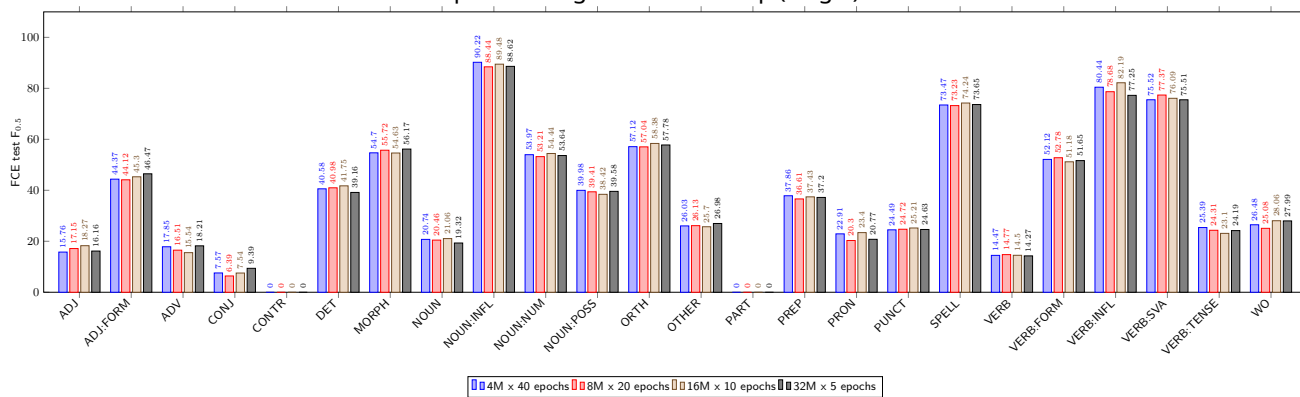
pre-training with fixed #step (ensemble)



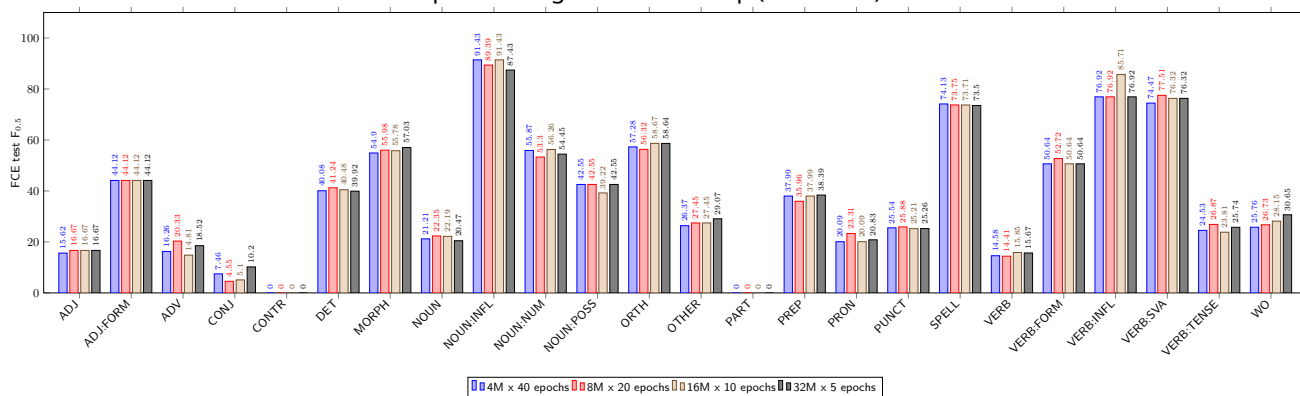
pre-training with fixed #step (ensemble + rescore)



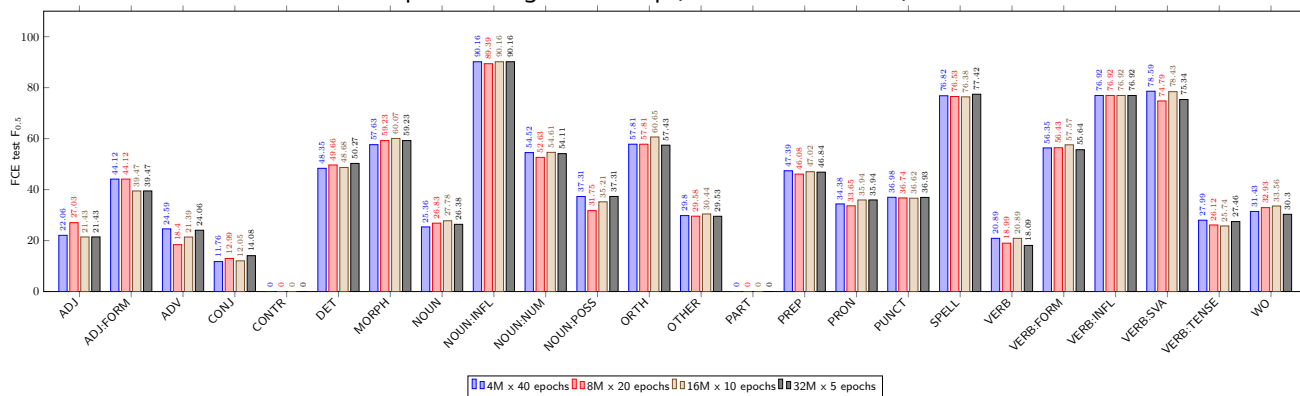
pre-training with fixed #step (single)



