

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

F. Antici<sup>1</sup>, A. Galassi<sup>1</sup>, **F. Ruggeri**<sup>1</sup>, K. Korre<sup>2</sup>, A. Muti<sup>2</sup>, A.  
Bardi<sup>2</sup>, A. Fedotova<sup>2</sup>, and A. Barrón-Cedeño<sup>2</sup>

<sup>1</sup>DISI, University of Bologna, Italy

<sup>2</sup>DIT, University of Bologna, Forlì, Italy

# Background

Subjectivity Detection (**SD**) contributes to several NLP tasks: sentiment analysis  
bias detection  
fact-checking

... [Riloff and Wiebe, 2003]

# Background

Subjectivity Detection (**SD**) contributes to several NLP tasks: sentiment analysis

bias detection

fact-checking

**However**, spotting subjectivity is **challenging**:

- Different interpretations of language
- Background knowledge
- Personal biases

... [Riloff and Wiebe, 2003]

# Background

Subjectivity Detection (**SD**) contributes to several NLP tasks: sentiment analysis

bias detection

fact-checking

**However**, spotting subjectivity is **challenging**:

- Different interpretations of language
- Background knowledge
- Personal biases

... [Riloff and Wiebe, 2003]

**Consequently**, creating corpora for SD is notably **difficult** and **costly**.

# Background

Subjectivity Detection (**SD**) contributes to several NLP tasks: sentiment analysis

bias detection

fact-checking

**However**, spotting subjectivity is **challenging**:

- Different interpretations of language
- Background knowledge
- Personal biases

... [Riloff and Wiebe, 2003]

**Consequently**, creating corpora for SD is notably **difficult** and **costly**.

**Past approaches**:

- Relied on spotting subjectivity cues via domain- and language-specific **lexicons**.
- Relied on annotation guidelines but obtained not satisfying **agreement** measures.
- Relied on **machine translation** to transfer their approach to other languages.

## Contribution

We **develop** a novel set of annotation guidelines that may be applied to any language.

# Contribution

We **develop** a novel set of annotation guidelines that may be applied to any language.

We **design** an annotation procedure based on the prescriptive paradigm. [Röttger et al., 2022]

*Discourage annotator subjectivity by enforcing a shared belief,  
formulated in the annotation guidelines*

# Contribution

We **develop** a novel set of annotation guidelines that may be applied to any language.

We **design** an annotation procedure based on the prescriptive paradigm. [Röttger et al., 2022]

*Discourage annotator subjectivity by enforcing a shared belief,  
formulated in the annotation guidelines*

We **manually** curate NewsSD-ENG, a high-quality English corpus concerning news articles



## Contribution

We **develop** a novel set of annotation guidelines that may be applied to any language.

We **design** an annotation procedure based on the prescriptive paradigm. [Röttger et al., 2022]

|  
*Discourage annotator subjectivity by enforcing a shared belief,  
formulated in the annotation guidelines*

We **manually** curate NewsSD-ENG, a high-quality English corpus concerning news articles

## Experiments

We **evaluate** NewsSD-ENG by employing several models for SD.

## Contribution

We **develop** a novel set of annotation guidelines that may be applied to any language.

We **design** an annotation procedure based on the prescriptive paradigm. [Röttger et al., 2022]

*Discourage annotator subjectivity by enforcing a shared belief,  
formulated in the annotation guidelines*

We **manually** curate NewsSD-ENG, a high-quality English corpus concerning news articles

## Experiments

We **evaluate** NewsSD-ENG by employing several models for SD.

We **evaluate** the applicability of our guidelines in multilingual and cross-language settings (Ita, Eng).

## Guideline Refinement

We define our initial **guidelines** based on those presented in Antici et al. (2021).

Annotators label a sentence either subjective (SUBJ) or objective (OBJ).

We recruit 7 annotators with near-to-native English knowledge.

# Guideline Refinement

We define our initial **guidelines** based on those presented in Antici et al. (2021).

Annotators label a sentence either subjective (SUBJ) or objective (OBJ).

We recruit 7 annotators with near-to-native English knowledge.

## First pilot study

- **Objective:** discover edge cases and validate guidelines.
- **Data:** 3 news articles (270 sentences) from a pool of selected outlets.
- **Result:** The IAA measured as Krippendorff's alpha is **0.40** ("*fair agreement*").

# Guideline Refinement

## Second pilot study

- **Objective:** annotating with or without context by splitting annotators into two groups.
- **Data:** 70 sentences from 5 articles.
- **Result:** 0.38 IAA with context and **0.53** (“*moderate*”) without.

# Annotation Guidelines

## 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 [...].”*

# Annotation Guidelines

## 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)”*

# Annotation Guidelines

## 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”*



# Annotation Guidelines

**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 [...]”*

# Annotation Guidelines

**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 like hyperboles 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.”*

# Annotation Guidelines

## 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.”*

# Annotation Guidelines

## 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.”*

# Annotation Guidelines

## 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.”*

# Annotation Guidelines

**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."*

# Annotation Guidelines

**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.”*

# Annotation Guidelines

**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** – When referring to an individual, any kind of well-known nickname or title is considered objective.

