A Unique utterance counts, intent counts and labels counts for pseudo Golden and Rare data.

Domain	Utterance Group	Unique Utterance Count	Unique Intent Count	Unique Label Count
alarm	**pseudo Goldens**	202	6	3
alarm	**Rare**	7006	6	7
reminder	**pseudo Goldens**	513	3	7
reminder	**Rare**	5580	3	7
weather	**pseudo Goldens**	228	3	5
weather	**Rare**	10929	3	8

B Summary of the results on Language Bootstrapping Experiments averaged over three runs for Generator with sentence-level reward.

Metric	Domain	Baseline	Generator with sentence-level reward with Monte Carlo rollout					
			TopX		Unique		All	
		Performance	Performance	% change	Performance	% change	Performance	% change
Domain accuracy		96.83	97.71	+0.91	95.5	-1.37	95.17	-1.72
Intent accuracy	**alarm**	73.33	68.53	-6.54	79.98	+9.07	80.46	+9.73
	reminder	72.93	81.72	+12.06	79.63	+9.19	79.39	+8.85
	weather	97.5	97.58	+0.08	97.72	+0.23	98.65	+1.18
Overall intent accuracy		84	84.89	+1.05	87.96	+4.71	88.4	+5.24
Slot F1	**alarm**	92.63	91.8	-0.89	92.54	-0.09	93.11	+0.52
	reminder	83.49	84.79	+1.55	84.72	+1.47	83.61	+0.14
	weather	83.63	82.55	-1.29	83.2	-0.52	83.63	+0.00
Overall slot F1		86.19	85.88	-0.36	86.08	-0.12	86.07	-0.14
Overall frame accuracy		46.44	45.64	-1.73	46.53	+0.19	46.51	+0.16
Metric	**Domain**	**Baseline**	**Generator with sentence-level reward without rollouts**					
			TopX		**Unique**		**All**	
		Performance	**Performance**	**% change**	**Performance**	**% change**	**Performance**	**% change**
Domain accuracy		96.83	94.62	-2.28	97.6	+0.80	96.58	-0.26
Intent accuracy	**alarm**	73.33	73.89	+0.77	79.77	+8.78	83.93	+14.45
	reminder	72.93	74.71	+2.44	87.53	+20.02	85.62	+17.40
	weather	97.5	97.64	+0.14	97.8	+0.31	97.76	+0.27
Overall intent accuracy		84	84.21	+0.25	89.86	+6.98	90.56	+7.81
Slot F1	**alarm**	92.63	91.48	-1.24	89.92	-2.93	91.53	-1.19
	reminder	83.49	80.64	-3.41	86.09	+3.11	84.5	+1.21
	weather	83.63	84.62	+1.18	81.78	-2.21	81.32	-2.77
Overall slot F1		86.19	85.33	-1.00	85.33	-1.00	85.13	-1.23
Overall frame accuracy		46.44	46.01	-0.93	45.66	-1.69	44.38	-4.43

C Summary of the results on Language Bootstrapping Experiments averaged over three runs for Generator with token-level reward.