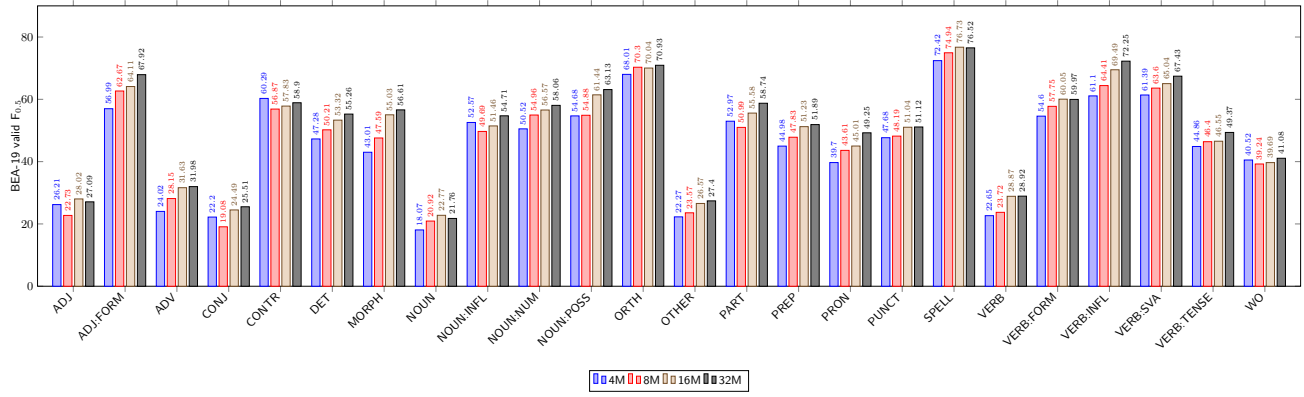
pre-training with fixed #step (ensemble)



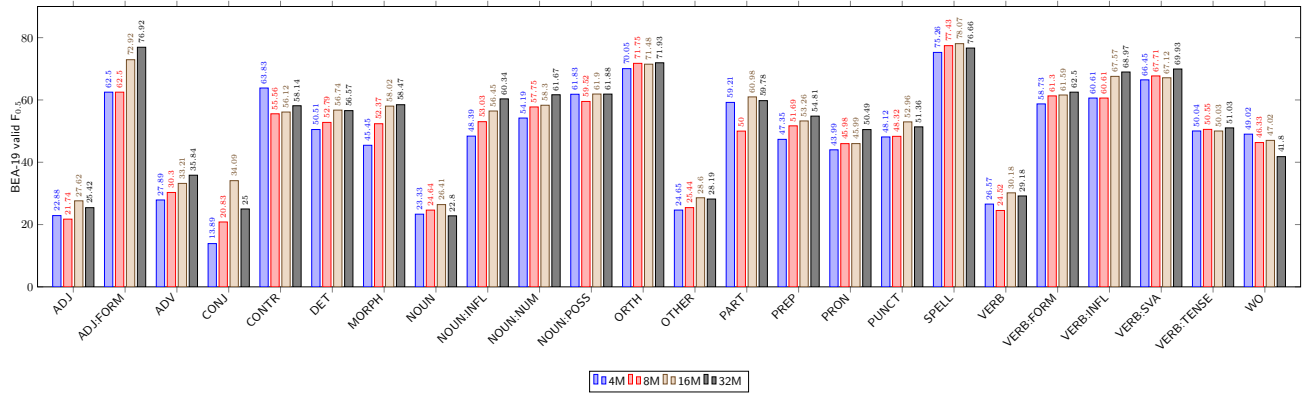
pre-training fixed #step (ensemble + rescore)



