DH-FBK at SemEval-2022

Leveraging annotators' disagreement and multiple data views for patronizing language detection
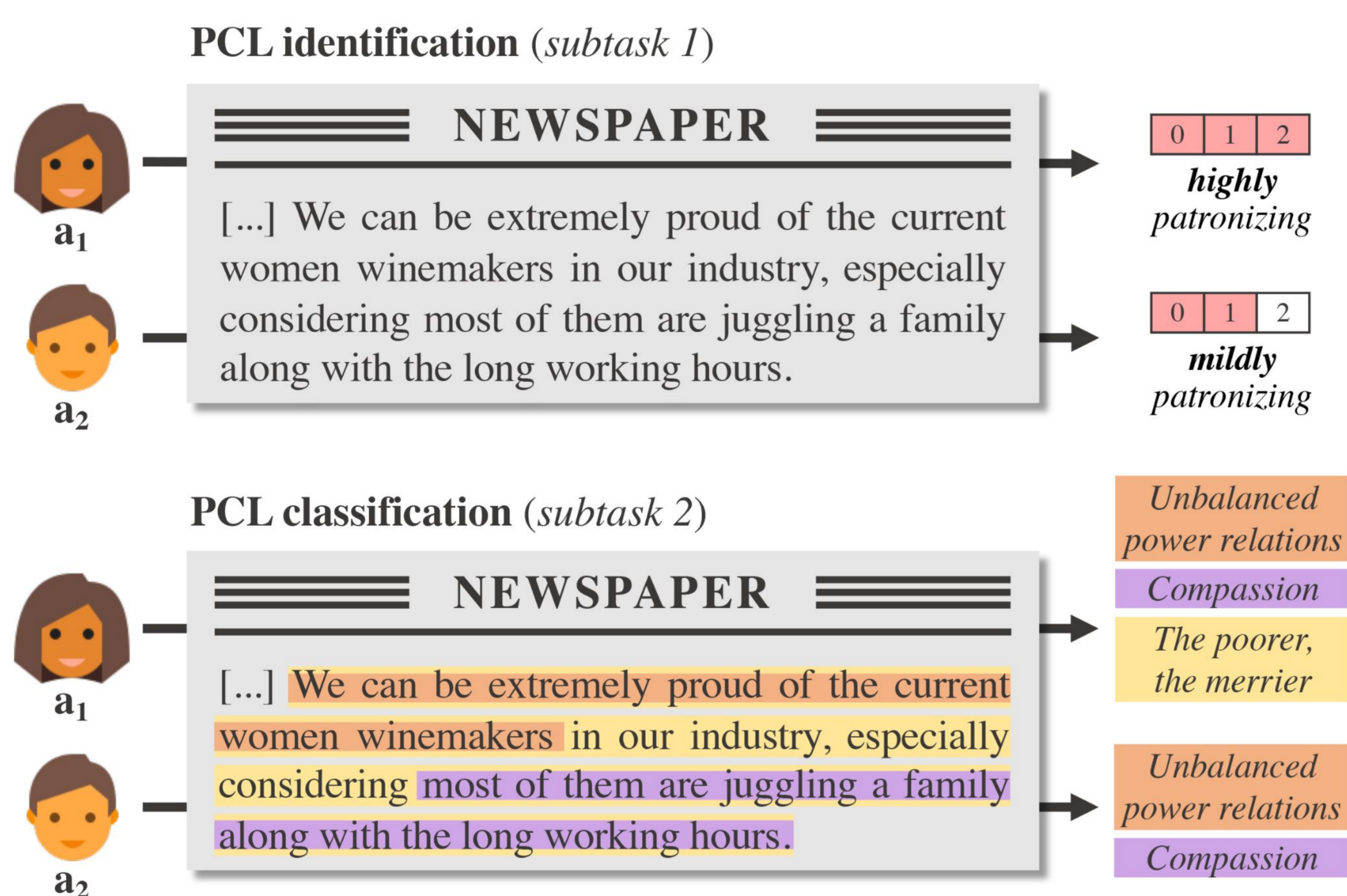
Alan **Ramponi** & Elisa **Leonardelli** – Fondazione Bruno Kessler, Trento, Italy

