

# Evaluation Scores

## Workshop SANTA

Sept. 19, 2018

# Conceptual Coverage

Group	C1		C2		C3		C4	
SANTA1	3.00±	1.10	1.00±	0.00	1.00±	0.00	2.83±	0.98
SANTA2	2.67±	1.03	3.00±	1.10	2.83±	0.75	2.67±	1.21
SANTA3	2.43±	0.98	2.86±	1.07	2.71±	1.38	2.29±	0.76
SANTA4	3.71±	0.49	4.00±	0.00	3.71±	0.49	3.00±	0.82
SANTA5	3.71±	0.49	4.00±	0.00	3.43±	0.53	3.00±	1.00
SANTA6	3.33±	0.82	3.33±	0.82	2.67±	0.52	3.00±	0.89
SANTA7	2.71±	1.11	3.33±	0.82	2.71±	0.95	2.83±	1.33
SANTA8	2.29±	1.11	1.00±	0.00	1.67±	1.21	3.14±	0.69

# Conceptual Coverage

Group	Points
SANTA4	14.43
SANTA5	14.14
SANTA6	12.33
SANTA7	11.60
SANTA2	11.17
SANTA3	10.29
SANTA8	8.10
SANTA1	7.83

# Usefulness

## Points

Group	U1		U2		U3		U4	
SANTA1	3.17±	0.75	3.00±	0.71	2.17±	0.75	1.67±	0.52
SANTA2	3.50±	0.55	3.40±	0.89	3.00±	0.89	2.67±	0.52
SANTA3	3.33±	0.82	3.17±	0.98	2.29±	0.76	2.17±	0.75
SANTA4	3.29±	0.76	2.83±	0.98	2.71±	0.76	2.43±	0.98
SANTA5	3.50±	0.55	3.40±	0.55	3.14±	0.69	2.83±	0.75
SANTA6	3.40±	0.55	3.40±	0.89	2.17±	0.75	2.40±	0.55
SANTA7	3.00±	0.63	2.60±	0.55	2.17±	0.98	2.00±	0.89
SANTA8	3.20±	0.84	2.80±	1.10	1.29±	0.76	1.83±	0.75

# Usefulness

## Ranking

Group	Points
SANTA5	12.88
SANTA2	12.57
SANTA6	11.37
SANTA4	11.26
SANTA3	10.95
SANTA1	10.00
SANTA7	9.77
SANTA8	9.12

# Applicability

## Points

Group	A1	A2
SANTA1	2.60± 0.55	2.67± 0.52
SANTA2	3.17± 0.41	2.17± 0.98
SANTA3	2.57± 1.13	1.43± 0.53
SANTA4	3.57± 0.53	2.14± 0.38
SANTA5	3.00± 1.15	2.29± 0.95
SANTA6	3.17± 1.17	2.00± 0.89
SANTA7	2.33± 0.52	1.17± 0.41
SANTA8	2.71± 1.38	3.43± 0.79

# Applicability

## Ranking

Group	Points
SANTA8	6.14
SANTA4	5.71
SANTA2	5.33
SANTA5	5.29
SANTA1	5.27
SANTA6	5.17
SANTA3	4.00
SANTA7	3.50

# Inter-Annotator Agreement

- ▶ SANTA<sub>1</sub>
  - ▶ embedded narrative, narrative
- ▶ SANTA<sub>2</sub>
  - ▶ level
  - ▶ degree
- ▶ SANTA<sub>3</sub>
  - ▶ quaternary n., primary n., quinary n., tertiary n., secondary n., senary n.
- ▶ SANTA<sub>4</sub>
  - ▶ narrative level
  - ▶ degree of embedding, sequence



- ▶ SANTA5
  - ▶ level 1-5, 1st narrative act (level 1-3), 2nd narrative act (level 1-3), 3rd narrative act (level 1-3)
  - ▶ upper level: crossed boundary, upper level: head
- ▶ SANTA6
  - ▶ Narrative Level Change, Narrative Level
  - ▶ Number
- ▶ SANTA7
  - ▶ voice, FOC\_INT, foc, narrative\_transmission, FOC\_UNR, CHARACTERS, FOC\_EXT, VOICE 1-3, AUTHOR, SCENE, NARRATOR
- ▶ SANTA8
  - ▶ nframe
  - ▶ level