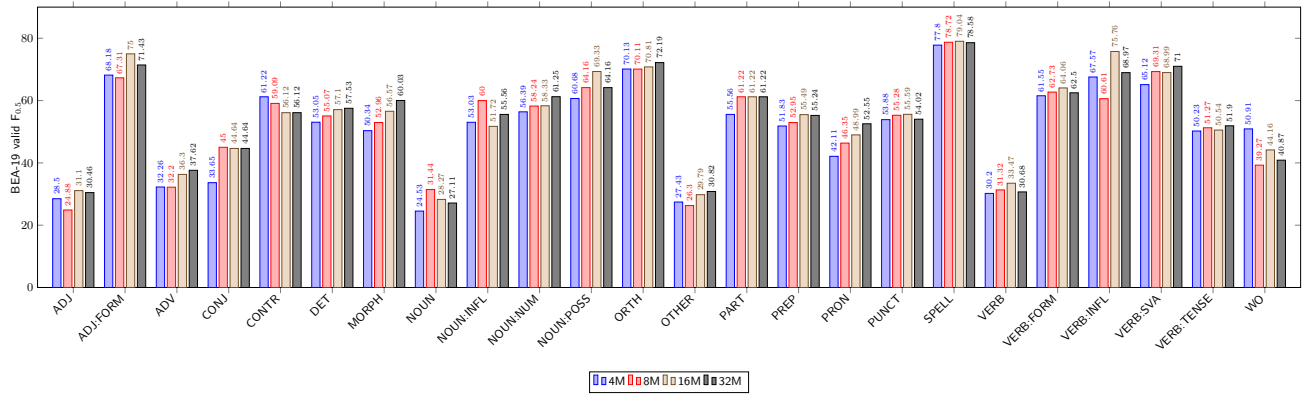
10 epochs pre-training + fine-tuning (single)



10 epochs pre-training + fine-tuning (ensemble)

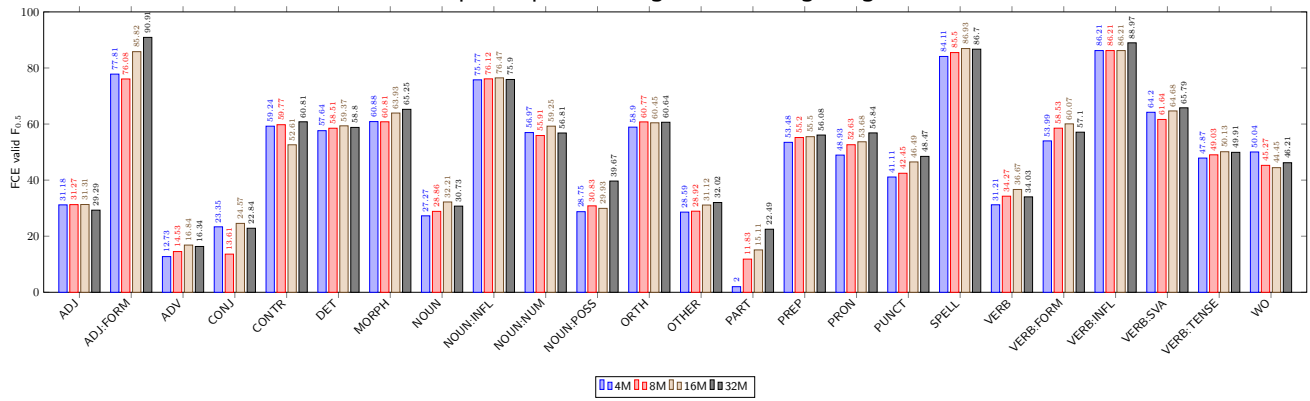


10 epochs pre-training + fine-tuning (ensemble + rescore)

