EVALITA 2018

EVALUATION OF NLP AND SPEECH TOOLS FOR ITALIAN

IronITA

Irony Detection in Italian Tweets

Alessandra Teresa Cignarella[1,2], **Simona Frenda**[1,2],
Valerio Basile[1], Cristina Bosco[1], Viviana Patti[2] and Paolo Rosso[2]

[1] Dipartimento di Informatica, Università degli Studi di Torino, Italy
[2] PRHLT Research Center, Universitat Politècnica de València, Spain

Introduction & Motivation

CONTINUITY WITH:

- SENTIPOLC @ EVALITA 2014 (Barbieri et al., 2014)
- SENTIPOLC @ EVALITA 2016 (Basile et al., 2016)

Introduction & Motivation

CONTINUITY WITH:

- SENTIPOLC @ EVALITA 2014 (Barbieri et al., 2014)
- SENTIPOLC @ EVALITA 2016 (Basile et al., 2016)

FOCUS ON:

- Italian social media texts (Twitter)

Introduction & Motivation

irony

noun [C/U] • US /ˈɑɪ·rə·ni, ˈɑɪ·ər·ni/

a type of usually humorous expression in which you say the opposite of what you intend:

[U] *He had a powerful sense of irony, and you could never be absolutely sure when he was serious.*

Introduction & Motivation

irony

noun [C/U] • US /ˈɑɪ·rə·ni, ˈɑɪ·ər·ni/

- **a type of usually humorous expression in which you say the opposite of what you intend:**

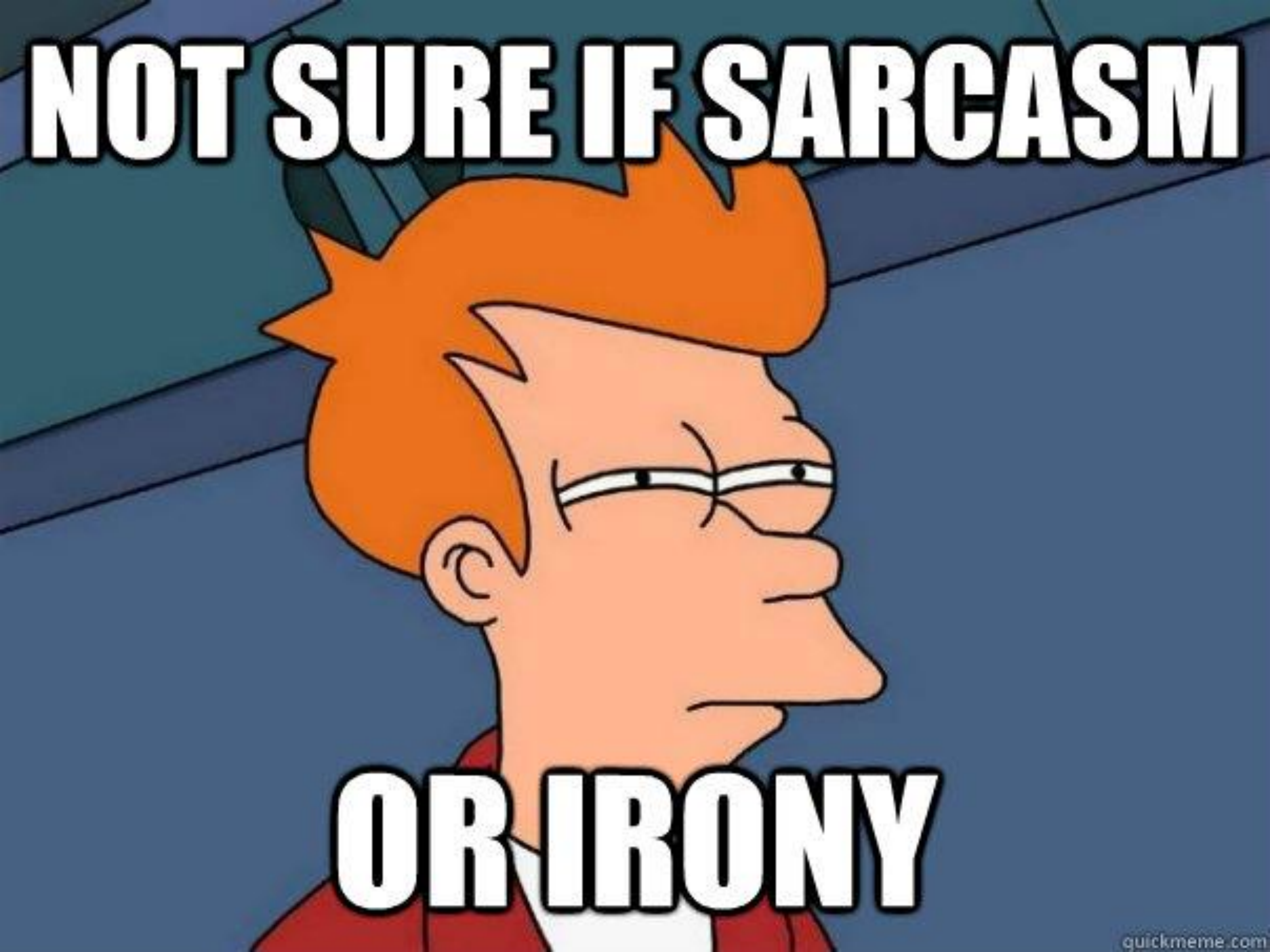
  [U] *He had a powerful sense of irony, and you could never be absolutely sure when he was serious.*

