

MRP 2020: The Second Shared Task on Cross-Framework and Cross-Lingual Meaning Representation Parsing

— Appendix: In-Depth Per-Framework Results —

Stephan Oepen[♣], Omri Abend[♣], Lasha Abzianidze[♡], Johan Bos[◇], Jan Hajič[◊], Daniel Hershcovich^{*}, Bin Li[•], Tim O’Gorman[◊], Nianwen Xue^{*}, and Daniel Zeman[◊]

[♣] University of Oslo, Department of Informatics

[♣] The Hebrew University of Jerusalem, School of Computer Science and Engineering

[♡] Utrecht University, UiL OTS

[◇] University of Groningen, Faculteit der Letteren

[◊] Charles University, Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

^{*} University of Copenhagen, Department of Computer Science

[•] Nanjing Normal University, School of Chinese Language and Literature

[◊] University of Colorado at Boulder, Department of Linguistics

^{*} Brandeis University, Department of Computer Science

mrp-organizers@nlpl.eu

Abstract

The 2020 Shared Task at the Conference for Computational Language Learning (CoNLL) was devoted to Meaning Representation Parsing (MRP) across frameworks and languages (?). This supplementary material provides in-depth evaluation results for each of the five frameworks in the cross-framework track, as well as for the four frameworks in the cross-lingual track. All technical information regarding the task, including system submissions, official results, and links to supporting resources and software are available from the task web site at:

<http://mrp.nlpl.eu>

A Per-Framework Results

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	0.98	0.98	0.98	0.96	0.96	0.96	0.96	0.98	0.97	0.98	0.98	0.98	0.96	0.96	0.96	-	-	-	0.97	0.97	0.97
	0.92	0.92	0.92	0.94	0.93	0.94	0.94	0.94	0.94	0.95	0.94	0.94	0.93	0.92	0.93	-	-	-	0.94	0.93	0.94
ÚFAL	0.96	0.96	0.96	0.95	0.94	0.95	0.95	0.94	0.94	0.97	0.97	0.97	0.95	0.94	0.94	-	-	-	0.96	0.95	0.95
	0.9	0.9	0.9	0.94	0.93	0.93	0.95	0.94	0.94	0.94	0.93	0.93	0.91	0.91	0.91	-	-	-	0.93	0.92	0.93
HIT-SCIR	0.92	0.92	0.92	0.91	0.91	0.91	0.65	0.61	0.63	0.97	0.97	0.97	0.93	0.93	0.93	-	-	-	0.9	0.89	0.89
	0.87	0.85	0.86	0.9	0.9	0.9	0.64	0.7	0.67	0.93	0.93	0.93	0.89	0.89	0.89	-	-	-	0.87	0.88	0.87
HUJI-KU	0.8	0.73	0.76	0.77	0.81	0.79	0.37	0.03	0.06	0.9	0.95	0.93	0.85	0.88	0.86	-	-	-	0.83	0.76	0.79
	0.83	0.73	0.78	0.81	0.84	0.82	0.72	0.15	0.24	0.86	0.89	0.88	0.83	0.84	0.84	-	-	-	0.83	0.76	0.8
ISCAS	0.93	0.92	0.92	0.91	0.91	0.91	0.53	0.68	0.6	0.96	0.97	0.97	0.93	0.93	0.93	-	-	-	0.86	0.9	0.88
	0.88	0.85	0.86	0.91	0.91	0.91	0.49	0.61	0.54	0.93	0.93	0.93	0.9	0.9	0.9	-	-	-	0.85	0.87	0.86
TJU-BLCU	0.88	0.88	0.88	0.77	0.56	0.65	-	-	-	0.9	0.66	0.76	0.83	0.55	0.66	-	-	-	0.83	0.51	0.64
	0.72	0.72	0.72	0.81	0.4	0.53	-	-	-	0.88	0.43	0.58	0.84	0.34	0.49	-	-	-	0.84	0.35	0.49
JBNU	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ÚFAL	0.96	0.96	0.96	0.95	0.94	0.95	0.95	0.94	0.94	0.97	0.97	0.97	0.95	0.94	0.94	-	-	-	0.96	0.95	0.95
	0.9	0.9	0.9	0.94	0.93	0.93	0.95	0.94	0.94	0.94	0.93	0.93	0.91	0.91	0.91	-	-	-	0.93	0.92	0.93
ERG	0.93	0.93	0.93	0.96	0.97	0.96	0.96	0.98	0.97	0.98	0.99	0.98	0.92	0.93	0.93	-	-	-	0.95	0.96	0.96
	0.9	0.89	0.89	0.95	0.93	0.94	0.96	0.95	0.95	0.94	0.92	0.93	0.91	0.89	0.9	-	-	-	0.94	0.91	0.93

