

Online supplement 1

A Discussion

A.1. *Inter-annotator agreement (IAA)*

The low agreement in UPMC radiology reports is due to the fact that A1 missed a number of pairs that involve markables of type anatomy. For example, as Table Supplement 1 (a) shows, A1 failed to identify the relations among all 13 anatomy pairs that are related to the spine. These pairs were, however, captured by A3, and reflected subsequently in the gold standard.

The complementary nature of the annotations is further supported by our error analysis. Examining the IAA scores for the files in the set with the lowest score (p6) (Table Supplement 1 (b)) shows that a Mayo Clinic pathology report doc210 had an IAA of 0.16, with only one overlapping annotation (true positive). The full set of annotations in this report is in Table Supplement (c). There was only one overlapping annotation before the consensus stage. After the consensus stage, all but three of the original annotations were included in the gold standard. Figures 2, 3 and 4 show the contributions of each annotator in the UPMC and Mayo reports. Overall in UPMC reports, A1 and A3 individually contributed 41% and 42% to the gold standard, respectively, and only 18% were from both annotators. In the Mayo reports, the contributions are 10%, 12%, 24% for A1, A2, and A3 respectively, and 54% from two annotators. In general, A3, who had more domain and linguistic knowledge than the other annotators, contributed more pairs than any other annotator.

IAA Examples

Savova “Anaphoric relations in the clinical domain”

Antecedent ID	Anaphor ID	Relation type	Antecedent span offsets	Antecedent text	Anaphor span offsets	Anaphor text
2.1	2.2	#identity	185-197	SPINE LUMBAR	307-317	lower back
2.2	2.3	#identity	307-317	lower back	420-436	the lumbar spine
2.3	2.4	#part/whole	420-436	the lumbar spine	475-495	the L5-S1 disc space
2.3	2.5	#identity	420-436	the lumbar spine	568-584	the lumbar
2.3	2.6	#part/whole	420-436	the lumbar spine	672-677	L1-L2
2.3	2.7	#part/whole	420-436	the lumbar spine	679-684	L2-L3
2.3	2.8	#part/whole	420-436	the lumbar spine	686-691	L3-L4
2.3	2.9	#part/whole	420-436	the lumbar spine	696-702	L4- L5
2.4	2.13	#identity	475-495	the L5-S1 disc space	1402-1407	L5-S1
2.7	2.1	#identity	679-684	L2-L3	843-848	L2-L3
2.8	2.11	#identity	686-691	L3-L4	956-961	L3-L4
2.9	2.12	#identity	696-702	L4- L5	1145-1150	L4-L5
2.12	2.14	#identity	1145-1150	L4-L5	1594-1599	L4-L5

(a) Annotations for UPMC report 101

	doc224	doc161	doc190	doc210
<i>True Positive</i>	5	2	4	1
<i>False Positive</i>	3	4	6	6
<i>False Negative</i>	4	7	13	4
<i>Precision</i>	0.6250	0.3333	0.4000	0.1429
<i>Recall</i>	0.5556	0.2222	0.2353	0.2000
<i>F-Score</i>	0.5882	0.2667	0.2963	0.1667
<i>Kappa</i>	0.5854	0.2623	0.2939	0.1622

(b) IAA results per document from p6 set (Mayo pathology notes set 6)

Antecedent ID	Anaphor ID	Relation type	Antecedent span offsets	Antecedent text	Anaphor span offsets	Anaphor text
1.1	1.4	#identity	35-40;42-49	Colon ... sigmoid	110-117	colonic
2.1	2.4	#part/whole	35-49	Colon, sigmoid	110-117	colonic
2.1	2.7	#part/whole	35-49	Colon, sigmoid	204-232	the closest resection margin
2.1	2.9	#part/whole	35-49	Colon, sigmoid	273-283	muscularis
2.1	2.12	#part/whole	35-49	Colon, sigmoid	318-355	Proximal and distal resection margins
1.3	1.6	#identity	95-132	Grade 3 (of 4) colonic adenocarcinoma	155-181	a mass, 5.5 x 3.8 x 1.3 cm
1.4	1.9	#part/whole	110-117	colonic	269-283	the muscularis
2.5	2.6	#identity	118-132	adenocarcinoma	155-161	a mass
1.6	1.8	#identity	155-181	a mass, 5.5 x 3.8 x 1.3 cm	247-252	Tumor
			155-161	a mass	247-252	Tumor
2.7	2.12	#set/subset	204-232	the closest resection margin	318-355	Proximal and distal resection margins
1.11	1.13	#set/subset	382-393;401-424	Multiple (3 ... pericolic lymph nodes	397-424	15) pericolic lymph nodes

(c) Annotations for Mayo clinical note doc210. All but the grey shaded annotations were included in the gold standard after the consensus

Table Supplement 1: IAA Examples