# All Annotations

Conceptual Coverage

Usefulness

Applicability

Inter-Annotator Agreement

**All Annotations**

Own vs. Foreign Annotations

Own vs. Student Annotations

Foreign vs. Student Annotations

# Results

## Observed Disorder (all annotations)

SANTA_	1	2	3	4	5	6	7	8
Buechner	2.20	1.92	2.66	2.92	1.71	1.95	1.70	1.28
Chekhov	1.72	1.18	2.45	2.64	2.41	2.00	1.17	2.50
Henry	1.69	1.71	1.88	2.34	2.19	1.93	1.35	1.70
Kafka	2.19	2.06	2.00	1.50	1.16	2.06	1.22	1.00
Kleist	1.84	1.22	1.17	1.04	1.89	1.13	1.52	0.75
Lagerloef	1.01	1.67	2.10	2.38	1.48	2.57	0.96	0.67
Storm	1.76	1.50	2.36	2.10	2.02	1.42	1.55	1.36
Tieck	1.25	2.10	2.75	2.72	2.55	2.06	1.08	1.85
Min	1.01	1.18	1.17	1.04	1.16	1.13	0.96	0.67
Mean	1.71	1.67	2.17	2.20	1.93	1.89	1.32	1.39
Max	2.20	2.10	2.75	2.92	2.55	2.57	1.70	2.50
Stddev.	0.41	0.35	0.51	0.64	0.47	0.44	0.26	0.61

**Table:** Observed Disorder (the lower the better)

# Results

## Inter-Annotator Agreement Gamma (all annotations)

SANTA_	1	2	3	4	5	6	7	8
Buechner	-0.08	0.04	-0.13	-0.41	0.25	0.15	0.12	
Chekhov	0.05	0.30	-0.11	-0.14	0.02	-0.30	0.35	
Henry	-0.02	-0.01	0.08	-0.07	0.02	-0.05	0.32	
Kafka	-0.08				0.50	0.02	0.43	
Kleist	0.32	0.51				0.56	0.06	0.50
Lagerloef	0.15	0.06	0.04	-0.10	0.25	0.05	0.42	
Storm	0.05	0.28	-0.05	-0.02	0.11	0.25	0.28	0.38
Tieck	0.10	-0.03	-0.13	-0.35	-0.06	-0.02	0.38	
Min	-0.08	-0.03	-0.13	-0.41	-0.06	-0.30	0.06	0.38
Mean	0.06	0.16	-0.05	-0.18	0.16	0.08	0.30	0.44
Max	0.32	0.51	0.08	-0.02	0.50	0.56	0.43	0.50
Stddev.	0.13	0.20	0.09	0.16	0.19	0.25	0.14	0.08

**Table:** Inter-Annotator Agreement Gamma (the higher the better)

# Own vs. Foreign Annotations

Conceptual Coverage

Usefulness

Applicability

Inter-Annotator Agreement

All Annotations

**Own vs. Foreign Annotations**

Own vs. Student Annotations

Foreign vs. Student Annotations

# Results

## Observed Disorder (own vs. foreign)

SANTA_	1	2	3	4	5	6	7	8
Buechner	1.63	1.92	1.50	1.86	1.14	1.11	0.91	0.86
Chekhov	1.45	1.18	1.81	1.96	1.68	2.00	0.54	1.84
Henry	1.33	1.71	2.00	2.02	0.00	1.30	0.98	2.12
Kafka	0.00	2.06	1.33	1.50	0.82	2.06	0.52	0.00
Kleist	1.23	0.85	1.00	0.85	2.00	1.19	0.00	0.00
Lagerloef	0.77	1.67	2.00	1.82	1.85	2.00	0.45	1.00
Storm	1.96	1.22	1.23	1.64	1.42	1.25	1.23	1.51
Tieck	1.11	1.49	1.40	1.92	1.67	1.68	0.00	0.93
Min	0.00	0.85	1.00	0.85	0.00	1.11	0.00	0.00
Mean	1.18	1.51	1.54	1.70	1.32	1.57	0.58	1.03
Max	1.96	2.06	2.00	2.02	2.00	2.06	1.23	2.12
Stddev.	0.41	0.35	0.51	0.64	0.47	0.44	0.26	0.61