Table 1: Official results using the cross-framework MRP metric for the EDS dataset in the cross-framework track. Results for each submission is evaluated in two settings: the top one presents results for *The Little Prince* corpus, and the bottom one presents results for the full test set. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top), unofficial submissions (middle), and a reference submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	1	1	1	0.82	0.9	0.86	0.68	0.75	0.71	0.88	0.88	0.88	0.84	0.81	0.83	0.88	0.9	0.89	0.8	0.84	0.82
	1.0	1.0	1.0	0.89	0.94	0.91	0.94	0.93	0.94	0.94	0.94	0.94	0.81	0.8	0.8	0.74	0.7	0.72	0.89	0.89	0.89
ÚFAL	1	1	1	0.92	0.91	0.92	0.7	0.74	0.72	0.82	0.89	0.85	0.83	0.81	0.82	0.9	0.88	0.89	0.81	0.84	0.83
	1.0	1.0	1.0	0.95	0.95	0.95	0.94	0.91	0.92	0.87	0.95	0.91	0.79	0.8	0.8	0.73	0.74	0.73	0.88	0.89	0.88
HIT-SCIR	1	1	1	0.88	0.89	0.88	0.59	0.59	0.59	0.88	0.88	0.88	0.8	0.78	0.79	0.8	0.84	0.82	0.78	0.78	0.78
	1	1	1	0.9	0.91	0.91	0.84	0.79	0.81	0.94	0.94	0.94	0.77	0.77	0.77	0.68	0.67	0.67	0.85	0.84	0.84
HUJI-KU	1	0.99	0.99	-	-	-	0.64	0.55	0.59	0.79	0.79	0.79	0.67	0.61	0.64	-	-	-	0.71	0.49	0.58
	1.0	0.94	0.96	-	-	-	0.81	0.64	0.71	0.73	0.71	0.72	0.52	0.46	0.49	-	-	-	0.69	0.44	0.54
ISCAS	0.21	0.21	0.21	0.33	0.5	0.39	-	-	-	0.38	0.65	0.48	-	-	-	-	-	-	0.12	0.25	0.16
	0.17	0.17	0.17	0.42	0.61	0.5	-	-	-	0.4	0.64	0.49	0.0	0.0	0.0	-	-	-	0.14	0.26	0.18
TJU-BLCU	0.53	0.53	0.53	0.38	0.3	0.33	-	-	-	0.41	0.4	0.41	0.43	0.29	0.35	-	-	-	0.41	0.24	0.3
	0.8	0.8	0.8	0.58	0.33	0.42	-	-	-	0.27	0.19	0.22	0.24	0.12	0.16	-	-	-	0.38	0.15	0.21
JBNU	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ÚFAL	1	1	1	0.91	0.92	0.92	0.7	0.75	0.72	0.89	0.89	0.89	0.84	0.82	0.83	0.93	0.9	0.91	0.83	0.84	0.84
	1.0	1.0	1.0	0.94	0.95	0.95	0.94	0.91	0.92	0.95	0.95	0.95	0.79	0.8	0.8	0.73	0.73	0.73	0.89	0.89	0.89
ERG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 2: Official results using the cross-framework MRP metric for the PTG dataset in the cross-framework track. Results for each submission is evaluated in two settings: the top one presents results for *The Little Prince* corpus, and the bottom one presents results for the full test set. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top), unofficial submissions (middle), and a reference submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	0.91	0.91	0.91	-	-	-	-	-	-	0.97	0.94	0.96	0.77	0.71	0.74	0.67	0.39	0.49	0.86	0.8	0.83
	0.98	0.98	0.98	-	-	-	-	-	-	0.97	0.88	0.92	0.66	0.62	0.64	0.48	0.45	0.47	0.78	0.72	0.75
ÚFAL	0.91	0.91	0.91	-	-	-	-	-	-	0.97	0.98	0.98	0.74	0.71	0.73	0.8	0.44	0.57	0.84	0.82	0.83
	0.97	0.97	0.97	-	-	-	-	-	-	0.97	0.96	0.96	0.62	0.65	0.64	0.44	0.48	0.46	0.75	0.78	0.76
HIT-SCIR	1	1	1	-	-	-	-	-	-	0.97	0.98	0.98	0.74	0.69	0.71	0.69	0.38	0.49	0.84	0.8	0.82
	1	1	1	-	-	-	-	-	-	0.96	0.95	0.96	0.61	0.6	0.61	0.44	0.42	0.43	0.75	0.74	0.75
HUJI-KU	0.96	0.94	0.95	-	-	-	-	-	-	0.96	0.95	0.96	0.68	0.64	0.66	0.6	0.3	0.4	0.8	0.76	0.78
	0.98	0.98	0.98	-	-	-	-	-	-	0.95	0.95	0.95	0.58	0.59	0.59	0.36	0.38	0.37	0.73	0.73	0.73
ISCAS	1	1	1	-	-	-	-	-	-	0.06	0.01	0.02	0.51	0.08	0.14	-	-	-	0.45	0.08	0.13
	1.0	1.0	1.0	-	-	-	-	-	-	0.13	0.01	0.02	0.41	0.03	0.05	-	-	-	0.42	0.03	0.06
TJU-BLCU	1	0.98	0.99	-	-	-	-	-	-	0.48	0.16	0.23	0.43	0.07	0.12	-	-	-	0.52	0.13	0.21
	1.0	0.98	0.99	-	-	-	-	-	-	0.49	0.07	0.13	0.37	0.03	0.05	-	-	-	0.5	0.06	0.1
JBNU	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ÚFAL	0.91	0.91	0.91	-	-	-	-	-	-	0.97	0.98	0.98	0.74	0.71	0.73	0.8	0.44	0.57	0.84	0.81	0.83
	0.97	0.97	0.97	-	-	-	-	-	-	0.97	0.96	0.96	0.62	0.65	0.64	0.44	0.48	0.46	0.75	0.78	0.76
