Presenting TWITTIRÒ-UD:
An Italian Twitter Treebank in Universal Dependencies

Alessandra Teresa Cignarella

Dipartimento di Informatica, Università degli Studi di Torino

PRHLT Research Center, Universitat Politècnica de València

`cigna@di.unito.it`

Cristina Bosco

Dipartimento di Informatica

Università degli Studi di Torino

`bosco@di.unito.it`

Paolo Rosso

PRHLT Research Center

Universitat Politècnica de València

`prosso@dsic.upv.es`

Abstract

In this paper we describe the early stage application of the Universal Dependencies to an Italian corpus from social media developed for shared tasks related to irony and stance detection. The development of this novel resource (TWITTIRÒ-UD) serves a twofold goal: it enriches the scenario of treebanks for social media and for Italian, and it paves the way for a more reliable extraction of a larger variety of morphological and syntactic features to be used by sentiment analysis tools. On the one hand, social media texts are especially hard to parse and the limited amount of resources for training and testing NLP tools further damages the situation. On the other hand, we thought that adding the Universal Dependencies format to the fine-grained annotation for irony, that was previously applied on TWITTIRÒ, might meaningfully help in the investigation of possible relationships between syntax and semantics of the uses of figurative language, irony in particular.

1 Introduction

In the last decade, the interest towards social networking sites has grown considerably and the NLP community has been relying more and more on data extracted from social media and micro-blogs. In particular, thanks to the APIs provided by the platform, and the fact there is a variety of expressions of people's sentiments and opinions, Twitter has become one of the most exploited sources for the retrieval of data, especially in the fields of Sentiment Analysis (SA) and Opinion Mining. Nevertheless, although humans can understand each other while they exchange social media contents, which are featured by non-standard word-forms, misspelled words, dialectal word-forms, emojis and elongated words, dealing with them still proves to be a very hard challenge for automatic analyses, especially concerning syntax and morphology.

In this paper we introduce a novel Twitter treebank for Italian, i.e. TWITTIRÒ-UD. The data come from a resource originally developed for training and testing irony detection systems, also exploited as a benchmark for the Italian irony detection task held in EVALITA 2018[1] (Cignarella et al., 2018b). In order to pave the way towards collecting evidences about the relationships between syntax and semantic knowledge involved in SA tasks we are developing this project of annotation which encompasses in TWITTIRÒ-UD both the fine-grained annotation for irony applied in a multilingual setting in Karoui et al. (2017) and that morphological and syntactic provided by Universal Dependencies (UD). An alike resource will allow us to extract morphological and syntactic features to be used to improve the performance in irony and stance detection tasks (Duric and Song, 2012; Sidorov et al., 2014). The UD resources available for Italian and social media meaningfully helped us in the morphological and syntactic analysis of the dataset (Bosco et al., 2014; Sanguinetti et al., 2017; Sanguinetti et al., 2018).

This paper is organized as follows. The next section briefly surveys the literature about Italian social media UD resources. Section 3 introduces the dataset used for our project and describes the various

---

[1] `http://www.evalita.it/2018`

annotation steps. In Sections 4 we discuss the creation of the gold standard set, and we highlight the findings of a quantitative analysis. Finally, in Section 5 we draw some considerations on the current state of the project and give some insights on future work.

2 Related Work

In recent years UD have become the standard for syntactic annotation (De Marneffe et al., 2014; Nivre et al., 2016) and the repository of UD projects enlarges by the day, also including data for under-resourced languages and less studied varieties, see e.g. Wang et al. (2017). As far as Italian is concerned, the main UD resources, that we exploited as reference, are two: namely, the UD-Italian treebank (Simi et al., 2014) and PoSTWITA-UD (Sanguinetti et al., 2017; Sanguinetti et al., 2018). The former entails standard texts drawn from newspapers, legal codes and Wikipedia, the latter texts from social media.