Background

Patronizing & condescending language (PCL) Language use denoting superior attitude towards others, who are depicted in a compassionate way [Pérez-Almendros et al., 2020]

- Drives exclusion of already vulnerable communities
- Detection has social impact (e.g., suggestion tools)

Challenges Annotators often perceive PCL differently due to background/sensibility, and thus annotate it in different ways

PCL identification (subtask 1)

NEWSPAPER

[...] We can be extremely proud of the current women winemakers in our industry, especially considering most of them are juggling a family along with the long working hours.

highly patronizing

mildly patronizing

PCL classification (subtask 2)

NEWSPAPER

[...] We can be extremely proud of the current women winemakers in our industry, especially considering most of them are juggling a family along with the long working hours.

Unbalanced power relations
Compassion
The poorer, the merrier
Unbalanced power relations
Compassion

Task, data and annotation process

Dataset "Don't Patronize Me!" [Pérez-Almendros et al., 2020]

- 10.4K en paragraphs from the news of 20 countries
- All mention one of 10 selected vulnerable communities

Task setup Given an input paragraph P:

- **PCL identification**: identify whether P entails a PCL form
- **PCL classification**: determine PCL forms expressed by P

A closer look at the annotation Annotations by a_1 and a_2:

Annotation task	Individual decisions (a$_1$,a$_2$)	Score	Instances	Gold label
Subtask 1: "Does the paragraph contain any form of PCL?" Values: 0, 1, 2	(0,0)	0	8,529	No
	(0,1), (1,0), *	1	947	
	(1,1), *	2	144	
	(2,1), (1,2), *	3	458	Yes
	(2,2)	4	391	
Subtask 2: "Which PCL category does the span express (if any)?" Values: $c_i, c_j \in C$, NONE	$(c_i, \text{NONE}), (\text{NONE}, c_i)$	1	1,359	c_i
	$(c_i, c_j)_{c_i \neq c_j}, (c_j, c_i)_{c_j \neq c_i}$			c_i, c_j
	(c_i, c_i)	2	1,401	c_i

💡 **Idea** Raw "score" values can be leveraged to capture different shades of PCL based on **annotators' interpretation and sensibility**, thus modeling their **uncertainty** and **disagreement** in detecting PCL

References

Don't Patronize Me! An Annotated Dataset with Patronizing and Condescending Language towards Vulnerable Communities (Perez Almendros et al., COLING 2020)

Massive Choice, Ample Tasks (MaChAmp): A Toolkit for Multi-task Learning in NLP (van der Goot et al., EACL 2021)

Methods

Based on MaChAmp [van der Goot et al., 2021]

General framework Multi-task learning with shared encoder

- Main task decoder: for the end task (e.g., subtask 1)
- Auxiliary task decoder(s): for providing useful signals

Multiple views Different forms (or views) of the dataset

- **Paragraph data view** (D_P): dataset in its standard form
- **Span data view** (D_S): dataset consisting of all PCL-expressing spans from D_P

Auxiliary tasks and associated data views

- *uncertainty*: labels: {0,1,2,3,4}, view: D_P – subtask 1
- *agreement*: labels: {1,2}, view: D_S – subtask 2
- *span*: labels: {UNB, SHA, PRE, …}, view: D_S – subtask 1, 2
- *country*: labels: {au, bd, ca, gb, …}, view: D_P – subtask 1, 2