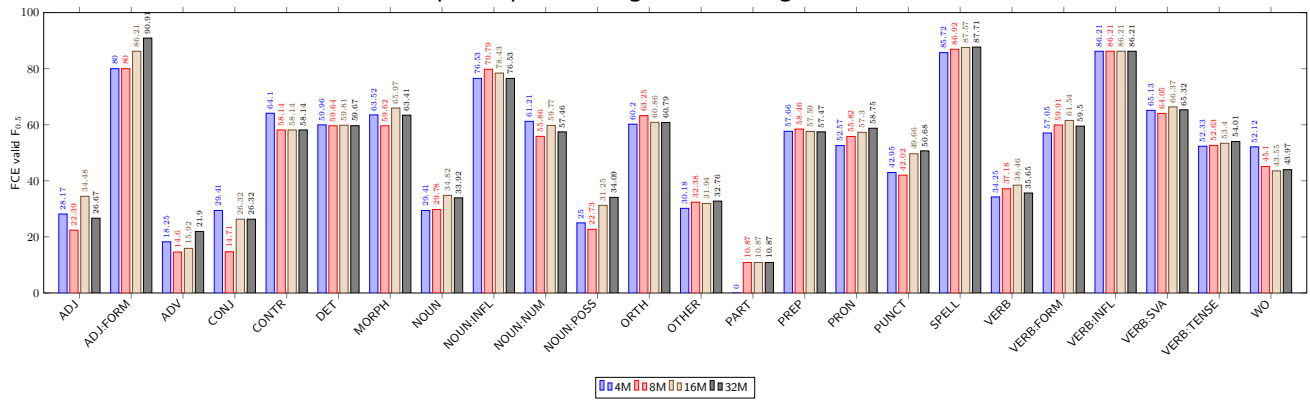




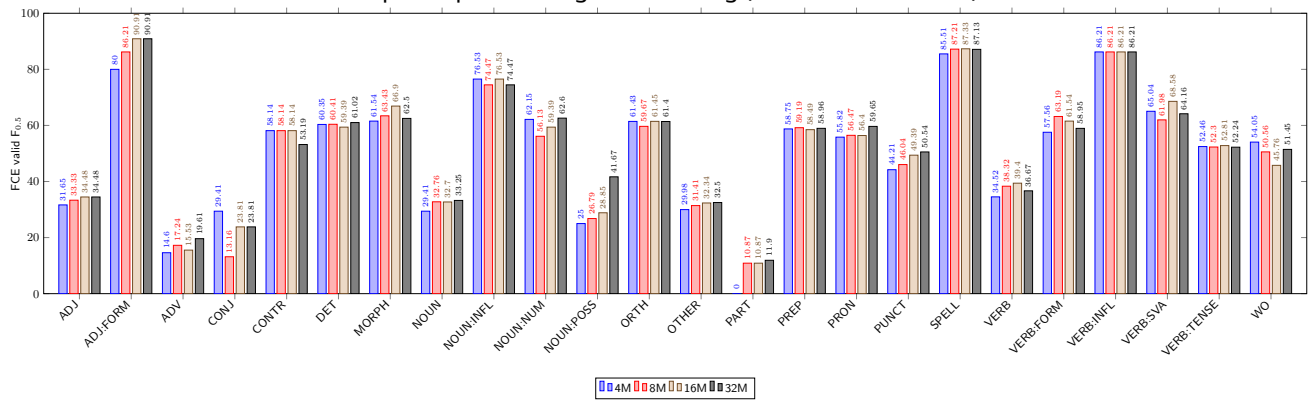
10 epochs pre-training + fine-tuning (single)



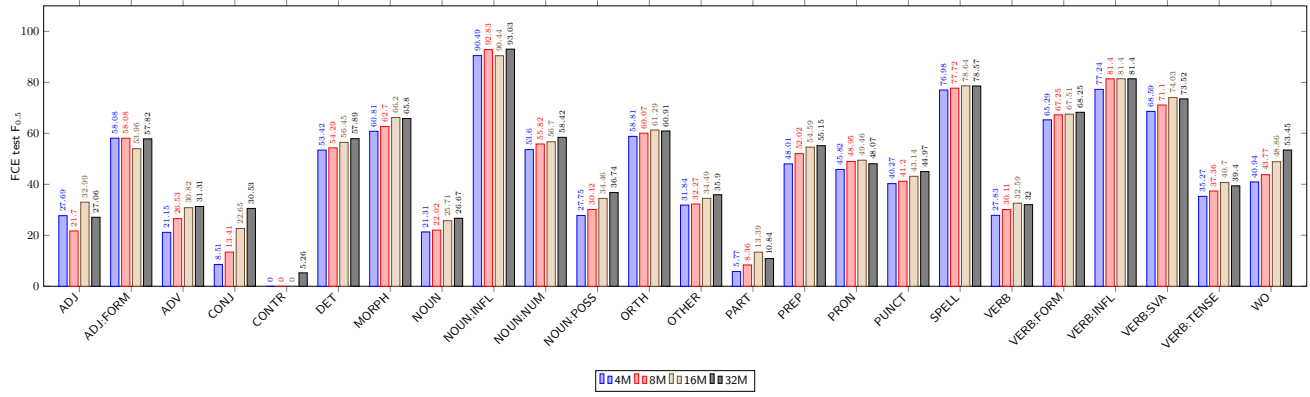
10 epochs pre-training + fine-tuning (ensemble)