Metric	Domain	Baseline	Generator with token-level reward						Upsampled				
			TopX		Uniques		All		TopX		Uniques		All
		Perf.	Perf.	% change	Perf.	% change	Perf.	% change	Perf.	% change	Perf.	% change	Perf.	% change
Domain accuracy		96.83	94.08	-2.84	96.49	-0.35	97.11	+0.29	95.41	-1.47	97.05	+0.23	97.50	+0.70
Intent accuracy	**alarm**	73.33	75.04	+2.33	81.26	+10.81	82.01	+11.84	70.30	-4.13	78.38	+6.88	81.26	+10.82
	reminder	72.93	70.64	-3.14	81.40	+11.61	83.60	+14.62	74.30	+1.87	80.07	+9.79	82.47	+13.08
	weather	97.5	97.29	-0.21	92.38	-5.25	93.71	-3.89	98.06	+0.58	99.00	+1.53	99.02	+1.56
Overall intent accuracy		84	83.33	-0.80	86.26	+2.69	87.70	+4.41	83.40	-0.71	88.06	+4.84	89.64	+6.72
Slot F1	**alarm**	92.63	92.77	+0.15	91.68	-1.03	91.66	-1.04	91.58	-1.13	92.40	-0.25	92.70	+0.08
	reminder	83.49	79.36	-4.95	84.23	+0.89	84.16	+0.80	82.28	-1.45	84.48	+1.18	84.87	+1.65
	weather	83.63	84.30	+0.80	84.26	+0.75	84.68	+1.25	83.80	+0.21	83.40	-0.27	82.94	-0.83
Overall Slot F1		86.19	85.12	-1.24	86.37	+0.20	86.53	+0.39	85.56	-0.73	86.28	+0.10	86.26	+0.09
Overall frame accuracy		46.44	46.61	+0.38	46.81	+0.80	47.95	+3.26	45.47	-2.08	47.86	+3.05	48.47	+4.37
Metric	**Domain**	**Baseline**	**Generator with token-level Monte Carlo rollout**								**Upsampled**		
			TopX		**Uniques**		**All**		**TopX**		**Uniques**		**All**
		Perf.	**Perf.**	**% change**	**Perf.**	**% change**	**Perf.**	**% change**	**[SAME AS ABOVE]**		**Perf.**	**% change**	**Perf.**	**% change**
Domain accuracy		96.83	94.06	-2.86	97.22	+0.41	94.86	-2.03			97.81	+1.02	96.18	-0.67
Intent accuracy	**alarm**	73.33	70.10	-4.41	84.20	+14.83	82.63	+12.68			77.35	+5.48	80.65	+9.98
	reminder	72.93	71.34	-2.18	83.45	+14.43	77.58	+6.38			83.39	+14.35	79.49	+9.00
	weather	97.5	97.33	-0.17	97.38	-0.12	97.44	-0.06			98.95	+1.48	99.14	+1.69
Overall intent accuracy		84	81.86	-2.55	90.06	+7.21	87.67	+4.36			88.64	+5.53	88.61	+5.49
Slot F1	**alarm**	92.63	90.94	-1.82	92.23	-0.43	90.28	-2.53			92.05	-0.63	92.15	-0.52
	reminder	83.49	79.79	-4.43	85.85	+2.82	81.87	-1.94			85.45	+2.35	83.94	+0.54
	weather	83.63	85.39	+2.10	83.84	+0.25	84.94	+1.56			82.62	-1.21	83.56	-0.09
Overall Slot F1		86.19	85.33	-1.00	86.77	+0.67	85.57	-0.73			86.12	-0.08	86.10	-0.10
Overall frame accuracy		46.44	45.36	-2.33	48.44	+4.30	47.75	+2.82			47.70	+2.71	48.73	+4.92

D Summary of the results on Low-Resource Domain Experiments averaged over three runs. Pre-trained embeddings were built using robust domains and used in GAN model with token-level reward to inform the data generation for the low-resource domain.