sarcasm

noun [U] • UK /ˈsɑː.kæz.ᵊm/ US /ˈsɑːr.kæz.ᵊm/

- **the use of remarks that clearly mean the opposite of what they say, made in order to hurt someone's feelings or to criticize something in a humorous way:**

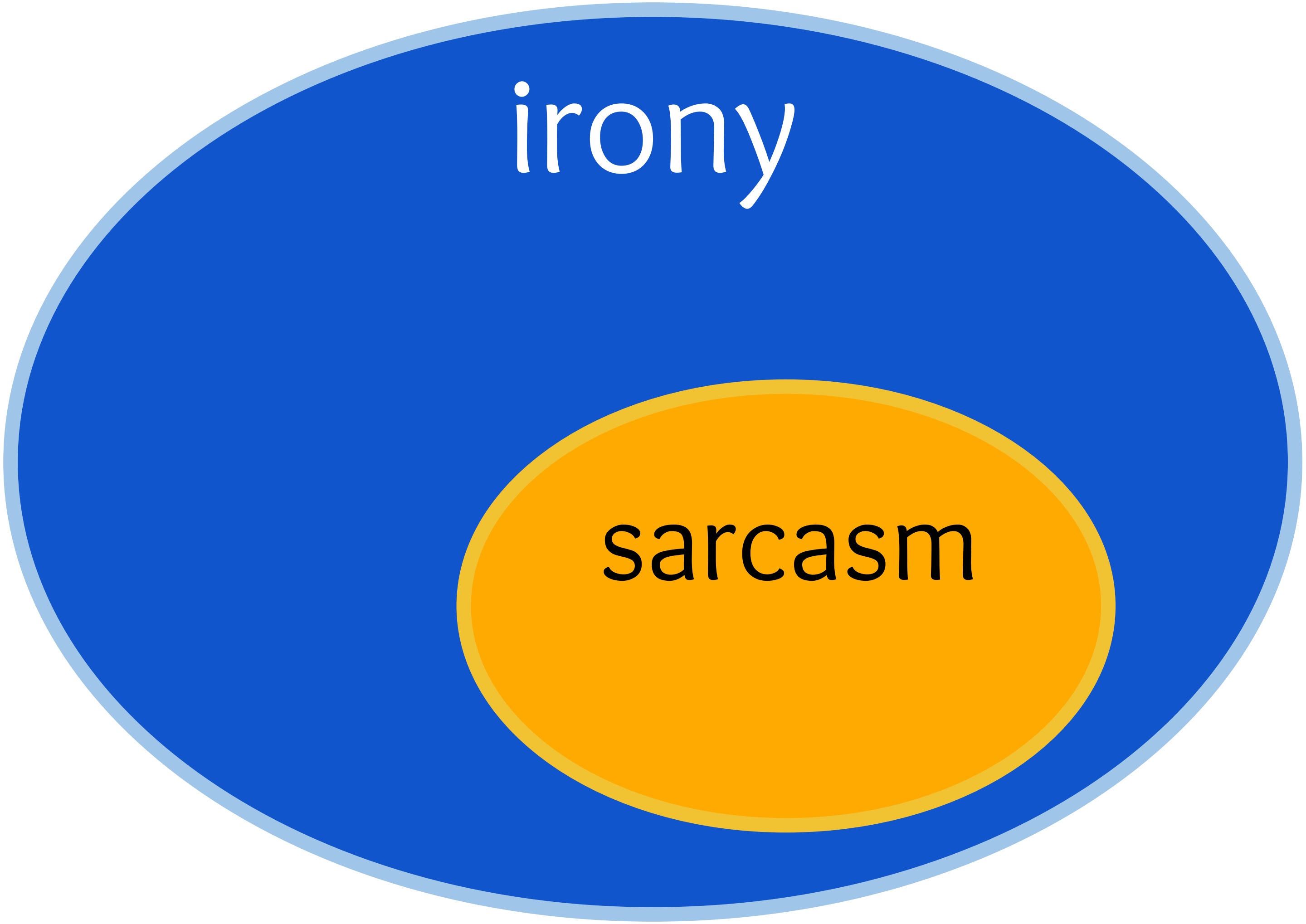
  "You have been working hard," he said with **heavy** *sarcasm, as he looked at the empty page.*