The genre of social media texts can be a bottleneck for morphological and syntactic analysis, but some experiments are reported in literature about parsing this type of data, see e.g. (Foster et al., 2011) and (Kong et al., 2014), who introduce the dependency parser TWEEBOPARSER and TWEEBANK, a Twitter treebank later extended in TWEEBANK V2 (Liu et al., 2018). In Albogamy and Ramsay (2017) an Arabic dependency treebank of tweets is converted in the UD format, while in (Blodgett et al., 2018) a treebank of tweets in African-American English is created, and in Bhat et al. (2018) a UD treebank of Hindi-English is created focusing on syntactic aspects of code-switching.

Finally, addressing the morphological analysis of social media, the task organized in the 2016's edition of EVALITA[2] can be cited. In this edition of the evaluation campaign for NLP and speech tools for Italian, a task about PoS-tagging of social media texts has been organized (Bosco et al., 2016) which was centered on the POSTWITA corpus, i.e. that later enriched with UD annotation for creating PoSTWITA-UD. This kind of experience encourages the community to adapt NLP tools to this different type of text domain, which is noisy and difficult to deal with automatically.

3 Data and Annotation

The data of TWITTIRÒ-UD are drawn from TWITTIRÒ (Cignarella et al., 2018a; Cignarella et al., 2019), a gold standard Italian corpus for irony detection. It has been firstly annotated according to the fine-grained schema for irony proposed in Karoui et al. (2017). Later it has been extended with the annotation for sarcasm exploited in the EVALITA 2018 task on irony detection in Italian tweets (IronITA[3]) (Cignarella et al., 2018b). The corpus includes 1,424 tweets annotated as follows.

```
# sent_id = 507111702744162304
# twittiro = EXPLICIT HYPERBOLE
# sarcasm = 0
# text = se sento ancora la parola merito vomito #labuonascuola #chenonèquelladirenzi
```

In the tweet[4] two features are marked for irony, i.e. the fact that all the elements necessary for interpreting the irony are lexically represented in the post (EXPLICIT), and that a particular device (HYPERBOLE) triggers irony, while a binary annotation has been applied for marking the (absence of) sarcasm. In TWITTIRÒ-UD, this annotation manually provided and revised in the original resource is enhanced by that for morphology and syntax according to UD (see examples in Sec. 3.1).

In order to create TWITTIRÒ-UD, we applied the full pipeline of tokenization, lemmatization, PoS-tagging and dependency parsing provided by *UDPipe*[5] (Straka and Straková, 2017). For this purpose, we trained UDPipe on two different gold benchmarks, namely PoSTWITA-UD (Sanguinetti et al., 2018) (6,712 tokens) and UD_Italian (Simi et al., 2014) (14,167 tokens). Considering the typology of text and the features of ironic messages, we followed the PoSTWITA-UD tenets, in particular for what concerns segmentation, which is at tweet level rather than at sentence level.

---

[2]http://www.evalita.it/2016/tasks/postwita.

[3]http://www.di.unito.it/~tutreeb/ironita-evalita18/index.html

[4]Translation: if I hear again the word merit I will throw up #labuonascuola #thatisnotthatofrenzi.

[5]In the UDPipe pipeline, the parsing is performed using Parsito (http://ufal.mff.cuni.cz/parsito).

3.1 Issues in Manual Correction

In this paper we focus on a subset of the original corpus, which includes 897 tweets only, while we plan a second release in the UD repository including the full corpus for November 2019. From the manual correction of this dataset[6] we have already learned some interesting lessons.

Tokenization

Several tokenization errors depend on misspelled words (i.e. not correctly separated by spaces) or punctuation irregularly used, like in the following example.

```
# sent_id = 516493351034826752
# twittiro = EXPLICIT RHETORICAL QUESTION
# sarcasm = 0
# text = @User #labuonascuola deve riconoscere il merito di chi ha superato il concorso...solo in Italia chi
vince perde?#dalleparoleaifatti

1 @User @User SYM SYM _ 4 vocative:mention _ _
2 #labuonascuola #labuonascuola SYM SYM _ 4 nsubj _ _
3 deve dovere AUX VM Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 4 aux _ _
4 riconoscere riconoscere VERB V VerbForm=Inf 0 root _ _
5 il il DET RD Definite=Def|Gender=Masc|Number=Sing|PronType=Art 6 det _ _
6 merito merito NOUN S Gender=Masc|Number=Sing 4 obj _ _
7 di di ADP E _ 8 case _ _
8 chi chi PRON PR PronType=Rel 6 nmod _ _
9 ha avere AUX VA Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 10 aux _ _
10 superato superare VERB V Gender=Masc|Number=Sing|Tense=Past|VerbForm=Part 8 acl:relcl _ _
11 il il DET RD Definite=Def|Gender=Masc|Number=Sing|PronType=Art 12 det _ _
12 concorso...solo concorso...solo NOUN S Gender=Masc|Number=Sing 10 obj _ _
13 in in ADP E _ 14 case _ _
14 Italia Italia PROPN SP _ 12 nmod _ _
15 chi chi PRON PR PronType=Rel 17 nsubj _ _
16 vince vincere VERB V Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 15 acl:relcl _ _
17 perde?#dalleparoleaifatti perde?#dalleparoleaifatti CCONJ CC _ 4 cc _ SpacesAfter=\n
```