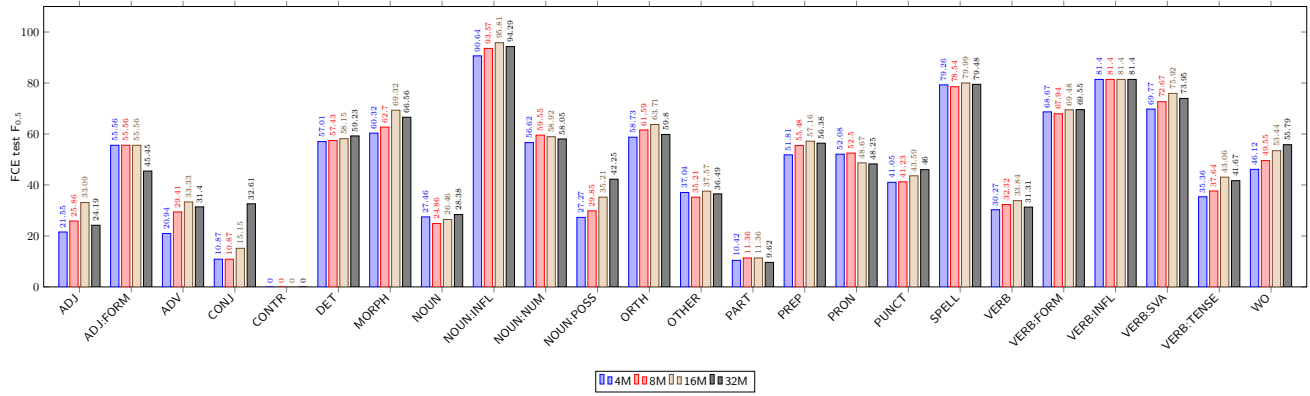
10 epochs pre-training + fine-tuning (ensemble + rescore)



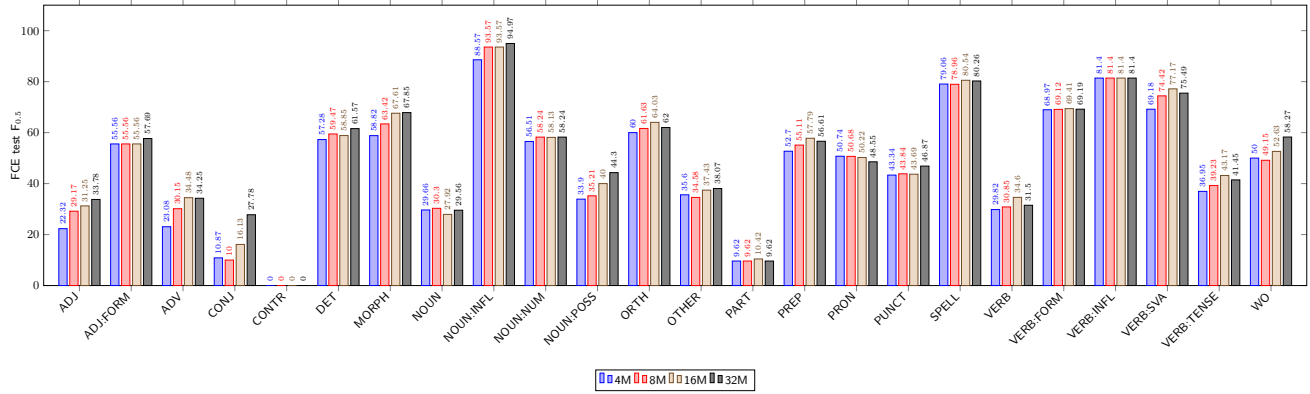
10 epochs pre-training + fine-tuning (single)



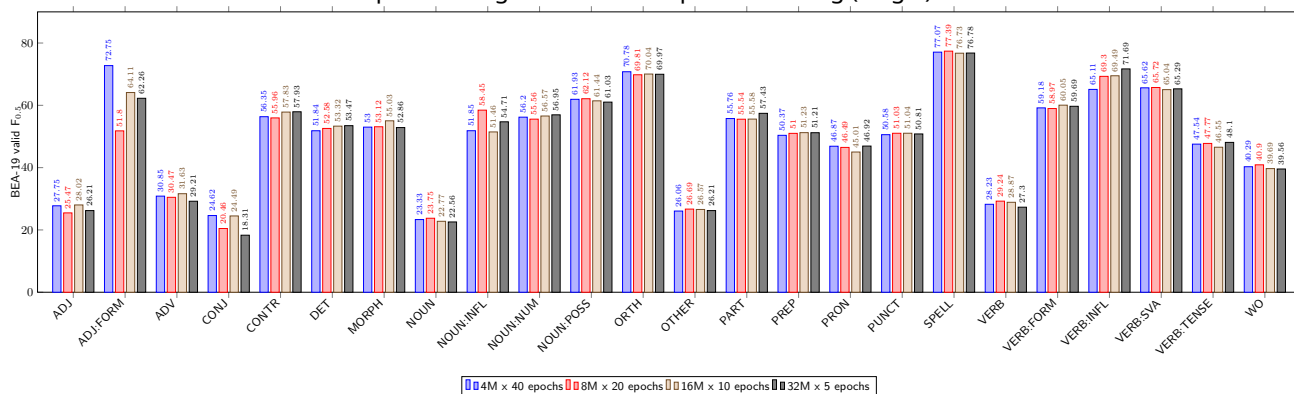
10 epochs pre-training + fine-tuning (ensemble)