Table: Observed Disorder (the lower the better)

# Results

## Inter-Annotator Agreement Gamma (own vs. foreign)

SANTA_	1	2	3	4	5	6	7	8
Buechner	-0.38	0.02	-0.06	-0.36	0.33	0.17	0.28	0.45
Chekhov	-0.11	0.30	-0.09	-0.28	0.02	-0.31	0.60	-0.06
Henry	-0.05	-0.08	-0.36	-0.35	1.00	0.13	0.34	-0.35
Kafka	1.00		0.00	0.02	0.54	0.01	0.64	1.00
Kleist	0.07	0.48		0.35		0.31	1.00	1.00
Lagerloef	0.29	0.04	-0.26	-0.22	-0.14	-0.26	0.62	
Storm	-0.65	0.28	0.19	-0.03	0.17	0.18	0.24	0.05
Tieck	-0.05	0.00	0.15	-0.36	0.00	-0.16	1.00	0.31
Min	-0.65	-0.08	-0.36	-0.36	-0.14	-0.31	0.24	-0.35
Mean	0.02	0.15	-0.06	-0.16	0.27	0.01	0.59	0.34
Max	1.00	0.48	0.19	0.35	1.00	0.31	1.00	1.00
Stddev.	0.13	0.20	0.09	0.16	0.19	0.25	0.14	0.08

**Table:** Inter-Annotator Agreement Gamma (the higher the better)

# Own vs. Student Annotations

Conceptual Coverage

Usefulness

Applicability

Inter-Annotator Agreement

All Annotations

Own vs. Foreign Annotations

**Own vs. Student Annotations**

Foreign vs. Student Annotations



# Results

## Observed Disorder (own vs. student)

SANTA_	1	2	3	4	5	6	7	8
Buechner	1.22	0.00	1.74	3.73	1.39	1.91	1.75	1.33
Chekhov	1.43	0.00	1.76	1.33	1.72	0.00	1.07	1.33
Henry	1.18	0.00	1.10	1.50	1.52	1.71	1.13	1.50
Kafka	0.00	0.00	1.33	0.00	0.91	0.00	1.01	1.00
Kleist	2.50	1.09	1.52	2.45	1.73	0.88	1.52	0.75
Lagerloef	0.65	0.00	1.10	0.67	0.58	1.94	0.95	0.00
Storm	0.77	1.05	1.77	0.57	1.63	1.17	1.13	0.25
Tieck	0.64	1.33	1.91	0.64	1.91	1.58	1.08	1.00
Min	0.00	0.00	1.10	0.00	0.58	0.00	0.95	0.00
Mean	1.05	0.43	1.53	1.36	1.42	1.15	1.21	0.90
Max	2.50	1.33	1.91	3.73	1.91	1.94	1.75	1.50
Stddev.	0.41	0.35	0.51	0.64	0.47	0.44	0.26	0.61

Table: Observed Disorder (the lower the better)

# Results

## Inter-Annotator Agreement Gamma (own vs. student)

SANTA_	1	2	3	4	5	6	7	8
Buechner	-0.04	1.00	-0.05		0.13	-0.13	-0.09	0.02
Chekhov	-0.10	1.00	-0.06	-0.04	0.03	1.00	0.28	-0.08
Henry	-0.18	1.00	0.18	-0.04	0.01	0.01	0.21	-0.05
Kafka	1.00	1.00	0.05	1.00	0.40	1.00	0.36	
Kleist	-0.33	0.39			0.03	0.49	0.03	0.50
Lagerloef	0.25	1.00	0.27	0.40	0.57	-0.33	0.31	1.00
Storm	0.13	0.27	-0.09	0.64	0.01	0.28	0.29	0.83
Tieck	0.25	0.04	-0.07		-0.16	-0.01	0.37	0.31
Min	-0.33	0.04	-0.09	-0.04	-0.16	-0.33	-0.09	-0.08
Mean	0.12	0.71	0.03	0.39	0.13	0.29	0.22	0.36
Max	1.00	1.00	0.27	1.00	0.57	1.00	0.37	1.00
Stddev.	0.13	0.20	0.09	0.16	0.19	0.25	0.14	0.08

