Detecting the use of Propaganda in the News

Giovanni Da San Martino

Work in collaboration with Israa Jaradat, Seunghak Yu, Alberto Barròn-Cedeño, Preslav Nakov

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Why Propaganda?

- “Expression **deliberately** designed to **influence** the opinions/actions of other individuals or groups with reference to predetermined ends.”

Institute for Propaganda Analysis

Computational Propaganda

- “The rise of the Internet […] has opened the **creation and dissemination of propaganda messages**, which were once the province of states and large institutions, to **a wide variety of individuals and groups**.”

(Bolsover and Howard, Big Data 5(4))

audience targeting

Bot armies

persuasive messages

anonymity

efficient dissemination of data

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Propaganda Analysis at Document-level

Supervised model to compute a propagandist index: the likelihood of a text to contain propagandistic mechanisms to deliberately influence the reader's opinion.

Information Processing & Management
Volume 56, Issue 5, September 2019, Pages 1849-1864

Proppy: Organizing the news based on their propagandistic content

Alberto Barrón-Cedeño [1, a], Israa Jaradat [1, b], Giovanni Da San Martino [c], Preslav Nakov [c]

Show more

https://doi.org/10.1016/j.ipm.2019.03.005

Get rights and content

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Related work

- **Aim**: differentiating real news from satire, hoaxes, and propaganda.
- **Corpus**: ~22K documents from the English Gigaword (real news) and from seven unreliable news sites.
- **Representation**: word n-grams, with $n \in [1, 3]$.
- **Model**: max entropy with L_2 regularization.

(Rashkin, et al., EMNLP 2017)

TSHP-17 Corpus (Rashkin, et al., EMNLP 2017)

kind	sources	articles	training	dev	test	length (tokens)
Trusted	4*	5,750	3,997	1,003	750	522 ± 429.13
Satire	3	5,750	3,981	1,019	750	324 ± 276.31
Hoax	2	5,750	4,014	986	750	262 ± 300.92
Propaganda	2	5,330	3,670	910	750	$1{,}047 \pm 1{,}156.87$
Total	**11**	**22,580**	**15,662**	**3,918**	**3,000**	$\mathbf{529 \pm 705.34}$

Sources	Trusted	Gigaword News*
	Satire	The Onion ● The Borowitz Report ● Clickhole
	Hoax	American News ● DC Gazette
	Propaganda	The Natural News ● Activist Report

- Gold labels obtained by distant supervision

- Representation:
  word n-grams
- In-domain data (dev):
  - **F1**: 94.48
  - **Accuracy** : 94.44
- Out-of-domain data (test):
  - **F1**: 69.26
  - **Accuracy**: 69.73

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Hypothesis

- The topic of a document and its topic-specific vocabulary are not relevant factors to decide whether it is propagandist or not.

- Representations based on writing style and complexity can generalize better than current approaches based on word-level representations

Proppy: features

1. Lexical features

Lexicon	Sample words
Wiktionary	
Modal, Action, Manner Adverbs	truly, apparently, accidentally, deliberately
Comparative, Superlative Forms	higher, less, purest, worst
LIWC	
First Person Singular, Second Person	my, I, you, yours
Hear, Money, Negation, Number, See, Sexual, Swear	says, costs, can't, quarter, watch, gay, dumb
Strong/Weak Subjectives (Wilson)	anti-semites, extremist
Hedges (Hyland)	appears, approximately, perhaps
Assertives (Hooper)	admit, hypothesize, certain

For each of the lexicons, the total number of words in the article is a feature

Proppy: features

2. Vocabulary richness features

feature	computation
TTR. Type–token ratio	$\|types\|/\|tokens\|$
Hapax legomena. Amount of tokens appearing once in a text.	$\|types_i\|$
Hapax dislegomena. Amount of tokens appearing twice in a text	$\|types_j\|$
Honore's R. Combination of types, tokens, and hapax legomenæ.	$\frac{100 \cdot log(\|tokens\|)}{1-\|hapax\_legomena\|/\|types\|}$.
Yule's characteristic K. Combination of types appearing with different frequencies and tokens. The chance of a word to occur in a text to follow a Poisson distribution	$10^4 \frac{\sum_i i^2 \|types_k\| - \|tokens\|}{\|tokens\|^2}$

where $i = 1$, $j = 2$, and $k = [1, 2, \ldots]$ are the different frequencies of types in the text.