ERG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 3: Official results using the cross-framework MRP metric for the UCCA dataset in the cross-framework track. Results for each submission is evaluated in two settings: the top one presents results for *The Little Prince* corpus, and the bottom one presents results for the full test set. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top), unofficial submissions (middle), and a reference submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	0.84	0.84	0.84	0.83	0.85	0.84	0.86	0.77	0.81	-	-	-	0.71	0.73	0.72	-	-	-	0.78	0.79	0.79
	0.86	0.86	0.86	0.88	0.86	0.87	0.83	0.81	0.82	-	-	-	0.77	0.74	0.76	-	-	-	0.83	0.8	0.82
ÚFAL	0.86	0.86	0.86	0.85	0.87	0.86	0.78	0.71	0.75	-	-	-	0.69	0.71	0.7	-	-	-	0.77	0.79	0.78
	0.84	0.84	0.84	0.88	0.87	0.88	0.86	0.85	0.85	-	-	-	0.73	0.7	0.71	-	-	-	0.81	0.79	0.8
HIT-SCIR	0.85	0.85	0.85	0.73	0.77	0.75	0.56	0.45	0.5	-	-	-	0.62	0.64	0.63	-	-	-	0.68	0.71	0.7
	0.85	0.85	0.85	0.81	0.77	0.79	0.71	0.39	0.5	-	-	-	0.65	0.6	0.63	-	-	-	0.74	0.66	0.7
HUJI-KU	0.71	0.69	0.7	0.65	0.62	0.64	0.15	0.14	0.15	-	-	-	0.47	0.37	0.41	-	-	-	0.56	0.5	0.53
	0.58	0.56	0.57	0.64	0.6	0.62	0.46	0.4	0.43	-	-	-	0.5	0.38	0.44	-	-	-	0.57	0.49	0.52
ISCAS	0.67	0.66	0.66	0.76	0.54	0.63	0.34	0.38	0.36	-	-	-	0.64	0.38	0.47	-	-	-	0.68	0.47	0.56
	0.7	0.7	0.7	0.84	0.62	0.71	0.59	0.48	0.53	-	-	-	0.66	0.43	0.52	-	-	-	0.74	0.53	0.61
TJU-BLCU	0.91	0.91	0.91	0.48	0.32	0.38	-	-	-	-	-	-	0.43	0.3	0.35	-	-	-	0.5	0.34	0.4
	0.86	0.86	0.86	0.59	0.24	0.34	-	-	-	-	-	-	0.44	0.17	0.24	-	-	-	0.54	0.21	0.3
JBNU	0.86	0.86	0.86	0.79	0.8	0.79	0.54	0.45	0.49	-	-	-	0.68	0.67	0.68	-	-	-	0.74	0.73	0.74
	0.84	0.84	0.84	0.79	0.73	0.76	0.68	0.39	0.5	-	-	-	0.61	0.54	0.57	-	-	-	0.71	0.62	0.66
ÚFAL	0.86	0.86	0.86	0.85	0.87	0.86	0.78	0.71	0.75	-	-	-	0.69	0.71	0.7	-	-	-	0.77	0.79	0.78
	0.84	0.84	0.84	0.88	0.87	0.88	0.86	0.85	0.85	-	-	-	0.73	0.7	0.71	-	-	-	0.81	0.79	0.8
ERG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 4: Official results using the cross-framework MRP metric for the AMR dataset in the cross-framework track. Results for each submission is evaluated in two settings: the top one presents results for *The Little Prince* corpus, and the bottom one presents results for the full test set. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top), unofficial submissions (middle), and a reference submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	1.0	1.0	1.0	0.91	0.89	0.9	-	-	-	-	-	-	0.96	0.94	0.95	-	-	-	0.94	0.92	0.93
ÚFAL	0.99	0.99	0.99	0.93	0.92	0.92	-	-	-	-	-	-	0.96	0.94	0.95	-	-	-	0.95	0.93	0.94
HIT-SCIR	0.99	0.99	0.99	0.87	0.87	0.87	-	-	-	-	-	-	0.91	0.89	0.9	-	-	-	0.9	0.89	0.89
HUJI-KU	0.99	0.94	0.96	-	-	-	-	-	-	-	-	-	0.83	0.82	0.83	-	-	-	0.84	0.5	0.63
ISCAS	0.99	0.99	0.99	0.63	0.6	0.61	-	-	-	-	-	-	0.9	0.62	0.74	-	-	-	0.78	0.63	0.69
TJU-BLCU	0.35	0.35	0.35	0.74	0.5	0.6	-	-	-	-	-	-	0.32	0.22	0.26	-	-	-	0.49	0.34	0.4
JBNU	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ÚFAL	0.99	0.99	0.99	0.93	0.92	0.92	-	-	-	-	-	-	0.96	0.94	0.95	-	-	-	0.95	0.93	0.94
ERG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 5: Official results using the cross-framework MRP metric for the DRG dataset in the cross-framework track. Results for each submission is evaluated in two settings: the top one presents results for *The Little Prince* corpus, and the bottom one presents results for the full test set. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top), unofficial submissions (middle), and a reference submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	1.0	1.0	1.0	0.86	0.89	0.88	0.93	0.9	0.91	0.95	0.93	0.94	0.8	0.74	0.77	0.72	0.61	0.66	0.89	0.86	0.87
ÚFAL	1.0	1.0	1.0	0.97	0.96	0.96	0.94	0.93	0.94	0.88	0.98	0.92	0.85	0.83	0.84	0.78	0.77	0.78	0.91	0.91	0.91
HIT-SCIR	1.0	1.0	1.0	0.85	0.83	0.84	0.85	0.74	0.79	0.92	0.92	0.92	0.68	0.62	0.65	0.51	0.45	0.48	0.82	0.75	0.78
HUJI-KU	0.99	0.91	0.95	-	-	-	0.69	0.76	0.72	0.73	0.71	0.72	0.47	0.43	0.45	-	-	-	0.65	0.53	0.58
TJU-BLCU	1	1	1	0.61	0.29	0.39	-	-	-	0.44	0.27	0.34	0.42	0.2	0.27	-	-	-	0.51	0.14	0.22
ÚFAL	1.0	1.0	1.0	0.97	0.96	0.96	0.94	0.93	0.94	0.98	0.98	0.98	0.86	0.83	0.85	0.8	0.77	0.79	0.93	0.92	0.92