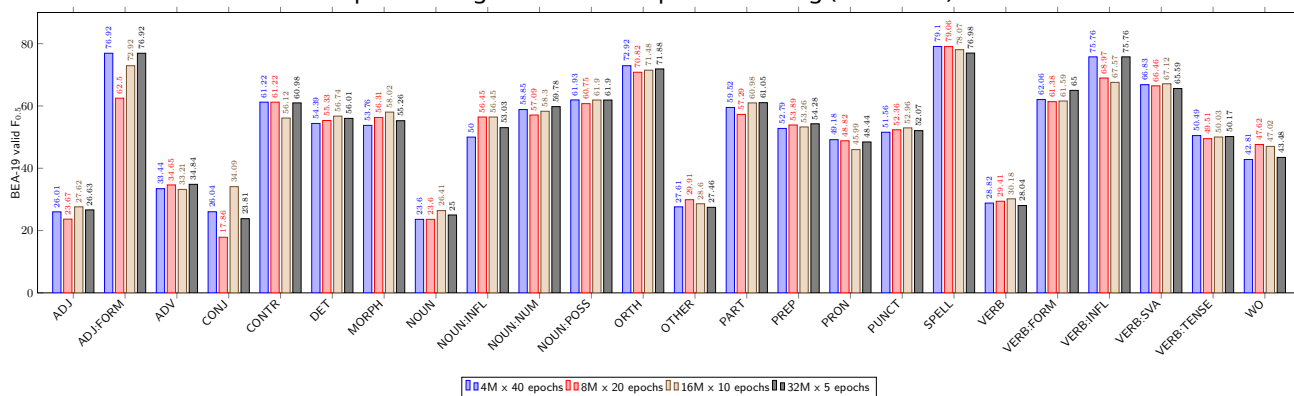
10 epochs pre-training + fine-tuning (ensemble + rescore)



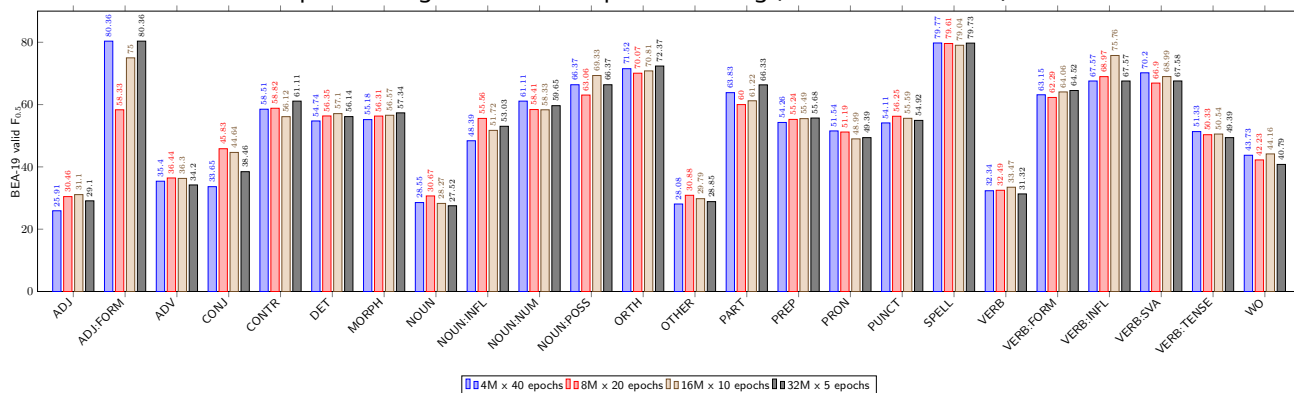
pre-training with fixed #step + fine-tuning (single)