NOT SURE IF SARCASM
OR IRONY
quickmeme.com

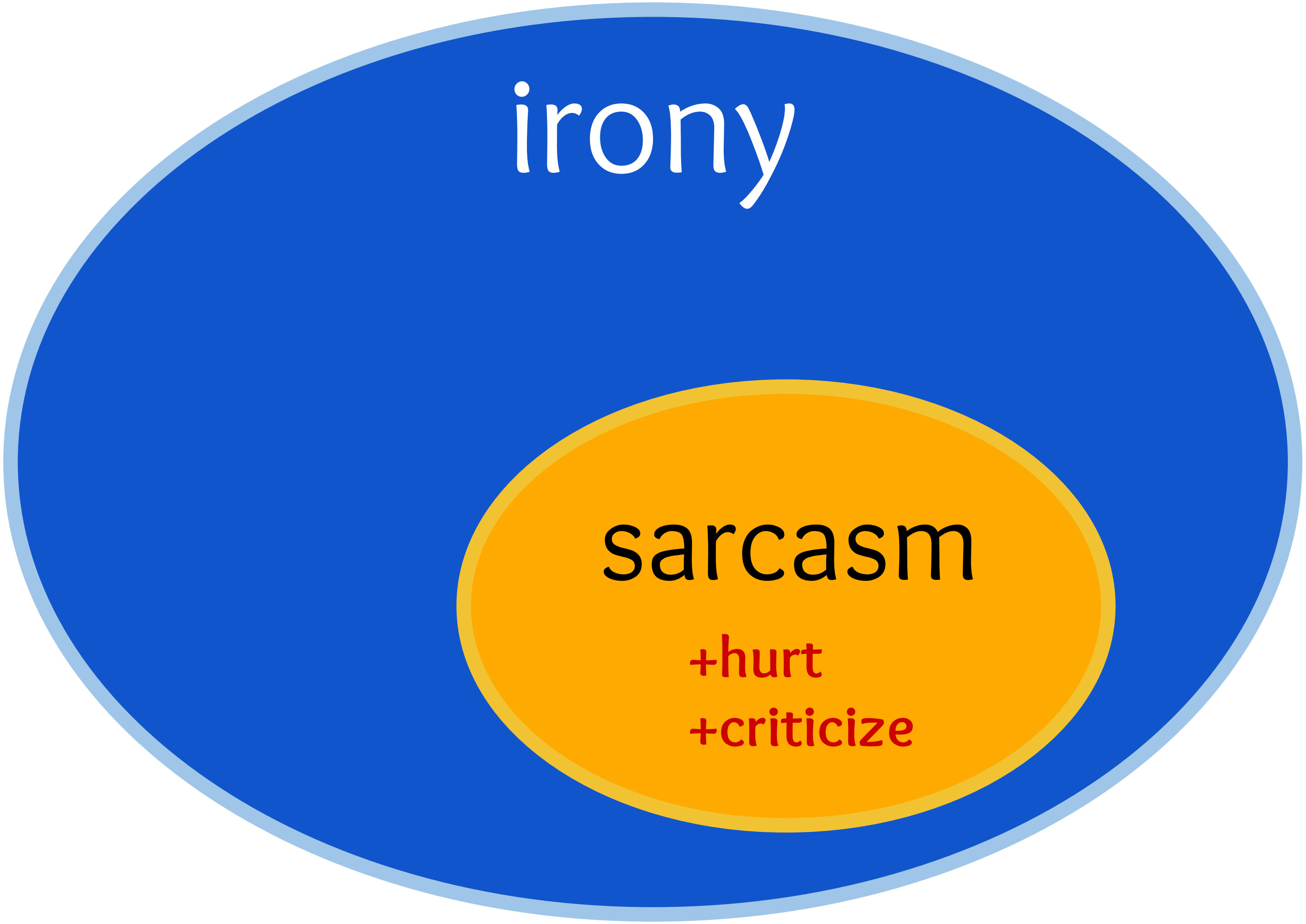
Introduction & Motivation

irony

Introduction & Motivation



Introduction & Motivation



Task Description

INSPIRED BY:

SemEval2018-Task3

Irony detection in English tweets (Van Hee et al., 2018)

Task Description

INSPIRED BY:

SemEval2018-Task3

Irony detection in English tweets (Van Hee et al., 2018)

OUR SHARED TASK:

- TASK A
  - ironic
  - not-ironic
- TASK B
  - sarcastic
  - ironic but not categorized as sarcastic
  - not-ironic

Datasets

OVERLAP WITH:

HSC: Hate Speech corpus (Sanguinetti et al., 2018)