Proppy: features

3. Readability features

feature	computation
Flesch–Kincaid grade level. US grade level necessary to understand a text.	$0.39 \cdot \frac{\lvert tokens \rvert}{\lvert syllables \rvert} + 11.9 \cdot \frac{\lvert syllables \rvert}{\lvert tokens \rvert} - 15.59$
Flesch reading ease. A scale in range $[0, 100]$ representing the complexity of a text. The latter is the easiest	$206.835 - 1.015 \cdot \frac{\lvert tokens \rvert}{\lvert sentences \rvert} - 84.6 \cdot \frac{\lvert syllables \rvert}{\lvert tokens \rvert}$
Gunning fog index. Amount of the years of formal education necessary to understand a text.	$0.4 \left(\frac{\lvert tokens \rvert}{\lvert sentences \rvert} + 100 \cdot \frac{\lvert tokens_c \rvert}{\lvert tokens \rvert} \right)$

$tokens_c$ stands for complex tokens; those with three syllables or more.

Proppy: features

4. Style features:
   - TF-IDF weighted Character 3-grams to capture different style markers, such as prefixes, suffixes, and punctuation marks.

Proppy: features

5. NELA* features :
   - **Structure** : POS counts, linguistic (LIWC), clickbaits (Chakraborty et al. 2016).
   - **Sentiment**: sentiment(Hutto and Gilbert 2014), emotion (Recasens et al. 2013) and (LIWC) , happiness (Mitchell et al. 2013).
   - **Topic-dependent**: bio, relativity, personal concerns (LIWC)
   - **Morality**: Moral (Haidt et al. 2009) and (Lin et al. 2017)
   - **Bias**: bias (Recasens et al. 2013) and (Mukherjee et al. 2015), subjectivity (Pang et al. 2004).

*(B. Horne, S. Khedr, S. Adal, "Sampling the news producers: A large news and feature data set for the study of the complex media landscape" AAAI-18)

Proppy: Corpus

- Qprop-18

Label	Sources	Articles	Train	Dev	Test	Length (tokens)
Propagandistic	10	5,737	4,021	575	1,141	1084.46 ± 890.59
Non-propagandistic	94	45,557	31,972	4,564	9,021	620.31 ± 518.92
Total	**104**	**51,294**	**35,993**	**5,139**	**10,162**	672.22 ± 590.98

- Collected using GDELT + MBFC

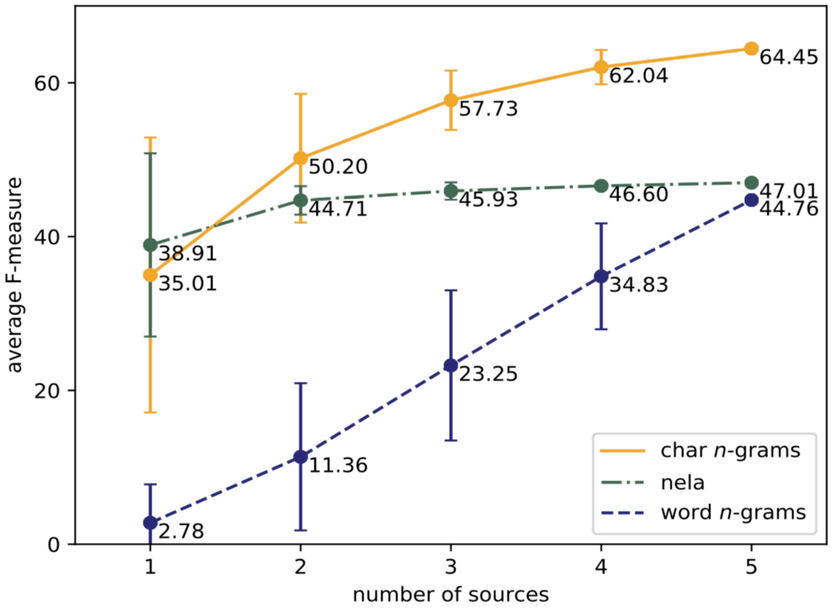
Experiment 1: Two-Class Classification on TSHP-17 and QProp-18