Models Leverage annotators' uncertainty & disagreement

(1) MTMW(UNC+SPAN) *model for subtask 1*

(2) MTMW(AGR+COU+SPAN) *model for subtask 2*
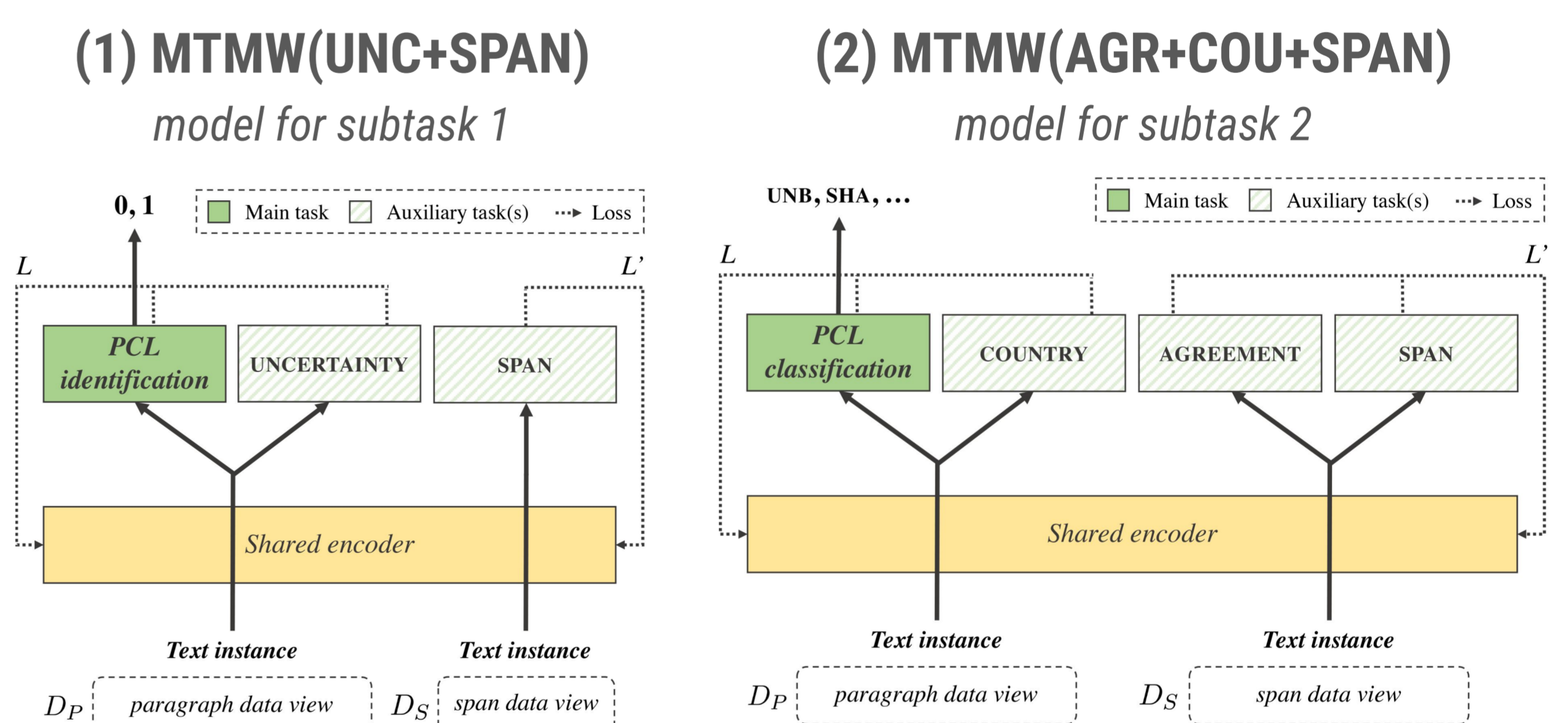
(3) SEQ. FINE-TUNING: On subtask 1, then 2 – *model for subtask 1 and 2*

Results & analysis

Test set results ↓ and →

PCL identification

	P	R	F$_1$
Organizers' baseline	39.35	65.30	49.11
MTMW(UNC+SPAN)	64.23	52.68	**57.89**
SEQ. FINE-TUNING	53.99	55.52	54.74

PCL classification

	UNB	SHA	PRE	AUT	MET	COM	THE	F$_1$
Organizers' baseline	35.35	0.00	16.67	0.00	0.00	20.87	0.00	10.41
MTMW(AGR+COU+SPAN)	52.46	36.22	26.95	**37.71**	**31.86**	**45.95**	**30.30**	**37.35**
SEQ. FINE-TUNING	**54.00**	**46.73**	**28.07**	22.22	29.73	44.28	20.69	35.10

Analysis Contribution of aux tasks and role of disagreement

- **Aux tasks** Subtask 1: *uncertainty* consistently improves results; Subtask 2: *agreement* is orthogonal to *country*
- **Uncert/Disagr** Subtask 1: *uncertainty* worsen results; Subtask 2: instances w/ disagreement are more difficult

Conclusion

- **Towards annotators-centric NLP** for subjective tasks
- **Competitive results**, without external data or ensembles