From **HSC:**

Di fronte a queste forme di terrorismo siamo tutti sulla stessa barca. A parte Briatore. Briatore ha la sua.

irony, no sarcasm

Basile et al. (2018)

Datasets

POLITICAL DOMAIN:

TWITTIRÒ corpus (Cignarella et al., 2018)

- LaBuonaScuola corpus (TW-BS) (Stranisci et al., 2016)
- Sentipolc corpus (TW-SENTIPOLC) (Barbieri et al., 2016)
- Spinoza corpus (TW-SPINO) (Barbieri et al., 2016)

From **TWITTIRÒ**:

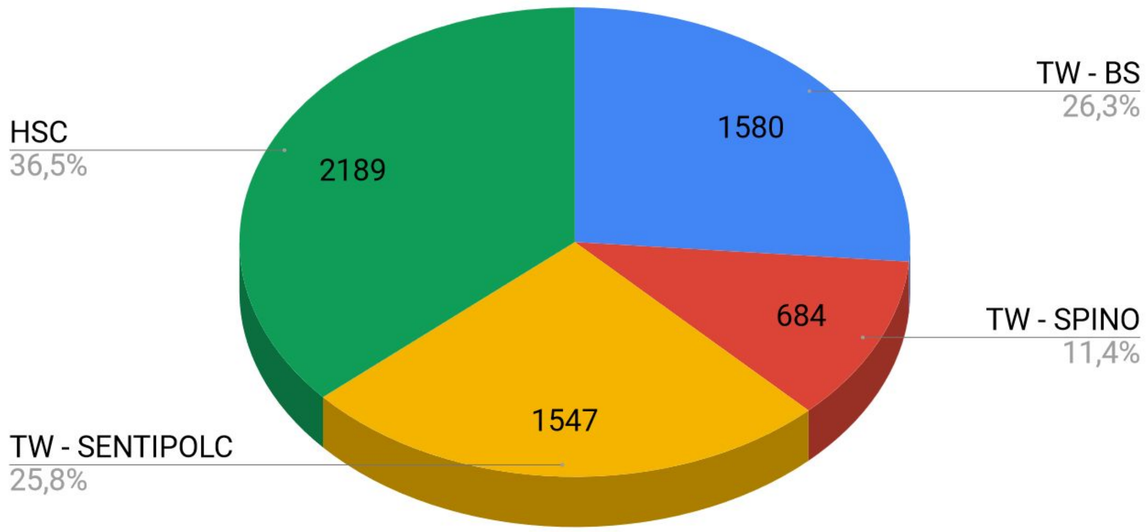
#labuonascuola Fornitura illimitata di rotoli di carta igienica e poi, piano piano, tutti gli altri aspetti meno importanti.