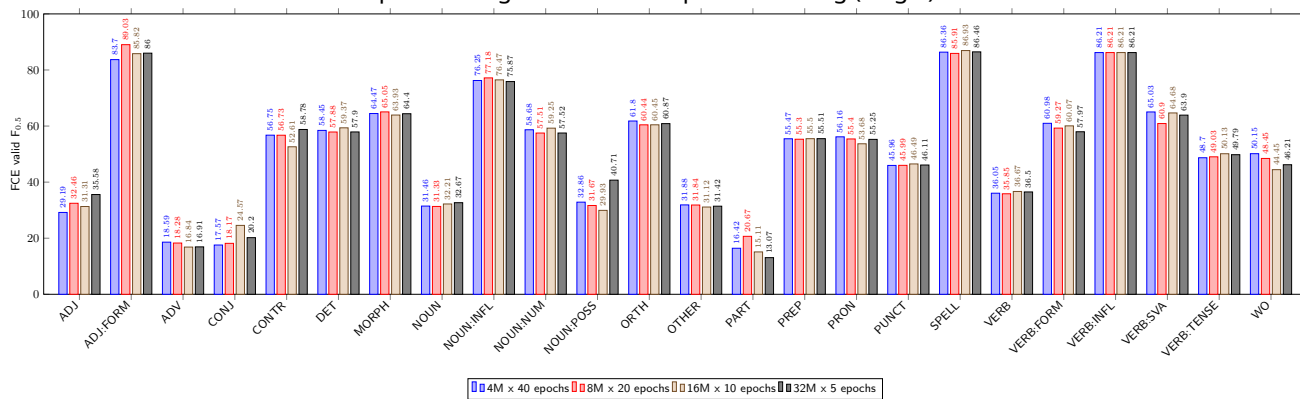
pre-training with fixed #step + fine-tuning (ensemble)



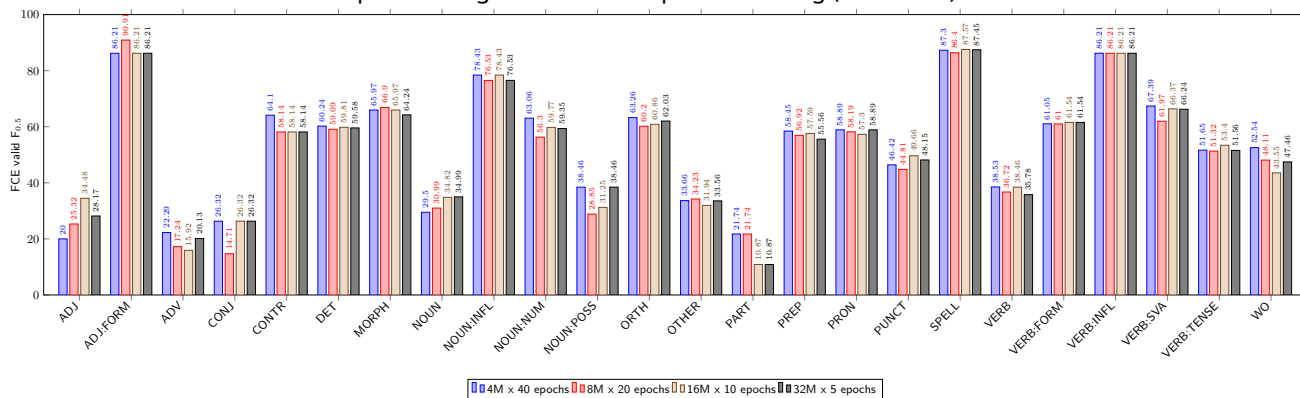
pre-training with fixed #step + fine-tuning (ensemble + rescore)



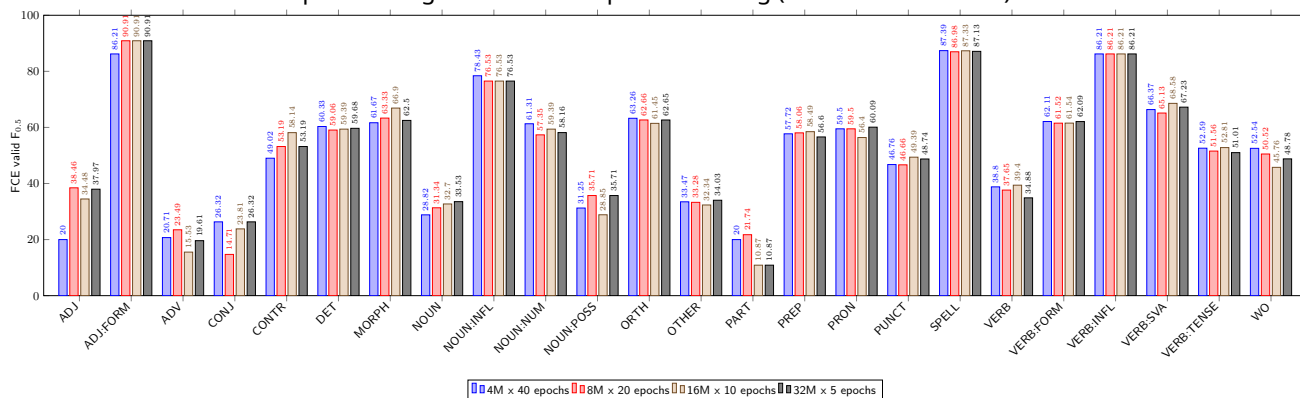
pre-training with fixed #step + fine-tuning (single)



