

all_metrics

	B ³			MUC			CEAF _E		
	R	P	F1	R	P	F1	R	P	F1
HDDCRP (all mentions)	67.3	85.6	75.4	41.7	74.3	53.4	79.8	65.1	71.1
HDDCRP (valid mentions)	83.82	96.84	89.86	60.4	88.96	71.95	91.72	78.36	84.52
HDDCRP (valid mentions)	83.03	96.85	89.41	59.41	88.96	71.24	91.48	78.23	84.34
OURS (valid mentions)									

all_metrics

CoNLL	
F1	gold truth
66.8	corpus
82.11	hddcrp's
81.66	corpus
	hddcrp's
	corpus

summary results

		gold truth		
		hddcrp's valid mentions	all corpus mentions	
model	samelemma	-	60.2	
	HDDCRP	82.1	66.8	
	Choubey '17	-	68.9	
	nothing (all singletons)	49.1	48.6	LOWERBOUND
	ours (min-merge max)	83.8	83.3	UPPERBOUND
	ours (min-merge)	74.0	74.8	ACTUAL
	ours (min-avgmerge max)	83.8	83.5	
	ours (min-avgmerge)	75.3	?	
WITHIN-DOC PERFORMANCE				

74.0 = /home/ctanner/gpuVRUN_nl2_ne3_ws2_neg10_bs128_sf_e400.out:conll - f1: 0.740006733379

74.8 = /home/ctanner/gpuCorpusRUN_nl1_ne10_ws2_neg10_bs64_sf_e400.out:conll - f1: 0.748030955473 (sp=0.55)

conll WD scores

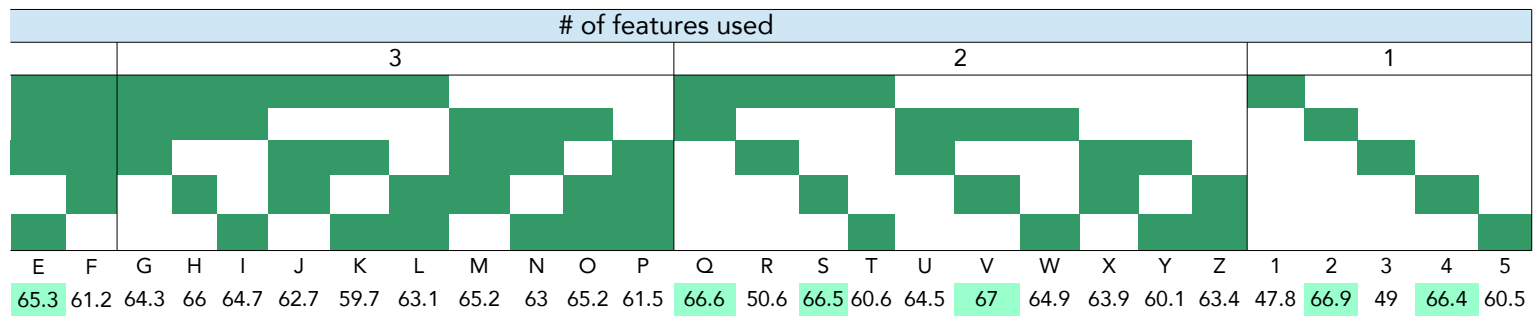
		token embeddings		max pooling		
feature	glove corpus	sum	avg			
word	400		60.60			
	6B		62.52			
	840B		62.74			
lemma	400	65.62	65.56			
	6B	66.57	66.39			
	840B	~65	~65			
dependency parent	word	6B	-		46.16	-
		840B	-		46.17	-
	lemma	6B	46.55		46.52	31.35
		840B	46.55		46.56	31.35
dependency parent+children	lemma	6B	48.64	31.35	48.88	48.24
		840B	49.26	31.35	49.37	31.35
character embeddings	word	concat	66.50	max pooling	66.57	
		sum	58.68		59.69	31.35
	lemma	concat	66.55		66.65	
		sum	65.07		63.52	31.35
semantic space	lemma	50				
		300				
		800				
	token location	%				
		1-hot				
bag of words	word	6B				
		840B				
	lemma	6B				
		840B				

