

Run ID:	Stanford1
3-way accuracy:	0.553
2-way accuracy:	0.614
Average precision:	0.4416

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	333	143	24	500
	Unknown	136	208	6	350
	Contradiction	83	55	12	150
Total		552	406	42	1000

Table 1: Confusion matrix for the 3-way task for Stanford1

Run ID:	Stanford2
3-way accuracy:	0.508
2-way accuracy:	0.57
Average precision:	0.5427

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	310	175	15	500
	Unknown	155	190	5	350
	Contradiction	85	57	8	150
Total		550	422	28	1000

Table 2: Confusion matrix for the 3-way task for Stanford2

Run ID:	Stanford3
3-way accuracy:	0.508
2-way accuracy:	0.57
Average precision:	0.5427

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	310	175	15	500
	Unknown	155	190	5	350
	Contradiction	85	57	8	150
Total		550	422	28	1000

Table 3: Confusion matrix for the 3-way task for Stanford3