Table 6: Official results using the cross-framework MRP metric for the PTG dataset in the cross-lingual track. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top) and an unofficial submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	1.0	1.0	1.0	-	-	-	-	-	-	0.96	0.92	0.94	0.7	0.69	0.69	0.35	0.46	0.4	0.79	0.79	0.79
ÚFAL	0.97	0.97	0.97	-	-	-	-	-	-	0.96	0.98	0.97	0.68	0.74	0.71	0.4	0.55	0.46	0.79	0.83	0.81
HIT-SCIR	0.99	0.99	0.99	-	-	-	-	-	-	0.94	0.96	0.95	0.69	0.72	0.7	0.39	0.5	0.44	0.78	0.82	0.8
HUJI-KU	0.98	0.98	0.98	-	-	-	-	-	-	0.93	0.95	0.94	0.61	0.62	0.61	0.32	0.32	0.32	0.74	0.76	0.75
ÚFAL	0.97	0.97	0.97	-	-	-	-	-	-	0.96	0.98	0.97	0.68	0.74	0.71	0.4	0.55	0.46	0.79	0.83	0.81

Table 7: Official results using the cross-framework MRP metric for the UCCA dataset in the cross-lingual track. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top) and an unofficial submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	0.86	0.86	0.86	0.84	0.81	0.83	0.82	0.77	0.79	0.9	0.88	0.89	0.73	0.69	0.71	-	-	-	0.82	0.79	0.8
ÚFAL	0.83	0.83	0.83	0.84	0.85	0.85	0.79	0.77	0.78	0.76	0.93	0.84	0.66	0.68	0.67	-	-	-	0.75	0.81	0.78
HIT-SCIR	0.8	0.8	0.8	0.64	0.64	0.64	0.0	0.0	0.0	-	-	-	0.58	0.56	0.57	-	-	-	0.6	0.42	0.49
HUJI-KU	0.64	0.53	0.58	0.6	0.62	0.61	0.56	0.53	0.54	-	-	-	0.48	0.44	0.46	-	-	-	0.55	0.38	0.45
TJU-BLCU	0.79	0.79	0.79	0.46	0.16	0.24	-	-	-	0.47	0.2	0.28	0.39	0.13	0.2	-	-	-	0.46	0.17	0.25
ÚFAL	0.82	0.82	0.82	0.84	0.84	0.84	0.79	0.79	0.79	0.91	0.91	0.91	0.68	0.68	0.68	-	-	-	0.81	0.8	0.81