In line 12 and line 17 of the tweet[7] we find "concorso...solo" and "perde?#dalleparoleaifatti", which should be split in three different tokens each. In order to avoid that the failures in tokenization propagate in the other annotation levels, before tokenization we applied an automatic data cleaning which consists in always adding a white space between words and punctuation signs (with the exception of the apostrophe which left attached to the preceding token). We only manually corrected the remaining cases of misspelled tokens, that is not separated by the necessary white space. The result of the correction of the example above can be seen below (where we also corrected the PoS tags).

```
# sent_id = 516493351034826752
# twittiro = EXPLICIT RHETORICAL QUESTION
# sarcasm = 0
# text = @User #labuonascuola deve riconoscere il merito di chi ha superato il concorso...solo in Italia chi
vince perde?#dalleparoleaifatti

1 @User @User SYM SYM _ 4 vocative:mention _ _
2 #labuonascuola #labuonascuola SYM SYM _ 4 nsubj _ _
3 deve dovere AUX VM Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 4 aux _ _
4 riconoscere riconoscere VERB V VerbForm=Inf 0 root _ _
5 il il DET RD Definite=Def|Gender=Masc|Number=Sing|PronType=Art 6 det _ _
6 merito merito NOUN S Gender=Masc|Number=Sing 4 obj _ _
7 di di ADP E _ 8 case _ _
8 chi chi PRON PR PronType=Rel 6 nmod _ _
9 ha avere AUX VA Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 10 aux _ _
10 superato superare VERB V Gender=Masc|Number=Sing|Tense=Past|VerbForm=Part 8 acl:relcl _ _
11 il il DET RD Definite=Def|Gender=Masc|Number=Sing|PronType=Art 12 det _ _
12 concorso concorso NOUN S Gender=Masc|Number=Sing 10 obj _ SpaceAfter=No
13 ... ... PUNCT FS _ 10 punct _ SpaceAfter=No
14 solo solo ADV B _ 16 advmod _ _
```

---

[6]We exploited the *Dependency Grammar Annotator*: `http://medialab.di.unipi.it/Project/QA/Parser/DgAnnotator/`.

[7]Translation: @User #labuonascuola needs to acknowledge the merit of whose who passed the competition...only in Italy who wins also loses? #fromwordstofacts.

15 in in ADP E _ 16 case _ _
16 Italia Italia PROPN SP _ 4 obl _ _
17 chi chi PRON PR PronType=Rel 19 nsubj _ _
18 vince vincere VERB V Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 17 acl:relcl _ _
19 perde perdere VERB V Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 4 acl:relcl _
SpaceAfter=No
20 ? ? PUNCT FS _ 19 punct _ SpaceAfter=No
21 #dalleparoleaifatti #dalleparoleaifatti SYM SYM _ 4 parataxis:hashtag _ SpaceAfter=\n

Lemmatization and PoS-tagging

Misspelled forms often occurring in social media contents cannot be recognized by lemmatizers and their analysis may result in a failure. Here, as it was done in the annotation of PoSTWITA-UD, we associated the non-standard forms with the lemmas of their normalized versions, thus allowing a correct PoS-tagging. For instance, the typo *anema* is paired with the lemma *anima* (soul), the abbreviation *ke* with *che* (that), the elongated *nooo* with *no* (no), and the abbreviations *X* and *h* respectively with *per* (for) and *ora* (hour). Emoticons, emojis, URLs, email addresses, and Twitter marks (hashtags and mentions) have been instead labelled with the tag SYM.