ALARM AS LOW-RESOURCE DOMAIN						
Metric	**Domain**	**Baseline**	**Pre-trained embeddings with token-level reward**		**Pre-trained embeddings with MC rollout token-level reward**	
		Performance	**Performance**	**% change**	**Performance**	**% change**
Domain accuracy		99.19	99.13	-0.06	99.21	+0.03
Intent accuracy	**alarm**	74.75	84.07	+12.46	85.32	+14.14
	reminder	97.70	97.89	+0.20	97.98	+0.29
	weather	99.10	98.86	-0.24	99.24	+0.14
Overall intent accuracy		91.56	94.26	+2.95	94.81	+3.54
Slot F1	**alarm**	93.00	91.34	-1.79	92.46	-0.58
	reminder	92.79	92.70	-0.10	92.79	+0.00
	weather	97.79	97.78	-0.01	97.86	+0.07
Overall Slot F1		95.13	94.64	-0.52	95.00	-0.13
Overall frame accuracy		73.95	74.14	+0.27	74.57	+0.84
REMINDER AS LOW-RESOURCE DOMAIN						
Metric	**Domain**	**Baseline**	**Pre-trained embeddings with token-level reward**		**Pre-trained embeddings with MC rollout token-level reward**	
		Performance	**Performance**	**% change**	**Performance**	**% change**
Domain accuracy		99.68	99.57	-0.12	99.40	-0.28
Intent accuracy	**alarm**	96.75	96.80	+0.05	96.59	-0.16
	reminder	80.81	91.48	+13.21	93.24	+15.39
	weather	99.75	99.57	-0.18	99.37	-0.38
Overall intent accuracy		94.56	96.90	+2.47	97.15	+2.74
Slot F1	**alarm**	96.29	96.32	+0.03	96.24	-0.05
	reminder	87.90	87.63	-0.31	88.00	+0.11
	weather	97.96	97.83	-0.14	97.84	-0.13
Overall Slot F1		94.94	94.82	-0.13	94.92	-0.02
Overall frame accuracy		75.70	76.05	+0.46	76.16	+0.61
WEATHER AS LOW-RESOURCE DOMAIN						
Metric	**Domain**	**Baseline**	**Pre-trained embeddings with token-level reward**		**Pre-trained embeddings with MC rollout token-level reward**	
		Performance	**Performance**	**% change**	**Performance**	**% change**
Domain accuracy		96.64	95.46	-1.22	96.55	-0.09
Intent accuracy	**alarm**	92.66	89.94	-2.93	91.64	-1.10
	reminder	92.30	91.63	-0.72	93.30	+1.08
	weather	97.73	97.80	+0.08	98.23	+0.51
Overall intent accuracy		94.75	93.64	-1.18	94.86	+0.12
Slot F1	**alarm**	94.57	92.92	-1.74	93.75	-0.86
	reminder	89.25	88.82	-0.49	90.09	+0.94
	weather	84.72	84.36	-0.42	83.13	-1.87
Overall Slot F1		89.02	88.32	-0.78	88.38	-0.71
Overall frame accuracy		60.38	59.04	-2.22	58.29	-3.46

E Data quality evaluation results.

Model	**Pre-trained embeddings**	**Domain**	**Unique utterances**	**Unique words**	**Mean utterance length**	**BLEU1**	**BLEU2**	**BLEU3**	**BLEU4**
Generator with sentence-level reward without rollouts	**No**	**alarm**	1183	113	5.49	96.76	91.23	82.38	61.45
		reminder	2792	539	6.23	94.64	85.51	73.44	60.9
		weather	1378	104	5.33	99.77	96.03	90.78	81.44
Generator with sentence-level reward with Monte Carlo rollout	**No**	**alarm**	1064	111	5.73	97.91	94.37	88.72	77.1
		reminder	3487	540	6.75	95.1	84.83	70.99	53.93
		weather	1839	105	5.59	99.65	91.98	85.06	75.22
Generator with token-level reward	**No**	**alarm**	2350	122	6.24	94.7	87.49	78.17	66.12
		reminder	4729	580	6.78	98.97	85.71	69.35	54.98
		weather	1479	108	5.45	99.62	96.13	91.16	86.33
Generator with token-level Monte Carlo rollout	**No**	**alarm**	2123	123	6.11	96.29	90.32	75.79	62.03
		reminder	5144	593	6.52	99.01	87.03	67.57	49.29
		weather	1975	109	5.44	99.22	95.09	87.75	76.11
Generator with token-level reward	**Yes**	**alarm**	2545	116	6.07	94.59	83.96	75.01	65.82
		reminder	7048	572	7.07	98.03	79.2	58.34	41.35
		weather	1323	104	5.59	99.64	89.51	83.89	74.8
Generator with token-level Monte Carlo rollout	**Yes**	**alarm**	3451	112	7.22	91.45	80.24	69.59	56.38
		reminder	6128	586	6.58	96.08	78.14	58.32	39.87
		weather	2291	108	5.43	98.08	87.39	75.49	58.32