A Corpus for Sentence-Level Subjectivity Detection on English News Articles

F. Antici, A. Galassi, F. Ruggeri, K. Korre, A. Muti, A. Bardi, A. Fedotova, A. Barrón-Cedeño

LT NLP LAB @ DISI DIT — UNIVERSITY OF BOLOGNA

MAIN CONFERENCE @ LREC-COLING 2024

Subjectivity Detection

Past Approaches

- Domain- and language-specific **lexicons**;
- Annotation guidelines **but not satisfying agreement** measures;
- **Machine translation** to transfer their approach across languages.

Challenges

- Subjectivity is subjective!
  - Interpretations of language;
  - Background knowledge and interpretation biases;
- Creating corpora is notably difficult and costly.

Contribution

Approach

- Information retrieval-oriented perspective
  - Objective: sentences from which information can be directly extracted
  - Subjective: sentences that must be further processed

- **Prescriptive paradigm [1]**:
  - Shared perspective between annotators
  - Discussions on disagreements
  - Rules grounded in examples, especially for controversial cases

NewsSD-ENG Corpus

- High-quality corpus concerning news articles in English
- 23 articles
- 1.049 sentences
- Curated human annotation
- At least 2 annotators per sentence

Experimental Results

We experiment with two languages to evaluate the applicability of our guidelines.

English: our corpus NewsSD-ENG
Italian: NewsSD-ITA, a re-annotated version of SubjectivITA [2].

Model	English Test Set Macro	OBJ	SUBJ	Italian Test Set Macro	OBJ	SUBJ
monolingual	en→en			it→it		
MAJ-B	0.33	0.65	0.00	0.42	0.85	0.00
RND-B	0.50	0.49	0.50	0.47	0.58	0.36
SVM	0.44	0.64	0.24	0.59	0.85	0.34
LR	0.55	0.63	0.48	0.60	0.77	0.42
M-SBERT	0.69	0.70	0.69	0.69	0.82	0.56
M-BERT	0.75	0.77	0.71	0.74	**0.88**	0.59

Monolingual - Takeaways

M-BERT is the **best** performing model.

M-BERT and M-SBERT have **comparable** performance concerning F1-SUBJ.

Model	English Test Set Macro	OBJ	SUBJ	Italian Test Set Macro	OBJ	SUBJ
monolingual	en→en			it→it		
MAJ-B	0.33	0.65	0.00	0.42	0.85	0.00
RND-B	0.50	0.49	0.50	0.47	0.58	0.36
SVM	0.44	0.64	0.24	0.59	0.85	0.34
LR	0.55	0.63	0.48	0.60	0.77	0.42
M-SBERT	0.69	0.70	0.69	0.69	0.82	0.56
M-BERT	0.75	0.77	0.71	0.74	**0.88**	0.59
multilingual	en+it→en			en+it→it		
SVM	0.49	0.63	0.34	0.60	0.85	0.35
LR	0.64	0.63	0.65	0.61	0.81	0.42
M-SBERT	0.71	0.67	0.76	0.69	0.81	0.56
M-BERT	**0.80**	**0.81**	**0.80**	**0.77**	**0.88**	**0.66**

Multilingual - Takeaways

Notable performance **improvement** for M-SBERT and M-BERT.

M-BERT is the **best** overall model.

These results suggest that the two corpora are **coherent in their annotation**.

Model	English Test Set Macro	OBJ	SUBJ	Italian Test Set Macro	OBJ	SUBJ
monolingual	en→en			it→it		
MAJ-B	0.33	0.65	0.00	0.42	0.85	0.00
RND-B	0.50	0.49	0.50	0.47	0.58	0.36
SVM	0.44	0.64	0.24	0.59	0.85	0.34
LR	0.55	0.63	0.48	0.60	0.77	0.42
M-SBERT	0.69	0.70	0.69	0.69	0.82	0.56
M-BERT	0.75	0.77	0.71	0.74	**0.88**	0.59
crosslingual	it→en			en→it		
M-SBERT	0.67	0.61	0.74	0.66	0.83	0.49
M-BERT	0.60	0.72	0.46	0.65	0.85	0.46

Cross-lingual - Takeaways

Notable performance **drop** for M-BERT compared to the **monolingual** setting.