irony, sarcasm

Basile et al. (2018)

Training Data

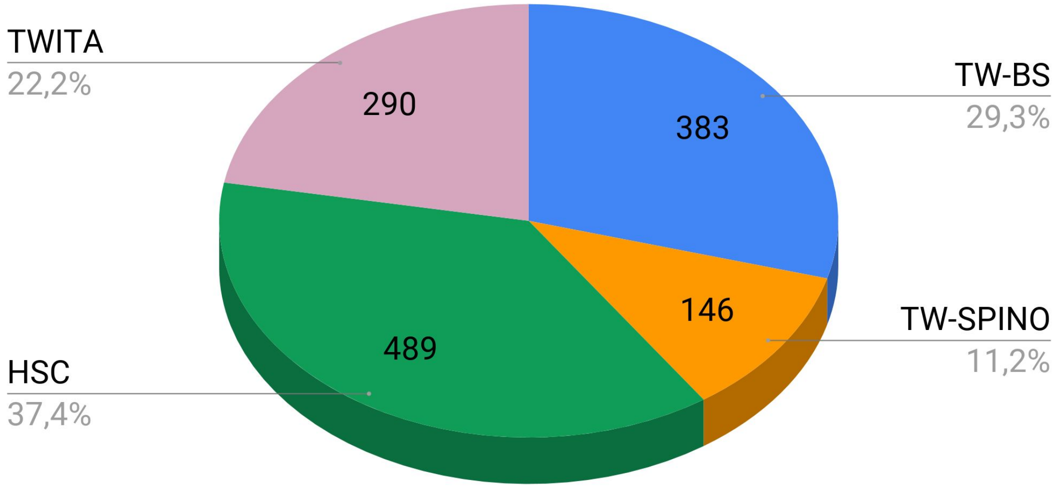
TRAINING SET	IRONIC	NOT-IRO	SARC	NOT-SARC
TW-BS	467	646	173	294
TW-SPINO	342	0	126	216
TW-SENTIPOLC	461	625	143	318
HSC	753	683	471	282
TWITA	0	0	0	0
TOTAL		**3,977**		



Basile et al. (2018)

Test Data

TEST SET				
	IRONIC	NOT-IRO	SARC	NOT-SARC
TW-BS	111	161	51	60
TW-SPINO	73	0	32	41
TW-SENTIPOLC	0	0	0	0
HSC	185	119	106	79
TWITA	67	156	28	39
TOTAL		**872**		



Basile et al. (2018)

Participants

24 submissions received by

1. ItaliaNLP
   ItaliaNLP group ILC-CNR
2. UNIBA
   University of Bari
3. 2XCheck
   App2Check srl
4. UNITOR
   University of Roma “Tor Vergata”
5. Aspie96
   University of Torino
6. UO_IRO
   CERPAMID, Santiago de Cuba
   and
   Havana University of Informatic Sciences
7. venses-itgetarun
   Ca’ Foscari University of Venice

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	ItaliaNLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

HaSpeeDe

TWITTIRÒ

SENTIPOLC 2016

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Participating Systems

	Italia NLP	UNIBA	X2 Check	UNITOR	Aspie96	UO_IRO	venses - itgetarun
Architecture	BiLSTM	SVM	SVM	SVM	GRU	BiLSTM	Rule-based
Preprocessing	✓			✓		✓	
Stylistic Features		✓	✓	✓	✓	✓	
Linguistic Features			✓	✓		✓	✓
Semantic Features	✓	✓		✓		✓	
Lexica	✓			✓		✓	✓
Add. training data	✓		✓	✓		✓	
Task A	✓	✓	✓	✓	✓	✓	✓
Task B	✓			✓	✓		✓
Interdependence	✓			✓			✓

Evaluation

$$precision_{class} = \frac{\#correct\_class}{\#assigned\_class}$$

$$recall_{class} = \frac{\#correct\_class}{\#total\_class}$$

$$F1_{class} = 2\frac{precision_{class} recall_{class}}{precision_{class} + recall_{class}}$$

