Metric		Dev	Test
Tokens		291,746	216,473
Sentences		15,084	11,623
Span	Precision	93.42	93.04
	Recall	94.21	94.34
	F1	93.81	93.69
Frame	Precision	93.47	93.20
	Recall	94.16	94.08
	F1	93.81	93.64
Type	Precision	85.56	85.67
	Recall	86.20	86.49
	F1	85.88	86.08
Role	Precision	70.21	69.59
	Recall	69.11	69.20
	F1	69.65	69.39
Label	Precision	96.51	95.02
	Recall	94.97	90.70
	F1	95.73	92.81
Slot	Precision	80.00	79.81
	Recall	79.90	80.10
	F1	79.95	79.96
Combined	Precision	87.46	87.20
	Recall	87.79	87.91
	F1	87.63	87.55