*“The Duke of York ‘plotted’ with Diana to ‘push Prince Charles aside’.”*



# Annotation Guidelines

**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.”*

# Corpus

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

- (i) Two annotators **label** their assigned sentences. (**After:** IAA 0.53)
- (ii) Each pair **discusses** ambiguous sentences to reach an agreement. (**After:** IAA 0.83)
- (iii) In case of disagreement, a **third annotator** labels the disputed sentences. (< **10%** of cases)

# Corpus

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

- (i) Two annotators **label** their assigned sentences. (**After:** IAA 0.53)
- (ii) Each pair **discusses** ambiguous sentences to reach an agreement. (**After:** IAA 0.83)
- (iii) In case of disagreement, a **third annotator** labels the disputed sentences. (< **10%** of cases)

	# Art.	# Sent.	# OBJ	# SUBJ
Train	16	731	487 (12)	244 (46)
Dev	3	99	45 (3)	54 (8)
Test	4	219	106 (4)	113 (16)
Total	23	1,049	638 (19)	411 (70)

## Experimental Setting

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

**English:** our corpus NewsSD-ENG

[Antici et al., 2021] **Italian:** NewSD-ITA, a re-annotated version of SubjectivITA

# Experimental Setting

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

**English:** our corpus NewsSD-ENG

[Antici et al., 2021] **Italian:** NewSD-ITA, a re-annotated version of SubjectivITA

We consider the following classifiers:

- A support vector machine (**SVM**) and a logistic regressor (**LR**) with tf-idf features.
- A multilingual SentenceBERT model (**M-SBERT**).
- A multilingual BERT model (**M-BERT**).

# Experimental Setting

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

**English:** our corpus NewsSD-ENG

[Antici et al., 2021] **Italian:** NewSD-ITA, a re-annotated version of SubjectivITA

We consider the following classifiers:

- A support vector machine (**SVM**) and a logistic regressor (**LR**) with tf-idf features.
- A multilingual SentenceBERT model (**M-SBERT**).
- A multilingual BERT model (**M-BERT**).

We consider three binary classification (SUBJ/OBJ) settings:

- **Monolingual:** both training and test data are in the same language.
- **Multilingual:** the training data combines both considered languages training data.
- **Cross-lingual:** the training data is in a language and the test data is in another language.

# Results - Monolingual F1-score

Model	English Test Set			Italian Test Set		
	Macro	OBJ	SUBJ	Macro	OBJ	SUBJ
<i>monolingual</i>	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	<b>0.88</b>	0.59
<i>multilingual</i>	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	<b>0.80</b>	<b>0.81</b>	<b>0.80</b>	<b>0.77</b>	<b>0.88</b>	<b>0.66</b>
<i>crosslingual</i>	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

## Takeaways

M-BERT is the best performing model.

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

# Results - Multilingual F1-score

Model	English Test Set			Italian Test Set		
	Macro	OBJ	SUBJ	Macro	OBJ	SUBJ
<i>monolingual</i>	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	<b>0.88</b>	0.59
<i>multilingual</i>	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	<b>0.80</b>	<b>0.81</b>	<b>0.80</b>	<b>0.77</b>	<b>0.88</b>	<b>0.66</b>
<i>crosslingual</i>	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

## 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.



# Results - Cross-lingual F1-score

Model	English Test Set			Italian Test Set		
	Macro	OBJ	SUBJ	Macro	OBJ	SUBJ
<i>monolingual</i>	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	<b>0.69</b>	<b>0.70</b>	<b>0.69</b>	<b>0.69</b>	<b>0.82</b>	<b>0.56</b>
M-BERT	<b>0.75</b>	<b>0.77</b>	<b>0.71</b>	<b>0.74</b>	<b>0.88</b>	<b>0.59</b>
<i>multilingual</i>	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	<b>0.80</b>	<b>0.81</b>	<b>0.80</b>	<b>0.77</b>	<b>0.88</b>	<b>0.66</b>
<i>crosslingual</i>	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

## Takeaways

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

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

# *Findings*

## **Methodology**

Our corpus demonstrates that our methodology yield high-quality annotations for SD

## **Language transfer**

The best performance was achieved when training on articles both in English and Italian.

# ***Findings***

## **Methodology**

Our corpus demonstrates that our methodology yield high-quality annotations for SD

## **Language transfer**

The best performance was achieved when training on articles both in English and Italian.

# **Research Directions**

## **Other languages**

Annotation of news articles in other languages.

## **Other domains**

Application of our methodology to other domains and tasks like claim verification.

# Thanks for the attention!



Paper



Github

## References

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pages 105–112.

Francesco Antici, Luca Bolognini, Matteo Antonio Inajetovic, Bogdan Ivasiuk, Andrea Galassi, and Federico Ruggeri. 2021. Subjectivita: An italian corpus for subjectivity detection in newspapers. In CLEF, volume 12880 of LNCS, pages 40–52. Springer.

P. Röttger, B. Vidgen, D. Hovy, J. B. Pierrehumbert, Two contrasting data annotation paradigms for subjective NLP tasks, in: M. Carpuat, M. de Marneffe, I. V. M. Ruíz (Eds.), Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022, Association for Computational Linguistics, 2022, pp. 175–190.

# Supplementary Material

## Annotators' Profile

---

<b>Profile</b>	<b>Description</b>
Gender	3 identify as male, 4 as female
Education	3 PhD, 4 PhD candidates
Origin	4 Central Europe, 1 Eastern Europe, 1 Latin-America

---

## SubjectvITA corpus

	# Art.	# Sent.	# OBJ	# SUBJ
Train	80	1,399	1,079	320
Val	13	214	152	62
Test	10	227	167	60
Total	103	1,840	1,398	442