Task A - Irony Detection

The average of F1-scores for the ironic and not-ironic classes

Evaluation

Task B: Different types of irony

Macro F1-score computed over the three classes

- ✓ not-ironic
  `irony=0, sarcasm=0`
- ✓ ironic-not-sarcastic
  `irony=1, sarcasm=0`
- ✓ sarcastic
  `irony=1, sarcasm=1`

Results Task A

	team name	type	run	not-ironic F1	ironic F1	avg. F1
1	ItaliaNLP	c	1	0.707	**0.754**	**0.731**
2	ItaliaNLP	c	2	0.693	0.733	0.713
3	UNIBA	c	1	0.689	0.730	0.710
4	UNIBA	c	2	0.689	0.730	0.710
5	X2Check	u	1	**0.708**	0.700	0.704
6	UNITOR	c	1	0.662	0.739	0.700
7	UNITOR	u	2	0.668	0.733	0.700
8	X2Check	u	2	0.700	0.689	0.695
9	Aspie96	c	1	0.668	0.722	0.695
10	X2Check	c	2	0.679	0.708	0.693
11	X2Check	c	1	0.674	0.693	0.683
12	UOIRO	u	2	0.603	0.700	0.651
13	UOIRO	u	1	0.626	0.665	0.646
14	UOIRO	c	2	0.579	0.678	0.629
15	UOIRO	c	1	0.652	0.577	0.614
16	*baseline-random*	*c*	*1*	*0.503*	*0.506*	*0.505*
17	venses-itgetarun	c	1	0.651	0.289	0.470
18	venses-itgetarun	c	2	0.645	0.195	0.420
19	*baseline-mfc*	*c*	*1*	*0.668*	*0.000*	*0.334*

Results Task A

Lorenzo De Mattei, Andrea Cimino, and Felice Dell'Orletta. 2018. **Multi-task learning in Deep Neural Networks for Irony Detection.**

			not-ironic	ironic	avg.
team name	type	run	F1	F1	F1
1 ItaliaNLP	c	1	0.707	**0.754**	**0.731**
2 ItaliaNLP	c	2	0.693	0.733	0.713
3 UNIBA	c	1	0.689	0.730	0.710
4 UNIBA	c	2	0.689	0.730	0.710
5 X2Check	u	1	**0.708**	0.700	0.704
6 UNITOR	c	1	0.662	0.739	0.700
7 UNITOR	u	2	0.668	0.733	0.700
8 X2Check	u	2	0.700	0.689	0.695
9 Aspie96	c	1	0.668	0.722	0.695
10 X2Check	c	2	0.679	0.708	0.693
11 X2Check	c	1	0.674	0.693	0.683
12 UOIRO	u	2	0.603	0.700	0.651
13 UOIRO	u	1	0.626	0.665	0.646
14 UOIRO	c	2	0.579	0.678	0.629
15 UOIRO	c	1	0.652	0.577	0.614
16 *baseline-random*	*c*	*1*	*0.503*	*0.506*	*0.505*
17 venses-itgetarun	c	1	0.651	0.289	0.470
18 venses-itgetarun	c	2	0.645	0.195	0.420
19 *baseline-mfc*	*c*	*1*	*0.668*	*0.000*	*0.334*

Results Task A

Emanuele Di Rosa and Alberto Durante. 2018. **Irony detection in tweets: X2Check at Ironita 2018.**

SENTIPOLC 2016: tweets annotated with irony

	team name	type	run	not-ironic F1	ironic F1	avg. F1
1	ItaliaNLP	c	1	0.707	**0.754**	**0.731**
2	ItaliaNLP	c	2	0.693	0.733	0.713
3	UNIBA	c	1	0.689	0.730	0.710
4	UNIBA	c	2	0.689	0.730	0.710
5	X2Check	u	1	**0.708**	0.700	0.704
6	UNITOR	c	1	0.662	0.739	0.700
7	UNITOR	u	2	0.668	0.733	0.700
8	X2Check	u	2	0.700	0.689	0.695
9	Aspie96	c	1	0.668	0.722	0.695
10	X2Check	c	2	0.679	0.708	0.693
11	X2Check	c	1	0.674	0.693	0.683
12	UOIRO	u	2	0.603	0.700	0.651
13	UOIRO	u	1	0.626	0.665	0.646
14	UOIRO	c	2	0.579	0.678	0.629
15	UOIRO	c	1	0.652	0.577	0.614
16	*baseline-random*	*c*	*1*	*0.503*	*0.506*	*0.505*
17	venses-itgetarun	c	1	0.651	0.289	0.470
18	venses-itgetarun	c	2	0.645	0.195	0.420
19	*baseline-mfc*	*c*	*1*	*0.668*	*0.000*	*0.334*