avg pooling

		5	4		
POS					
lemma					
dependency lemma					
char emb					
semantic space					
		A	B	C	D
		63.7	64.2	64.9	62.4

conll WD scores

conll WD scores

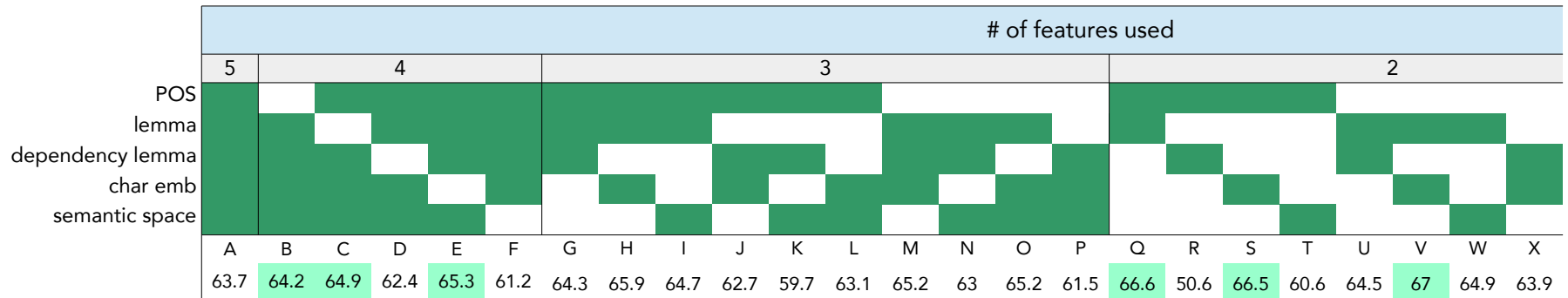


		AGG	
	global	65.8	
	dir	65.9	
	doc	65.8	
			FFNN
		AGG	A B
6629	23-fold	66.29	
6633	23-fold	66.33	
7	7 – 25	64.82	
11	11 – 25	66.16	
15	15 – 25	66.18	
19	19 – 25	66.37	

conll WD scores

23 23 – 25 66.99

FEATURES' EFFECTIVENESS:



HOW TOKENS AFFECT PERFORMANCE:

RAW COUNTS	single token mentions (both are)									
	contains lemma substring					does not contain lemma substring				
			pred					pred		
			T	F	gold truth			T	F	gold truth
			690	5				69	268	
			4	16				59	9614	
% s	single token mentions (both are)									
	contains lemma substring					does not contain lemma substring				
			pred					pred		
			T	F	gold truth			T	F	gold truth
			0.97	0.01				0.01	0.03	
			0.01	0.02				0.01	0.96	
RAW COUNTS	multi-token mentions (at least one is)									
	contains lemma substring					does not contain lemma substring				
			pred					pred		
			T	F	gold truth			T	F	gold truth
			55	11				2	11	
			159	176				14	1864	
% s	multi-token mentions (at least one is)									
	contains lemma substring					does not contain lemma substring				
			pred					pred		
			T	F	gold truth			T	F	gold truth
			0.14	0.03				.001	.006	
			0.40	0.44				.007	.986	

conll WD scores_2

		1					
Y	Z	1	2	3	4	5	
60.1	63.4	47.8	66.9	49	66.4	60.5	

paper

training
dev
test

		MUC	CEAF	BCUB	CoNLL F1
	agg				66.8
HDDCRP	HDDCRP				
	agg				69?
CCNN	deep clustering				
test data: HDDCRP's parsed, predicted mentions (not gold truth mentions)					
	Choubey 17				68
	agg				70?
CCNN	deep clustering				72?
test data: Choubey's mentions (not gold truth mentions)					