Dependency Relations Attachment

As said above, following the strategy applied in POSTWITA-UD, we did not perform any sentence splitting in the novel dataset. Each syntax tree of TWITTIRÒ-UD corresponds to a tweet in its entirety, and may consist of multiple sentences too. At the same time, provided that the UD scheme poses a single-root constraint, the internal connections between different sentences occurring in a tweet have to be annotated and labeled by the dependency relation parataxis. This relation is quite hard to be provided by the parser, which often fails in recognizing this kind of structure. See for instance, Figure 1 where we display a tweet[8] containing more paratactic structures.
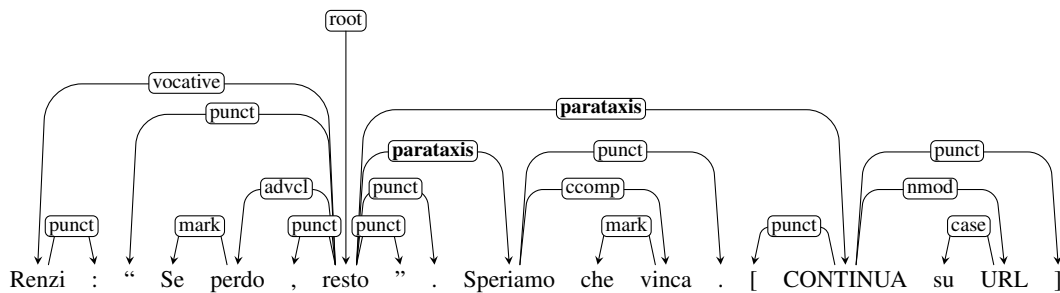

Figure 1: Example of tweet containing multiple sentences.

Another issue is related to the wide presence of Twitter marks. The current limited amount of adequate training data prevents the parser from dealing with them successfully. Within the manual correction phase, we resort to the label vocative:mention for Twitter mentions, the label discourse:emo for emojis, and dep for URLs.
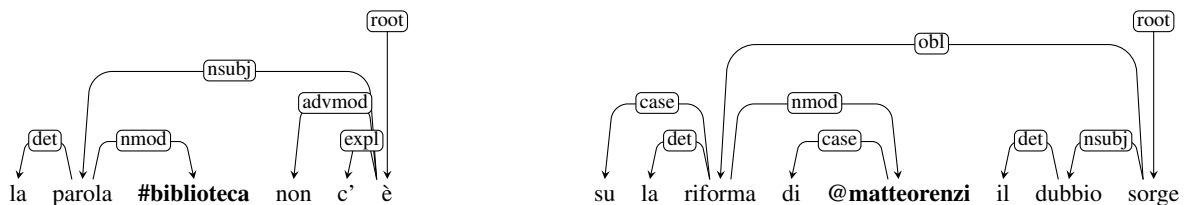

Figure 2: Examples of tweets containing a hashtag and a mention with syntactic function.

Moreover, hashtags and mentions could be either used at the end of the tweet, to create more emphasis, or with a full syntactic function. In the first case, we resort to the relation (parataxis:hashtags and

---

[8]Translation: Renzi: "If I lose, I stay". Let's hope he wins then. [Follows on URL].

`vocative:mention`), while in the second we annotate accordingly to the syntactic role, see for example in Fig. 2 the hashtag and the mention[9] labelled as `nmod`.

		PoSTWITA-UD	TWITTIRÒ-UD
hashtags	`parataxis:hashtag`	40.89%	54.79%
	`nmod`	19.64%	11.55%
	`nsubj`	13.48%	8.59%
	other	25.99%	25.07%
mentions	`vocative:mention`	92.37%	87.41%
	other	7.63%	12.59%

Table 1: Distribution of deprel labels for hashtags and mentions.

Table 1 shows the distribution of the dependency relations (*deprels*), and confirms that there is a syntactic correlate of the peculiar semantic role that hashtags and mentions play in tweets. The labels that are mostly exploited for linking the hashtags to the sentence structure PoSTWITA-UD and in TWITTIRÒ-UD are mostly two: `nmod` and `nsubj`.