	TSHP-17	QProp	
Features	in-domain	Dev	Test
word *n*-grams	90.76	74.42	75.55
lexicon	68.74	46.55	44.87
voc. richness	55.62	29.45	29.72
readability	40.16	21.96	21.50
char *n*-grams	**96.22**	**82.93**	**82.13**
nela	82.27	54.60	50.98
word *n*-grams + char *n*-grams	**97.21**	78.37	79.01
char *n*-grams + lexicon	97.14	83.02	81.94
char *n*-grams + nela	96.64	**83.21**	82.75
readability + nela	82.30	75.34	76.83
char *n*-grams + lexicon + voc. richness + nela	96.97	83.17	**82.89**
word & char *n*-grams + lexicon + voc. richness + nela	97.10	79.04	79.50

Features	TSHP-17 corpus out-of-domain
word *n*-grams	50.68
lexicon	61.54
voc. richness	54.29
readability	45.68
char *n*-grams	52.51
nela	64.00
word *n*-grams + char *n*-grams	63.66
char *n*-grams + lexicon	52.89
char *n*-grams + nela	53.66
readability + nela	53.66
char *n*-grams + lexicon + voc. richness + nela	**64.14**
word & char *n*-grams + lexicon + voc. richness + nela	63.47

Experiment 2: Learning Propaganda vs. Learning the Source

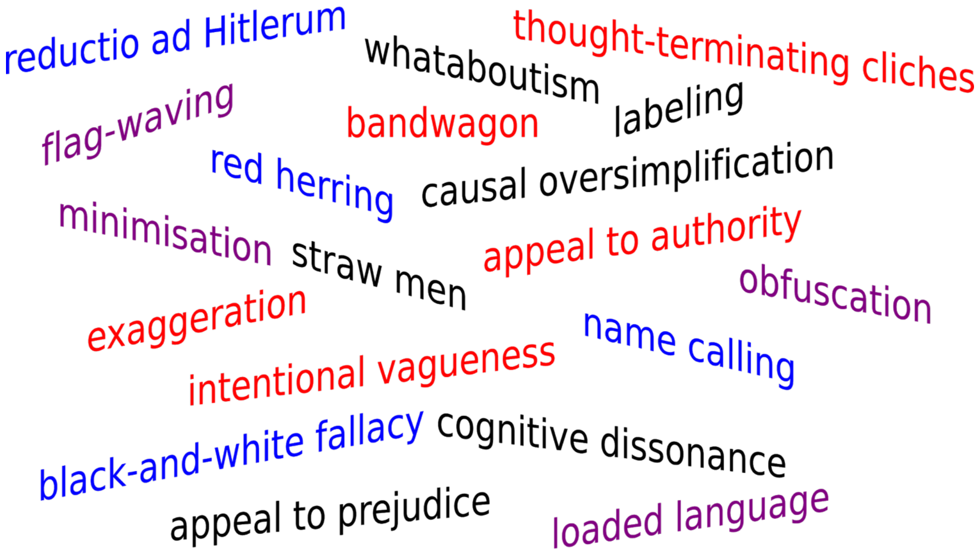
Test set (fixed): selected all examples from 5 propagandistic sources
Training: randomly selecting n propagandistic sources, random sampling the non-propagandistic ones such that the distribution is similar to the one of the full dataset



QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Fine-Grained Propaganda Analysis