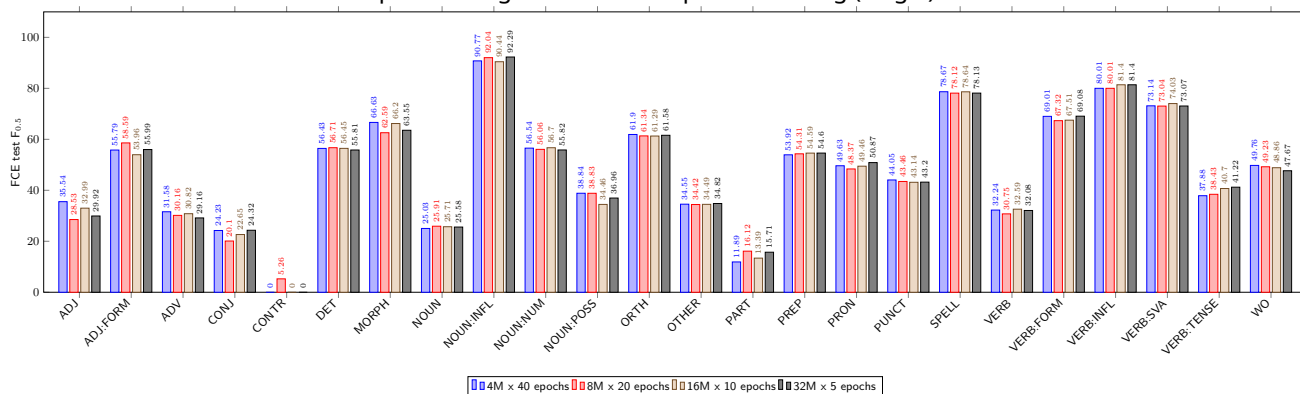
pre-training with fixed #step + fine-tuning (ensemble)



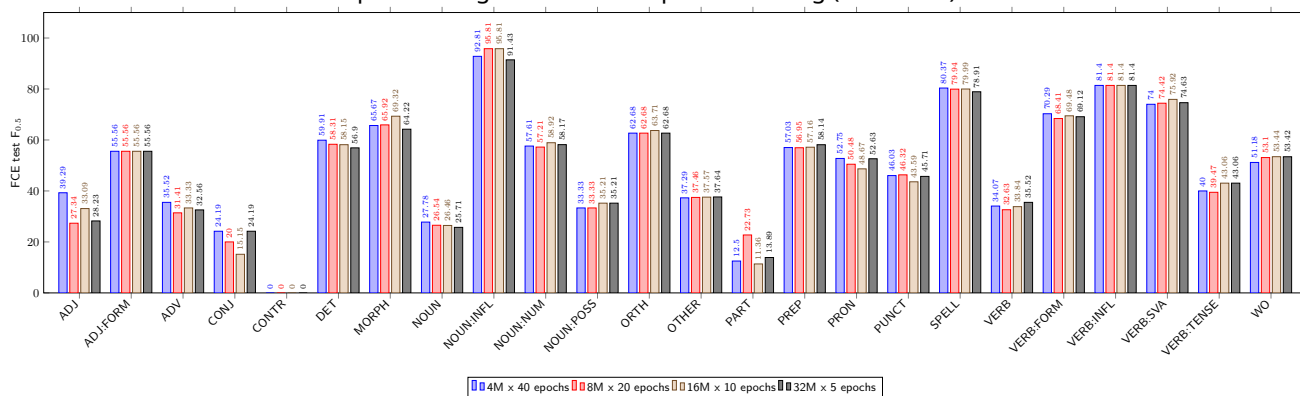
pre-training with fixed #step + fine-tuning (ensemble + rescore)



pre-training with fixed #step + fine-tuning (single)



pre-training with fixed #step + fine-tuning (ensemble)



pre-training fixed #step + fine-tuning (ensemble + rescore)