4   Analysis and Discussion

Table 2 shows the distribution of deprels in UD_Italian, PoSTWITA-UD and TWITTIRÒ-UD.

	UD_Ita	PoSTW	TWIT		UD_Ita	PoSTW	TWIT
`acl`	0.99	0.48	0.65	`flat`	0.19	0.35	0.10
`acl:relcl`	1.06	0.68	0.71	`flat:foreign`	0.05	0.28	0.05
`advcl`	1.26	1.00	0.90	`flat:name`	1.17	2.18	0.85
`advmod`	3.53	4.85	4.21	`goeswith`	0.00	0.03	-
`amod`	5.59	2.75	3.49	`iobj`	0.23	0.75	0.52
`appos`	0.31	0.43	0.16	`list`	-	0.22	-
`aux`	2.02	1.67	1.80	`mark`	2.11	2.23	2.10
`aux:pass`	0.75	0.12	0.18	`nmod`	8.01	6.84	5.68
`case`	14.03	9.42	10.23	`nsubj`	4.30	4.50	4.40
`cc`	2.73	2.26	1.80	`nsubj:pass`	0.77	0.16	0.26
`ccomp`	0.49	0.80	0.67	`nummod`	1.20	0.88	0.93
`compound`	0.25	0.17	0.27	`obj`	3.43	4.10	4.64
`conj`	3.39	2.95	1.72	`obl`	5.77	4.03	4.80
`cop`	1.15	1.75	1.54	`obl:agent`	0.38	0.12	0.13
`csubj`	0.11	0.17	0.07	`orphan`	0.01	0.05	-
`csubj:pass`	0.00	-	-	`parataxis`	0.14	4.02	4.62
`dep`	0.00	2.34	0.89	`parataxis:appos`	-	0.10	0.01
`det`	15.54	10.97	10.98	`parataxis:discourse`	-	0.02	0.01
`det:poss`	0.63	0.48	0.31	`parataxis:hashtag`	-	1.81	2.15
`det:predet`	0.14	0.12	0.11	`parataxis:insert`	-	0.03	-
`discourse`	0.02	1.18	0.75	`parataxis:nsubj`	-	0.03	-
`discourse:emo`	-	0.59	0.13	`parataxis:obj`	-	0.07	-
`dislocated`	0.01	0.11	0.01	`punct`	11.36	12.08	17.24
`expl`	0.73	0.85	0.96	`root`	4.75	5.39	4.77
`expl:impers`	0.14	0.15	0.13	`vocative`	0.03	0.38	0.09
`expl:pass`	0.13	0.05	0.04	`vocative:mention`	-	2.06	2.89
`fixed`	0.32	0.19	0.30	`xcomp`	0.76	0.76	0.78

Table 2: Dependency relations' distribution across the three main Italian treebanks. The values are expressed in percentage %.

We can observe, despite the sparseness of relations, how their frequency and distribution characterizes the language exploited in the social media data collected in TWITTIRÒ-UD and PoSTWITA-UD with respect to the standard language collected in UD_Italian. As expected, meaningful differences emerge for parataxis and punctuation. Punctuation is indeed exploited more extensively in the two social media datasets (12.08% and 17.24%) than in UD_Italian (11.36%), and the frequency of the `parataxis` deprel is 4.02% and 4.62% in PoSTWITA and TWITTIRÒ-UD, while it is only 0.14% in UD_Italian, marking a

---

[9]Translation: `about the reform of @matteorenzi a doubt rises`.

significant difference. The distributions of the relations `vocative:mention` and `parataxis:hashtag` especially features the two social media treebanks. The mentions' deprel is 2.06% in PoSTWITA-UD and 2.89% in TWITTIRÒ-UD, while the hashtags are respectively 1.81% and 2.15%.Furthermore, it is interesting to notice how the use of passive voices (`aux:pass`) is 0.75% in the UD_Italian treebank while only 0.12% in PoSTWITA-UD and only 0.18% in TWITTIRÒ-UD, indicating a preference for the exploitation of active voices in the language used in social media, as it happens in spoken language.