- Proppy is not able to provide explanations for its scores
- Distant supervision is problematic, but avoiding it by labeling each article is not feasible
- We tackle the problem from a different angle
- Propaganda is conveyed through a series of rhetorical and psychological techniques



QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY


Weather: Sunny and warm, 79/62 SPORTS ★ FINAL Monday, September 24, 2007

DAILY NEWS

50¢ NEW YORK'S HOMETOWN NEWSPAPER NYDailyNews.com

THE EVIL HAS LANDED

Mahmoud Ahmadinejad

STORIES ON PAGES 4-7
EDITORIAL PAGE 22

Weather: Sunny and warm, 79/62 SPORTS ★ FINAL Monday, September 24, 2007

DAILY NEWS

50¢ New York's Hometown Newspaper NYDailyNews.com

THE EVIL HAS LANDED

Mahmoud Ahmadinejad

STORIES ON PAGES 4-7
EDITORIAL PAGE 22

Name Calling

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY


WOULD YOU VOTE FOR
CLINTON AS PRESIDENT?
57%
SAY YES

Bandwagon: “Attempting to persuade the target audience to join in and take the course of action because "everyone else is taking the same action".

“We are in the middle of the sixth mass extinction, with more than 200 species getting extinct every day"

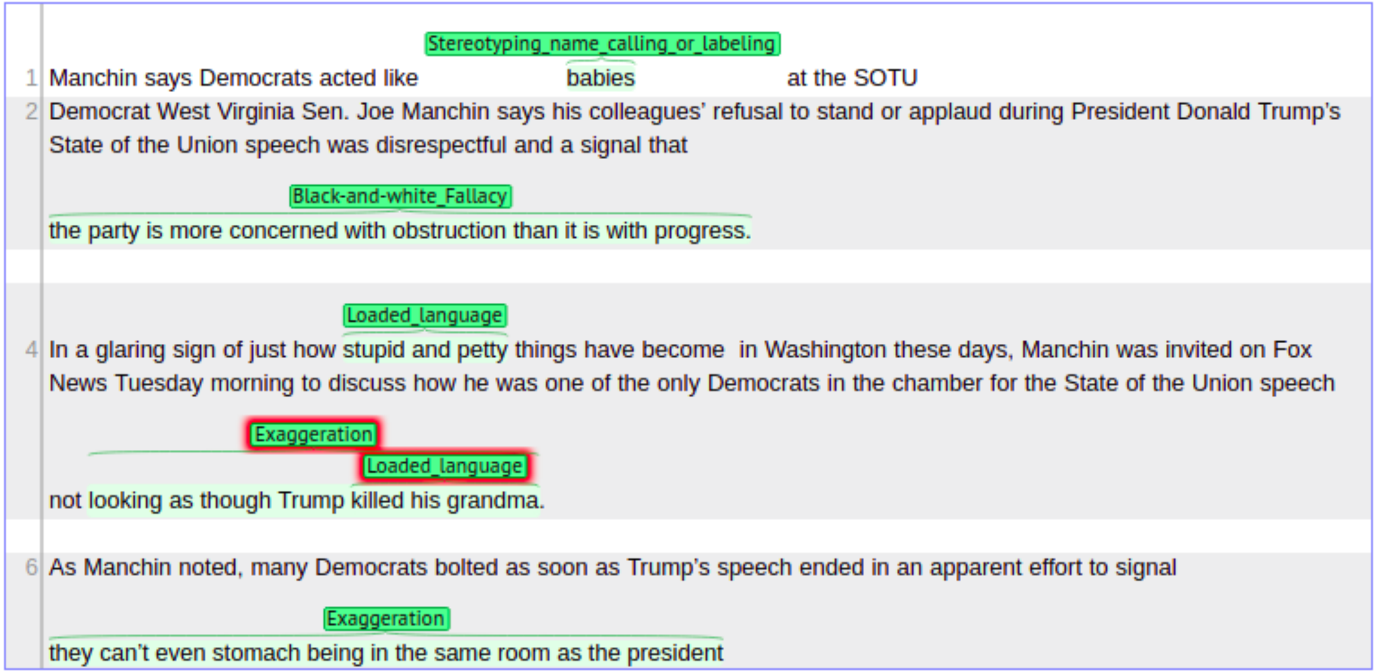
Greta Thunberg

"We are in the middle of the sixth mass extinction, with more than 200 species getting extinct every day"