Table 8: Official results using the cross-framework MRP metric for the AMR dataset in the cross-lingual track. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top) and an unofficial submission (bottom).

Teams	Tops			Labels			Properties			Anchors			Edges			Attributes			All		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hitachi	0.99	0.99	0.99	0.88	0.89	0.88	-	-	-	-	-	-	0.96	0.97	0.97	-	-	-	0.93	0.94	0.93
ÚFAL	0.99	0.99	0.99	0.85	0.84	0.84	-	-	-	-	-	-	0.94	0.92	0.93	-	-	-	0.9	0.89	0.9
HIT-SCIR	1	1	1	0.62	0.62	0.62	-	-	-	-	-	-	0.71	0.71	0.71	-	-	-	0.68	0.69	0.68
HUJI-KU	0.98	0.93	0.95	-	-	-	-	-	-	-	-	-	0.81	0.83	0.82	-	-	-	0.82	0.5	0.62
TJU-BLCU	0.44	0.43	0.43	0.58	0.37	0.46	-	-	-	-	-	-	0.31	0.21	0.25	-	-	-	0.42	0.28	0.34
ÚFAL	0.99	0.99	0.99	0.85	0.84	0.84	-	-	-	-	-	-	0.94	0.92	0.93	-	-	-	0.9	0.89	0.9

Table 9: Official results using the cross-framework MRP metric for the DRG dataset in the cross-lingual track. Results are broken down by ‘atomic’ component pieces. For each component we report precision (P), recall (R), and F₁ score (F). Entries are split into the same three blocks as in the main paper: official submissions (top) and an unofficial submission (bottom).