

Task	Technique	Metric	Skip-Gram	Embed-Align	Bayesian Skip-Gram
MR	Averaging	Accuracy	68.75	64.73	64.71
MR	SIF	Accuracy	69.84	66.55	69.45
MR	GRAN	Accuracy	67.83	61.92	63.78
CR	Averaging	Accuracy	74.91	70.41	68.08
CR	SIF	Accuracy	74.09	70.07	68.48
CR	GRAN	Accuracy	73.51	67.26	67.01
MPQA	Averaging	Accuracy	86.26	82.93	82.45
MPQA	SIF	Accuracy	86.19	83.51	83.77
MPQA	GRAN	Accuracy	84.62	79.65	81.47
SUBJ	Averaging	Accuracy	85.53	79.82	78.54
SUBJ	SIF	Accuracy	85.83	80.00	81.34
SUBJ	GRAN	Accuracy	83.18	75.21	77.78
SST	Averaging	Accuracy	71.66	67.71	63.87
SST	SIF	Accuracy	70.46	65.18	64.91
SST	GRAN	Accuracy	70.18	65.62	64.69
TREC	Averaging	Accuracy	69.40	49.60	58.00
TREC	SIF	Accuracy	62.60	45.20	46.40
TREC	GRAN	Accuracy	67.40	46.00	58.80
MRPC	Averaging	Accuracy / F1	70.78 / 80.06	71.54 / 80.78	70.67 / 80.22
MRPC	SIF	Accuracy / F1	71.48 / 80.84	69.80 / 79.98	69.86 / 79.78
MRPC	GRAN	Accuracy / F1	71.11 / 80.76	71.48 / 81.21	71.59 / 80.59
SICK	Averaging	Accuracy	75.20	74.00	70.41
SICK	SIF	Accuracy	71.85	73.74	64.85
SICK	GRAN	Accuracy	73.23	66.23	72.80
STS	Averaging	Spearman / Pearson	0.5326 / 0.5137	0.5788 / 0.5810	0.4623 / 0.4532
STS	SIF	Spearman / Pearson	0.5062 / 0.4867	0.5784 / 0.5809	0.4350 / 0.4380
STS	GRAN	Spearman / Pearson	0.3987 / 0.3643	0.5015 / 0.4745	0.3907 / 0.3795

Table 1: All results for the SentEval tasks, per model, per sentence representation technique. For tasks where two metrics were considered, both are mentioned.