Results **Task B**

	team name	type	run	not ironic F1	ironic F1	sarcastic F1	avg. F1
1	UNITOR	u	2	0.668	0.447	0.446	0.520
2	UNITOR	c	1	0.662	0.432	0.459	0.518
3	ItaliaNLP	c	1	0.707	0.432	0.409	0.516
4	ItaliaNLP	c	2	0.693	0.423	0.392	0.503
5	Aspie96	c	1	0.668	0.438	0.289	0.465
6	*baseline-random*	*c*	*1*	*0.503*	*0.266*	*0.242*	*0.337*
7	venses-itgetarun	c	1	0.431	0.260	0.018	0.236
8	*baseline-mfc*	*c*	*1*	*0.668*	*0.000*	*0.000*	*0.223*
9	venses-itgetarun	c	2	0.413	0.183	*0.000*	0.199

Results Task B

team name	type	run	not ironic F1	ironic F1	sarcastic F1	avg. F1
1 UNITOR	u	2	0.668	0.447	0.446	0.520
2 UNITOR	c	1	0.662	0.432	0.459	0.518
3 ItaliaNLP	c	1	0.707	0.432	0.409	0.516
4 ItaliaNLP	c	2	0.693	0.423	0.392	0.503
5 Aspie96	c	1	0.668	0.438	0.289	0.465
6 *baseline-random*	*c*	*1*	*0.503*	*0.266*	*0.242*	*0.337*
7 venses-itgetarun	c	1	0.431	0.260	0.018	0.236
8 *baseline-mfc*	*c*	*1*	*0.668*	*0.000*	*0.000*	*0.223*
9 venses-itgetarun	c	2	0.413	0.183	*0.000*	0.199

Andrea Santilli, Danilo Croce, and Roberto Basili. 2018. **A Kernel-based Approach for Irony and Sarcasm Detection in Italian.**

Dataset of **6,000 tweets** (searching for the hashtag #ironia — #irony).

Remarks

- The high performance of deep learning methods
- In general the participating systems work better on Hate domain than on Political domain in both tasks
- A dedicated task on irony detection led to a remarkable improvement of the scores (SENTIPOLC in 2014 `F=0.575` and 2016 `F=0.5412`)
- The scores for Italian are in line with those obtained at SemEval2018-Task3 on irony detection in English tweets (`F= 0.724`)

Remarks

- The high performance of deep learning methods
- In general the participating systems work better on Hate domain than on Political domain in both tasks
- A dedicated task on irony detection led to a remarkable improvement of the scores (SENTIPOLC in 2014 `F=0.575` and 2016 `F=0.5412`)
- The scores for Italian are in line with those obtained at SemEval2018-Task3 on irony detection in English tweets (`F= 0.724`)

Remarks

- The high performance of deep learning methods
- In general the participating systems work better on Hate domain than on Political domain in both tasks
- A dedicated task on irony detection led to a remarkable improvement of the scores (SENTIPOLC in 2014 F=0.575 and 2016 F=0.5412)
- The scores for Italian are in line with those obtained at SemEval2018-Task3 on irony detection in English tweets (F= 0.724)

Remarks

- The high performance of deep learning methods
- In general the participating systems work better on Hate domain than on Political domain in both tasks
- A dedicated task on irony detection led to a remarkable improvement of the scores (SENTIPOLC in 2014 `F=0.575` and 2016 `F=0.5412`) **+0.19**
- The scores for Italian are in line with those obtained at SemEval2018-Task3 on irony detection in English tweets (`F= 0.724`)

Remarks

- The high performance of deep learning methods
- In general the participating systems work better on Hate domain than on Political domain in both tasks
- A dedicated task on irony detection led to a remarkable improvement of the scores (SENTIPOLC in 2014 `F=0.575` and 2016 `F=0.5412`)
- The scores for Italian are in line with those obtained at SemEval2018-Task3 on irony detection in English tweets (`F= 0.724`)

Thank you :)

We will wait for you at the poster!