Greta Thunberg

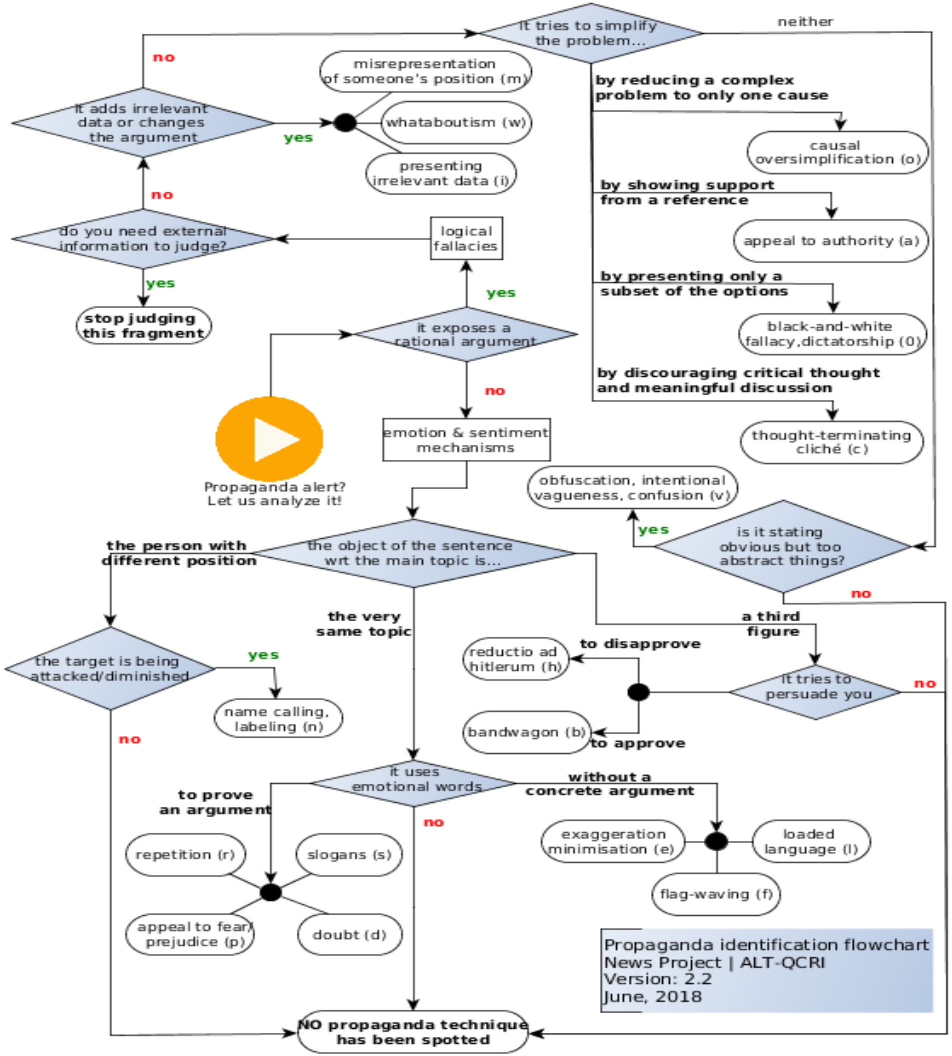
Appeal to Fear

Propaganda Techniques Corpus



We created a new dataset with 18 techniques annotated at fragment level (450 articles from 48 sources, 350k words, 400 man hours for annotating it)

- Articles are annotated at fragment level by experts
- Annotators choose between 18 techniques for a fragment



QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Annotation Process

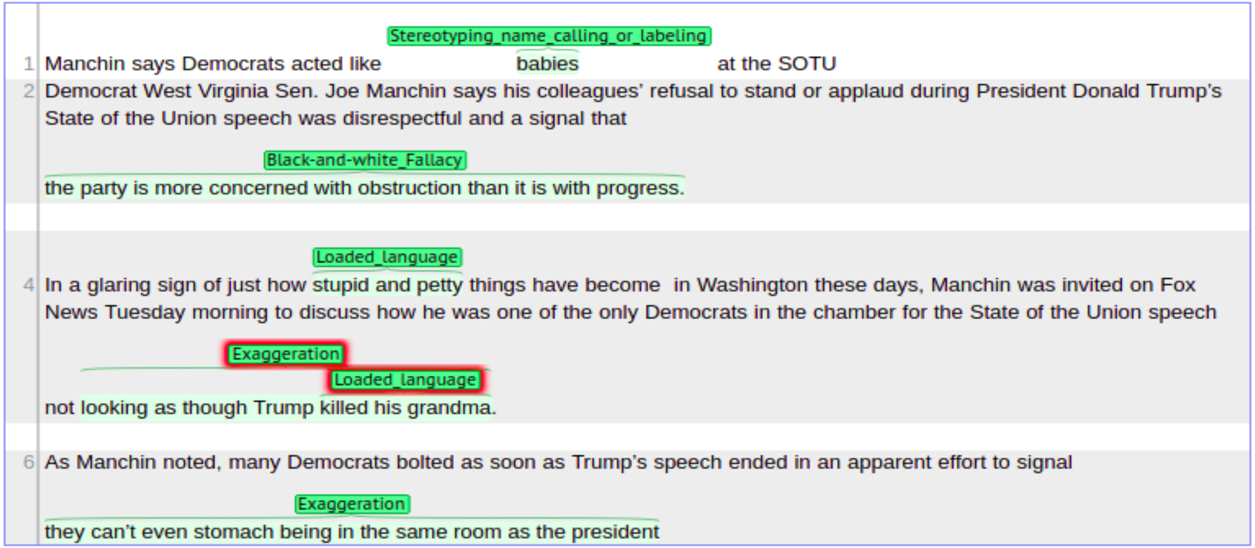
- Phase 1: two annotators, a_i, a_j, independe ntly annotate the same article
- Phase 2: they gather with a consolidator c_k to discuss all instances and to come up with a final annotation.

Annotations		spans (γ_s)	+labels (γ_{sl})
a_1	a_2	0.30	0.24
a_3	a_4	0.34	0.28
a_1	c_1	0.58	0.54
a_2	c_1	0.74	0.72
a_3	c_2	0.76	0.74
a_4	c_2	0.42	0.39

Propaganda Technique	**inst**	**avg. length**
loaded language	2,547	23.70 ± 25.30
name calling, labeling	1,294	26.10 ± 19.88
repetition	767	16.90 ± 18.92
exaggeration, minimization	571	45.36 ± 35.55
doubt	562	123.21 ± 97.65
appeal to fear/prejudice	367	93.56 ± 74.59
flag-waving	330	61.88 ± 68.61
causal oversimplification	233	121.03 ± 71.66
slogans	172	25.30 ± 13.49
appeal to authority	169	131.23 ± 123.2
black-and-white fallacy	134	98.42 ± 73.66
thought-terminating cliches	95	34.85 ± 29.28
whataboutism	76	120.93 ± 69.62
reductio ad hitlerum	66	94.58 ± 64.16
red herring	48	63.79 ± 61.63
bandwagon	17	100.29 ± 97.05
obfusc., int. vagueness, confusion	17	107.88 ± 86.74
straw man	15	79.13 ± 50.72
all	7,485	46.99 ± 61.45

Tasks

- **FLC** - detect the text-fragments in which a propaganda technique is used and identify the technique.
- **Spans** is a lighter version of the task in which only the span has to be identified.
- **SLC** a binary task at sentence-level: a sentence is considered as propagandistic if it contains one or more propagandistic fragments.