4.1 A Parsing Experiment

In order to preliminary evaluate the similarities between the three datasets, we performed an evaluation of UDPipe using the TWITTIRÒ-UD gold corpus as a test set. The following three settings were exploited.

> 1) training UDPipe using only UD_Italian (UD_It),
> 2) training UDPipe using only PoSTWITA-UD (PoSTW),
> 3) and training UDPipe using both resources (UD_It+PoSTW).

For evaluation we used the script made available for the CoNLL 2018 Shared Task 5[10] with the default setting parameters. Table 3 surveys the resulting scores for precision (P), recall (R) and averaged F1-score (F1).

	UD_It			PoSTW			UD_It+PoSTW		
	P	R	F1	P	R	F1	P	R	F1
Tokens	66.85	67.28	67.07	66.50	65.15	65.82	67.62	67.63	67.62
Sentences	66.18	66.18	66.18	66.18	66.18	66.18	66.18	66.18	66.18
Words	66.73	67.12	66.92	66.36	65.01	65.67	67.54	67.56	67.55
UPOS	57.10	57.44	57.27	62.71	61.44	62.07	65.75	65.77	65.76
XPOS	56.30	56.63	56.47	62.23	60.97	61.59	65.59	65.61	65.60
Feats	59.35	59.70	59.52	62.17	60.91	61.53	65.64	65.66	65.65
AllTags	55.11	55.43	55.27	60.59	59.36	59.97	65.04	65.06	65.05
Lemmas	60.88	61.23	61.05	62.17	60.91	61.53	65.48	65.50	65.49
UAS	66.73	67.12	66.92	66.36	65.01	65.67	67.54	67.56	67.55
LAS	50.12	50.42	50.27	54.07	52.97	53.51	56.84	56.85	56.85

Table 3: Evaluation of UDPipe.

First of all, it is interesting to notice the variation of the Unlabelled Attachment Score (UAS) and Labelled Attachment Score (LAS). For what concerns UAS, the first setup, where only the data from UD_Italian have been used for training, allowed a better result than the second one, where PoSTWITA-UD is the training dataset. But the opposite can be seen for LAS. We can hypothesize that the larger amount of data in UD_Italian allowed to build a more representative statistical model. Nevertheless, training on a resource which includes the same typology of data may be crucial for collecting an adequate knowledge about the specific relations exploited. This motivates the best scores for LAS an UAS, which were obtained in the third setup benefiting of both the resources for training. This encourages us to develop more and better gold standard treebanks also for social media to be used for training.

5 Conclusion and Future Work

In this paper we presented an ongoing project for the development of a novel Italian treebank from Twitter in the UD format: TWITTIRÒ-UD. Focusing on the 897 tweets currently annotated for the first release, we discuss the annotation of this resource which encompasses a fine-grained representation of irony and the UD morpho-syntactic analysis.

The preliminary analysis we applied shows some difference in the distribution of dependency relations in standard Italian and social media language, e.g. in the use of verbal active/passive voices, confirming that the language used in social media presents a strong preference for the exploitation of active voices. Furthermore, a simple parsing experiment and a comparison among the novel resource, UD_Italian (Simi et al., 2014) and PoSTWITA-UD (Sanguinetti et al., 2018) are provided, in order to shed light on the

---

[10]`http://universaldependencies.org/conll17/evaluation.html`

syntactic features of social media texts. Also considering the perspective of the future release of the complete resource (1,424 tweets) to be accomplished before the next UD release in November 2019, the work serves a twofold goal: it enriches the scenario of available resources for a text genre which is especially hard to parse (social media text), and helps in the investigation of possible relationships between syntax and semantics of the uses of figurative language (irony in particular). The availability of a resource whose annotation encompasses both UD relations and a fine-grained description of irony may indeed pave the way for the investigation of whether syntactic knowledge might help in SA and other related tasks.

Acknowledgments

References

Fahad Albogamy and Allan Ramsay. 2017. Universal Dependencies for Arabic Tweets. In *RANLP*, pages 46–51.

Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Manish Shrivastava, and Dipti Misra Sharma. 2018. Universal Dependency parsing for Hindi-English code-switching. *arXiv preprint arXiv:1804.05868*.