In contrast, M-SBERT achieves **comparable** performance in both corpora.

Findings & Future Work

Methodology

- Yields high-quality annotations for SD
- Our corpus as demonstration
- Transferable to other languages with minor efforts (*examples*)

Language Transfer

- Best performance in multilingual setting (Ita + Eng)

Future Work

- Annotation of news articles in other languages
- Application to other domains and tasks (e.g., claim verification)

Code

Paper

Annotation Process

Seven annotator with near-to-native English knowledge annotate in three stages:

Two pilot studies

- **First:** to discover edge cases and refine guidelines
  - IAA **0.40**, "fair agreement"
- **Second:** context vs contextless annotation
  - with context: IAA 0.38
  - without context: IAA **0.53**, "moderate agreement"'

Final annotation

- Two annotators per sentence (IAA 0.53)
- Discussion on disagreements (IAA 0.83)
- Third annotator for residual disagreement (**< 10%** of cases).

Guidelines

We define our initial guidelines based on those presented in Antici et al., 2021 [2].

Subjective (SUBJ)

Opinion – Explicitly reports the personal opinion of its author.

It has everything you could want in a holiday: beautiful sandy beaches, delicious food, the Greek salads are simply on another level [...].

Sarcasm – Contains sarcastic or ironic expressions attributable to its author.

It's no lie that the USA is one heck of a big country (said in a southern twang).

Auspice – Contains exhortations or personal auspices made by its author.

The West should arm Ukraine faster

Downgrade – Contains discriminating or downgrading expression made by its author.

How did we reach the stage where priests and bishops cowered like frightened puppies before a common flu [...]

Rhetoric – Contains rhetorical figures explicitly made by its author to convey their opinion.

Barcelona where it all began, Messi was a king in Catalonia and he lived like one too.

Objective (OBJ)

Factual – Reports on news or historical facts that are quoted by the author of the sentence.

In the modern era electroconvulsive therapy, first used in 1938, became a treatment for some serious forms of depression in the post-war decades.'

Emotion – Describes the personal feelings, emotions or moods of the writer, without conveying opinions on other matters.

I was definitely surprised at how emotional I felt watching the service.

Third-party – Expresses a point of view explicitly attributable to a third-party.

Frank Drake believed that the universe had to contain other intelligent beings.

Note: The presence of quotation marks (" "), when used to quote a third person, represents an explicit third-party opinion, even if it is not clearly stated in the sentence.

"Crosbie is an extremely violent man who has no place in society, and we welcome the jury's verdict today."

No conclusion – Contains an author's comment that does not draw any conclusion.

It is not clear yet which of the couples from the E4 reality show remain together and who have now, because the series has not concluded.

No stance – Contains factual conclusions that do not convey any stance or personal opinion.

In years gone by, travel to Japan was notoriously expensive, but the devaluing of the yen has made it more accessible.

Titles – Any kind of well-known nickname or title is considered objective.

The Duke of York 'plotted' with Diana to 'push Prince Charles aside'.

Expressions – Any kind of common expression or proverb is considered objective.

Home sweet home: George poses in one of the rooms at his sprawling Hampstead home during a photoshoot in 2002.

Ethics, Limitations, and Risks

Annotation Bias

- Impossible to exclude that annotators may share similar biases

Data Selection

- Some topics may be more frequently addressed in SUBJ sentences
- Cannot exclude a model learning a certain bias [3]

Corpus Size

- Limited size is symptomatic of the challenging scenario.

Language Adaptation

- Challenges in adapting guidelines to linguistically diverse contexts
- May affect how an annotator perceives the annotation
- Adapting to a new language mainly requires changing the examples

References

[1]. P. Rottger, B. Vidgen, D. Hovy, J. Pierrehumber, 2022, Two contrasting data annotation paradigms for subjective NLP tasks, NAACL 2022.

[2] F. Antici, L. Bolognini, M. A. Inajetovic, B. Ivasiuk, A. Galassi, F. Ruggeri, 2021, SubjectivITA: An italian corpus for subjectivity detection in newspapers, CLEF 2021.

[3]. M. Wiegand, J. Ruppenhofer, T. Kleinbauer, 20219, Detection of Abusive Language: the Problem of Biased Datasets, NAACL 2019.