

Run ID:	boeing1
3-way accuracy:	0.377
2-way accuracy:	0.515

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	35	463	2	500
	Unknown	13	337	0	350
	Contradiction	7	138	5	150
Total		55	938	7	1000

Table 1: Confusion matrix for the 3-way task for boeing1

Run ID:	boeing2
3-way accuracy:	0.459
2-way accuracy:	0.565

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	176	316	8	500
	Unknown	77	271	2	350
	Contradiction	34	104	12	150
Total		287	691	22	1000

Table 2: Confusion matrix for the 3-way task for boeing2

Run ID:	boeing3
3-way accuracy:	0.481
2-way accuracy:	0.547

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	284	203	13	500
	Unknown	160	182	8	350
	Contradiction	77	58	15	150
Total		521	443	36	1000

Table 3: Confusion matrix for the 3-way task for boeing3