Su Lin Blodgett, Johnny Wei, and Brendan O'Connor. 2018. Twitter Universal Dependency parsing for African-American and mainstream American English. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1415–1425.

Cristina Bosco, Felice Dell'Orletta, Simonetta Montemagni, Manuela Sanguinetti, and Maria Simi. 2014. The EVALITA 2014 dependency parsing task. In *EVALITA 2014 Evaluation of NLP and Speech Tools for Italian*, pages 1–8.

Cristina Bosco, Tamburini Fabio, Bolioli Andrea, and Alessandro Mazzei. 2016. Overview of the EVALITA 2016 part of speech on Twitter for Italian task. In *CEUR Workshop Proceedings*, volume 1749, pages 1–7.

Alessandra Teresa Cignarella, Cristina Bosco, Viviana Patti, and Mirko Lai. 2018a. Application and Analysis of a Multi-layered Scheme for Irony on the Italian Twitter Corpus TWITTIRÒ. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC-2018)*.

Alessandra Teresa Cignarella, Simona Frenda, Valerio Basile, Cristina Bosco, Viviana Patti, Paolo Rosso, et al. 2018b. Overview of the EVALITA 2018 task on Irony Detection in Italian Tweets (IronITA). In *Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018)*, volume 2263, pages 1–6.

Alessandra Teresa Cignarella, Cristina Bosco, Viviana Patti, and Mirko Lai. 2019. TWITTIRÒ: an Italian Twitter Corpus with a Multi-layered Annotation for Irony. *IJCoL - Italian Journal of Computational Linguistics*, 4(2):25–44.

Marie-Catherine De Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In *LREC*, volume 14, pages 4585–92.

Adnan Duric and Fei Song. 2012. Feature selection for sentiment analysis based on content and syntax models. *Decision support systems*, 53(4):704–711.

Jennifer Foster, Ozlem Cetinoglu, Joachim Wagner, Joseph Le Roux, Stephen Hogan, Joakim Nivre, Deirdre Hogan, and Josef Van Genabith. 2011. #hardtoparse: POS Tagging and Parsing the Twitterverse. In *Workshops at the 25th AAAI Conference on Artificial Intelligence*.

Jihen Karoui, Benamara Farah, Véronique Moriceau, Viviana Patti, Cristina Bosco, and Nathalie Aussenac-Gilles. 2017. Exploring the impact of pragmatic phenomena on irony detection in tweets: A multilingual corpus study. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 262–272.

Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archna Bhatia, Chris Dyer, and Noah A. Smith. 2014. A dependency parser for tweets. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1001–1012.

Yijia Liu, Yi Zhu, Wanxiang Che, Bing Qin, Nathan Schneider, and Noah A Smith. 2018. Parsing Tweets into Universal Dependencies. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 965–975.

Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan T. McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016)*.

Manuela Sanguinetti, Cristina Bosco, Alessandro Mazzei, Alberto Lavelli, and Fabio Tamburini. 2017. Annotating Italian Social Media Texts in Universal Dependencies. In *Proceedings of the 4th International Conference on Dependency Linguistics (DepLing 2017)*, pages 229–239.

Manuela Sanguinetti, Cristina Bosco, Alberto Lavelli, Alessandro Mazzei, Oronzo Antonelli, and Fabio Tamburini. 2018. PoSTWITA-UD: an Italian Twitter Treebank in Universal Dependencies. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC-2018)*.

Grigori Sidorov, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández. 2014. Syntactic n-grams as Machine Learning Features for Natural Language Processing. *Expert Systems with Applications*, 41(3):853–860.

Maria Simi, Cristina Bosco, and Simonetta Montemagni. 2014. Less is More? Towards a Reduced Inventory of Categories for Training a Parser for the Italian Stanford Dependencies. In *Proceedings of LREC 2014*, page 83–90.

Milan Straka and Jana Straková. 2017. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99.

Hongmin Wang, Yue Zhang, GuangYong Leonard Chan, Jie Yang, and Hai Leong Chieu. 2017. Universal Dependencies Parsing for Colloquial Singaporean English. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1732–1744.