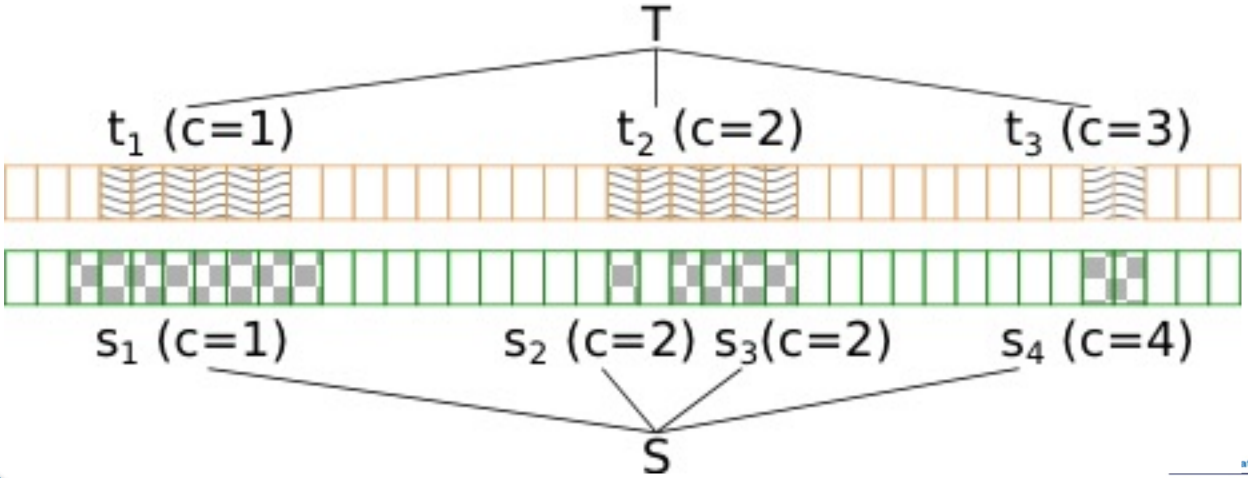


1	propaganda
2	non-propaganda
3	propaganda
4	propaganda
5	non-propaganda
6	non-propaganda