**Table:** Inter-Annotator Agreement Gamma (the higher the better)

# Foreign vs. Student Annotations

Conceptual Coverage

Usefulness

Applicability

Inter-Annotator Agreement

All Annotations

Own vs. Foreign Annotations

Own vs. Student Annotations

Foreign vs. Student Annotations

# Results

## Observed Disorder (foreign vs. student)

SANTA_	1	2	3	4	5	6	7	8
Buechner	1.92	0.00	1.96	1.86	1.48	1.94	1.44	1.02
Chekhov	1.20	0.00	1.12	1.68	2.17	0.00	1.08	1.51
Henry	1.36	0.00	1.94	1.54	1.52	1.54	1.04	1.52
Kafka	2.19	0.00	0.00	1.50	1.17	0.00	1.25	1.00
Kleist	1.67	0.85	0.97	0.85	1.32	0.77	0.00	0.00
Lagerloef	0.87	0.00	1.92	1.64	1.73	1.70	0.85	1.00
Storm	2.09	1.33	1.98	1.76	1.69	1.16	1.38	1.38
Tieck	1.18	1.88	2.00	1.98	1.88	1.63	0.00	1.20
Min	0.87	0.00	0.00	0.85	1.17	0.00	0.00	0.00
Mean	1.56	0.51	1.49	1.60	1.62	1.09	0.88	1.08
Max	2.19	1.88	2.00	1.98	2.17	1.94	1.44	1.52
Stddev.	0.41	0.35	0.51	0.64	0.47	0.44	0.26	0.61

Table: Observed Disorder (the lower the better)

# Results

## Inter-Annotator Agreement Gamma (foreign vs. student)

SANTA_	1	2	3	4	5	6	7	8
Buechner	-0.42	1.00	-0.33	-0.32	0.10	-0.08	0.01	0.16
Chekhov	0.16	1.00	0.29	-0.04	-0.21	1.00	0.26	
Henry	0.01	1.00	-0.34	0.02	0.06	-0.14	0.38	0.01
Kafka	-0.12	1.00	1.00	0.00	0.33	1.00	0.30	
Kleist	0.17	0.53		0.39		0.56	1.00	1.00
Lagerloef	0.18	1.00	-0.38	-0.12	-0.10	-0.01	0.41	
Storm	-0.43	0.23	-0.25	-0.12	-0.03	0.22	0.21	0.17
Tieck	0.09	-0.18	-0.11	-0.43	-0.01	-0.19	1.00	0.19
Min	-0.43	-0.18	-0.38	-0.43	-0.21	-0.19	0.01	0.01
Mean	-0.04	0.70	-0.02	-0.08	0.02	0.30	0.45	0.31
Max	0.18	1.00	1.00	0.39	0.33	1.00	1.00	1.00
Stddev.	0.13	0.20	0.09	0.16	0.19	0.25	0.14	0.08

**Table:** Inter-Annotator Agreement Gamma (the higher the better)

# Applicability

## Ranking

Guidelines	A1	A2	IAA	IAA	Sum
SANTA8	2.71	3.43	4.00	4.00	14.14
SANTA2	3.17	2.17	2.65	2.65	10.62
SANTA5	3.00	2.29	2.65	2.65	10.58
SANTA7	2.33	1.17	3.32	3.32	10.15
SANTA6	3.17	2.00	2.26	2.26	9.68
SANTA1	2.60	2.67	2.16	2.16	9.59
SANTA3	2.57	1.43	1.63	1.63	7.26
SANTA4	3.57	2.14	1.00	1.00	7.71

# All Rankings

Guideline	Conc.	Appl.	Usab.	Sum
SANTA1	7.83	9.59	10.00	27.42
SANTA2	11.17	10.62	12.57	34.36
SANTA3	10.29	7.26	10.95	28.50
SANTA4	14.43	7.71	11.26	33.40
SANTA5	14.14	10.58	12.88	37.60
SANTA6	12.33	9.68	11.37	33.38
SANTA7	11.60	10.15	9.77	31.51
SANTA8	8.10	14.14	9.12	31.36