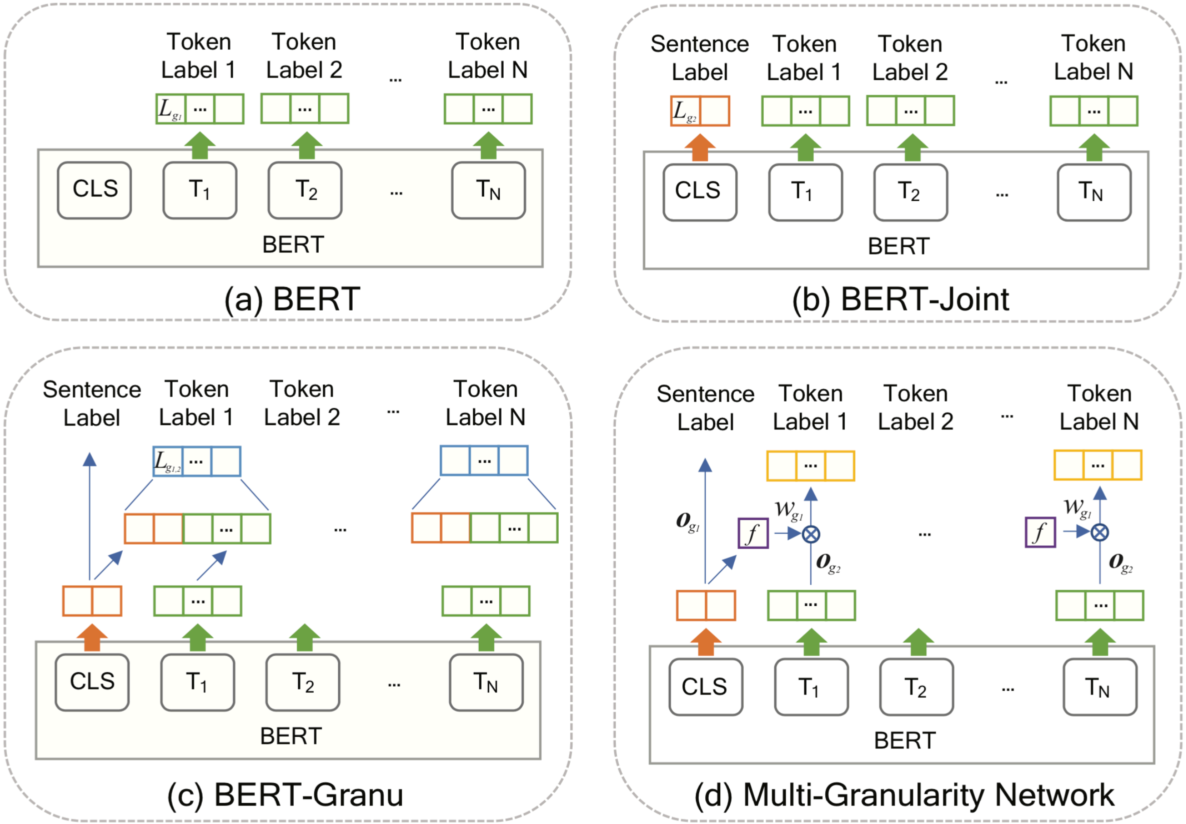
Evaluation Measures

- **SLC:** standard F_1 measure
- **FLC** - we adapted a measure for NER to account for overlapping gold spans

$$C(s,t,h) = \frac{|(s \cap t)|}{h}\delta\left(l(s), l(t)\right), \quad P(S,T) = \frac{1}{|S|}\sum_{\substack{s \in S,\\ t \in T}} C(s,t,|s|), \quad R(S,T) = \frac{1}{|T|}\sum_{\substack{s \in S,\\ t \in T}} C(s,t,|t|),$$



Models



Token Label 1
Token Label 2
...
Token Label N
CLS
T1
T2
TN
BERT
(a) BERT
Sentence Label
Token Label 1
Token Label 2
Token Label N
BERT
(b) BERT-Joint
Sentence Label
Token Label 1
Token Label 2
Token Label N
BERT
(c) BERT-Granu
Sentence Label
Token Label 1
Token Label 2
Token Label N
f
BERT
(d) Multi-Granularity Network

Results: Fragment-Level

Model	Spans P	Spans R	Spans F_1	Full Task P	Full Task R	Full Task F_1
BERT	39.57	36.42	37.90	21.48	**21.39**	21.39
Joint	39.26	35.48	37.25	20.11	19.74	19.92
Granu	43.08	33.98	37.93	23.85	20.14	21.80
Multi-Granularity						
ReLU	43.29	34.74	38.28	23.98	20.33	21.82
Sigmoid	**44.12**	**35.01**	**38.98**	**24.42**	21.05	**22.58**

Results: Sentence-Level

Model	Precision	Recall	F1
All-Propaganda	23.92	1.00	38.61
BERT	**63.20**	53.16	57.74
BERT-Granu	62.80	55.24	58.76
BERT-Joint	62.84	55.46	58.91
MGN Sigmoid	62.27	59.56	60.71
MGN ReLU	60.41	**61.58**	**60.98**

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, Preslav Nakov
Fine-Grained Analysis of Propaganda in News Articles. EMNLP 2019

Interested in the Task?

TASK | RULES/DATES | REGISTER | LEADERBOARD | ORGANISERS | Passcode | Team Page

SEMEVAL 2020 TASK 11
"DETECTION OF PROPAGANDA TECHNIQUES IN NEWS ARTICLES"

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY