



Valerio Basile, Danilo Croce, Maria Maro and Lucia C. Passaro (dir.)

**EVALITA Evaluation of NLP and Speech Tools for Italian
- December 17th, 2020**
Proceedings of the Seventh Evaluation Campaign of Natural
Language Processing and Speech Tools for Italian Final Workshop

Accademia University Press

KIPoS @ EVALITA2020: Overview of the Task on KIParla Part of Speech Tagging

Cristina Bosco, Silvia Ballarè, Massimo Cerruti, Eugenio Gorla and Caterina Mauri

DOI: 10.4000/books.aaccademia.7743
Publisher: Accademia University Press
Place of publication: Torino
Year of publication: 2020
Published on OpenEdition Books: 11 May 2021
Serie: Collana dell'Associazione Italiana di Linguistica Computazionale
Electronic ISBN: 9791280136329



<http://books.openedition.org>

Electronic reference

BOSCO, Cristina ; et al. *KIPoS @ EVALITA2020: Overview of the Task on KIParla Part of Speech Tagging* In: *EVALITA Evaluation of NLP and Speech Tools for Italian - December 17th, 2020: Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian Final Workshop* [online]. Torino: Accademia University Press, 2020 (generated 18 mai 2021). Available on the Internet: <<http://books.openedition.org/aaccademia/7743>>. ISBN: 9791280136329. DOI: <https://doi.org/10.4000/books.aaccademia.7743>.

KIPoS @ EVALITA2020: Overview of the Task on KIParla Part of Speech Tagging

Cristina Bosco^{*}, Silvia Ballarè[◇], Massimo Cerruti[⊕], Eugenio Goria[⊕], Caterina Mauri[⊖]

^{*}Dipartimento di Informatica, Università degli Studi di Torino

[◇]Dipartimento di Filologia Classica e Italianistica, Università degli Studi di Bologna

[⊕]Dipartimento di Studi Umanistici, Università degli Studi di Torino

[⊖]Dipartimento di Lingue, Letterature e Culture Moderne, Università degli Studi di Bologna
{cristina.bosco, massimosimone.cerruti, eugenio.goria}@unito.it,
{silvia.ballare, caterina.mauri}@unibo.it

Abstract

English. The paper describes the first task on Part of Speech tagging of spoken language held at the Evalita evaluation campaign, KIPoS. Benefiting from the availability of a resource of transcribed spoken Italian (i.e. the KIParla corpus), which has been newly annotated and released for KIPoS, the task includes three evaluation exercises focused on formal versus informal spoken texts. The datasets and the results achieved by participants are presented, and the insights gained from the experience are discussed.

Italiano. *L'articolo descrive il primo task sul Part of Speech tagging di lingua parlata tenutosi nella campagna di valutazione Evalita. Usufruento di una risorsa che raccoglie trascrizioni di lingua italiana (il corpus KIParla), annotate appositamente per KIPoS, il task è stato focalizzato intorno a tre valutazioni con lo scopo di confrontare i risultati raggiunti sul parlato formale con quelli ottenuti sul parlato informale. Il corpus di dati ed i risultati raggiunti dai partecipanti sono presentati insieme alla discussione di quanto emerso dall'esperienza di questo task.*

1 Motivation

Even (Bosco et al., 2020) though in the last decades we have witnessed an increase in the resources available for the study of spoken Italian, a great unbalance can still be observed between spoken and written corpora, from different angles.

Written corpora are generally larger, are able to provide a lot of information about the texts they include, and may count on a vast array of computational tools for morphological analysis and syntactic parsing. Conversely, spoken corpora of Italian are generally smaller, often give a minimum of information concerning the speakers and the context in which the interaction takes place and, finally, provide at most basic PoS-tagging and lemmatization tools. This, of course, poses considerable limitations on the searches that may be performed on these resources, eventually leading to a possible *written language bias* due to the different availability and richness of information of written vs. spoken corpora (Linell, 2005).

As a consequence of this unbalance, corpus-based sociolinguistic analyses of spoken Italian, which need a comprehensive set of metadata, have rarely been put to the test on publicly available speech corpora. In fact, most sociolinguistic studies have been conducted on ad hoc-collected datasets, see *inter al.* (Alfonzetti, 2002; Mereu, 2019).

The KIParla corpus (Mauri et al., 2019) (661k tokens approximately), which is available at the website www.kiparla.it, has been designed to overcome some shortcomings of previous resource tools. KIParla is a corpus of spoken Italian which encompasses various types of interactions between speakers of different origins and socio-economic backgrounds. It consists of speech data collected in Bologna and Turin between 2016 and 2019, and contains two independent modules, i.e. KIP (cf. sec. 3) and ParlaTO. Among other things, KIParla provides a wide range of metadata, including situational characteristics (such as the symmetrical vs. asymmetrical relationship between the participants) and socio-demographic information for each speaker (such as age and level of education). Nevertheless, the lack of PoS-tagging and lemmatization currently places severe limits on its

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

application.

In order to enrich the scenario of investigation to be applied on the KIParla corpus, we proposed the KIPoS task. Following the experience of the Evalita 2016 PoSTWITA task on PoS tagging Italian Social Media Texts (Bosco et al., 2016) and the subsequent development of an Italian treebank for social media (Sanguinetti et al., 2017; Sanguinetti et al., 2018), where the issues related to a particularly challenging written text genre were addressed, KIPoS offers the opportunity of addressing the theoretical and methodological challenges related to PoS tagging of Italian spontaneous speech texts. Carrying out this task means processing a type of data that is known to be problematic for computational treatment, that is unplanned spoken language (as opposed to experimental speech data). PoS tagging of this corpus entails dealing with both a wide range of spontaneous speech phenomena and a great amount of sociolinguistic variation.

The most challenging aspects to be addressed in the unconstrained speech of KIParla are:

- To identify mode-specific phenomena, such as repetitions, reformulations, fillers, incomplete syntactic structures, etc.
- To trace a relevant set of non-standard alternatives back to the same linguistic phenomenon (e.g. the presence of socio-geographically marked forms like *annà* or *andà*, equal to standard Italian *andare* "to go"), either assigning them to the correct part-of-speech, or working out an ad-hoc solution.
- To deal with different types of interaction and registers (casual conversations, interviews, office hours, etc.) with a variable number of participants (1 to 5), each transcribed on a separate line and corresponding to an autonomous text string.

PoS-tagging of data from KIParla corpus is intended to bring an improvement to the current practices in use for tagging and parsing spoken Italian. Furthermore, this result is also significant for the purposes of (socio)linguistic research, in that the availability of annotated spoken corpora enables the researcher to validate previous assumptions based on smaller or less informative datasets, but also to collect knowledge to be

meaningfully used in the development of automatic conversation systems and chatbots.

2 Definition of the task

Given the innovative features of KIParla, we proposed KIPoS as a task for EVALITA 2020 (Basile et al., 2020) to address the issues involved in the adaptation of a PoS tagger to the specific features of oral text, in order to systematically represent those features and to provide the mean to access to their specificities. We provided therefore data for training (i.e. Development Set, henceforth DEVSET) and testing (Test Set, henceforth TESTSET) systems organized in two ensembles which respectively represent formal (DEVSET-formal and TESTSET-formal) and informal texts (DEVSET-informal and TESTSET-informal). This allowed us to consider one main task and two subtasks, which are described as follows:

- Main task - general: training on all given data (both DEVSET-formal and DEVSET-informal) and testing on all test set data (both TESTSET-formal and TESTSET-informal)
- Subtask A - crossFormal: training on data from DEVSET-formal only, and testing separately on data from formal texts (TESTSET-formal) and from informal texts (TESTSET-informal)
- Subtask B - crossInformal: training on data from DEVSET-informal only, and testing separately on data from formal texts (TESTSET-formal) and from informal texts (TESTSET-informal).

While all tasks are oriented to investigate how challenging can it be to PoS-tag spontaneous speech data, the cross ones are especially useful for validating the hypothesis that some differences occur between the tagging of formal conversations and that of informal conversations. As we will see in section 5 and 6, this hypothesis is partially confirmed by results. Some example useful to draw the difference among the registers is provided in the next section.

3 Datasets

All the data provided for the KIPoS task are extracted from the KIP module (see Section 1),

Dataset	Register	Speakers	Turns	Tokens
DEVSET	Formal	5	1.998	13.864
	Informal	11	3.804	19.259
TESTSET	Formal	2	459	3.642
	Informal	2	582	3.532

Table 1: The sizes of the dataset.

which includes various communicative situations occurring in the academic context. As explained in detail in (Mauri et al., 2019), the recordings involve five different types of interactions, each of which is assigned for the aims of KIPoS either to the section of formal texts or to the section of informal texts (mainly on the basis of the relationship between the participants, i.e. asymmetrical vs. symmetrical).

The KIP corpus structure can thus be outlined as follows:

- Formal dataset:
 - lessons
 - office hours
 - oral examinations
- Informal dataset:
 - semi-structured interviews
 - casual conversations.

Below are examples of formal (1) and informal (2) texts.

(1)¹

BO088: una volta che carlo magno
conquistò l'italia fu permesso ad
anselmo di tornare eh a mantova
BO088: nel settecentosettantaquattro
BO088: ehme così po pote' riprendere
la sua attività prima eh di creazione
della biblioteca
BO088: perché secondo appunto l'uso eh
delle biblioteche eh
BO088: medioev medievali diciamo prima
eh vi era
BO088: mh la insomma la raccolta di
libri dall'esterno

(2)²

BO003: povero cristo sono andata a
beccare questo
BO002: ma poi scusa il piu' carino di
tutti lo cornifichi
BO003: si' si' si' esa poi secondo me
lui e' il piu' carino di tutti

¹KIP Corpus, BOC1001, oral examination

²KIP Corpus, BOA3001, casual conversation

BO003: cioè tra per i miei gusti tra il
gruppo
BO002: no eh
BO002: carino sia
BO002: di viso ma anche
BO003: poi e' anche il piu' si' si' si'
e' cornificatissimo non cornificato

Both excerpts feature spontaneous speech phenomena, such as fillers, repetitions and reformulations. However, example 1 shows several characteristics of formal styles, either cross-linguistically shared (e.g. clausal subordination, passive construction, abstract and specific terms) or language-specific (e.g. existential construction with *vi* as pre-copular proform); while example 2 displays various features which are typical of informal styles, such as simple sentence structure and pragmatically-marked word orders (e.g. *il più carino di tutti lo cornifichi*), multi-functional words (e.g. *carino*), colloquialisms (e.g. *povero cristo*, *beccare*, *cornifichi*, *cornificato*), elatives (e.g. *cornificatissimo*), deictics (e.g. *questo*, *lui*) and discourse markers (e.g. *cioè*, *scusa*).

All speakers were informed of the aims of the project, agreed to the recording and signed a consent form.

The set of data exploited for KIPoS precisely consists of around 200K tokens, corresponding to approximately one-third of the whole KIParla corpus, with an equal proportion of informal and formal speech data.

For the purposes of KIPoS, the UDpipe trained on all the treebanks available for Italian within the Universal Dependencies repository³ has been applied on this 200K tokens portion of the KIParla corpus. Among these data, approximately 30K tokens have been submitted to a careful manual check and correction⁴ and released as training sets of the KIPoS task (i.e. DEVSET–formal and

³<https://universaldependencies.org/it/index.html>

⁴We thank three students for their precious help: Filippo Mulinacci, Martina Pittalis and Roberto Russo of the Department of Modern Languages, Literatures and Cultures of the University of Bologna.

Team	Affiliation
UniBO	FICLIT – University of Bologna
UniBA	University of Bari "Aldo Moro"
KLUMSy	Friedrich Alexander Universität Erlangen-Nürnberg & Universität Stuttgart

Table 2: The teams which participated to KIPoS and their affiliation.

DEVSET–informal). From the remaining automatically annotated data, we extracted the formal-TESTSET and informal-TESTSET, and we also manually checked and validated them. Finally, we released as a silver standard (i.e. SILVERSET) the remaining data. They have been also made available together with the other data⁵ to be used for training participants’ systems.

3.1 Annotation

As far as the annotation is concerned, for the purpose of the task, the original orthographic transcriptions were provided in a tab-delimited .txt format. Three are the main identifiers we used in this format, respectively indicating the conversation (alphanumeric), the speaker’s ID (alphanumeric) and the position of the turn (numeric) within the context of the conversation. For instance, the example below includes the first three turns of the conversation "BOD2018"⁶, in which three different speakers are involved ("1_MP_BO118", "2_MP_BO118" and "3_AM_BO140"):

```
# conversation = BOD2018
# speaker = 1_MP_BO118
# turn = 1
# text = dovresti parlarmi della tua casa
1 dovresti AUX
2-3 parlarmi VERB_PRON
2 parlar VERB
3 mi PRON
4-5 della ADP_A
4 di ADP
5 la DET
6 tua DET
7 casa NOUN

# conversation = BOD2018
# speaker = 2_MP_BO118
# turn = 2
# text = attuale
1 attuale ADJ
```

⁵All the data annotate for KIPoS are available at <https://github.com/boscoc/kipos2020>, with the licence and the annotation guidelines.

⁶The alphanumeric code used to name the KIP’s conversations provides information about the city in which the data has been collected (BO= Bologna, TO=Turin) and the kind of interaction (A1=office hours, A3=free conversation, C1=exams, D1=lessons, D2=interviews). For example, BOD2018 is a semistructured interview recorded in Bologna.

```
# conversation = BOD2018
# speaker = 3_AM_BO140
# turn = 3
# text = mh sì
1 mh PARA
2 sì INTJ
```

The format and the labels for tagging the part of speech of the KIPoS data are compliant with that provided in the Universal Dependencies Italian treebanks. Data were indeed released in a CoNNL-U - like format, but which only includes the three first columns of it, separated by tab keys as usually. For a detailed list and description of the tagset used in KIPoS datasets, see the Appendix at the end of this paper.

3.2 Tokenization Issues

For what concerns words including multiple tokens, in the data released for the development and training of participant systems (DEVSET–formal and DEVSET–informal), we annotated their compound and splitting both. See for instance, in the first turn of the example above lines 2-3, 2 and 3: a verb with clitic suffix occurs and it is annotated as a compound in line 2-3, while its components, i.e. the verb and the clitic, are separately annotated on line 2 and 3 respectively.

In contrast, for the purpose of the evaluation, the format applied on the test set (TESTSET–formal and TESTSET–informal) only includes a word for each line, regardless of the fact that a word may be composed of more than one token. This makes the format of the test set slightly different from that used in the development data, but more compliant with the evaluation scripts and procedures. An example of this format follows, which consists in the first turn of the example above:

```
# conversation = BOD2018
# speaker = 1_MP_BO118
# turn = 1
# text = dovresti parlarmi della tua casa
1 dovresti AUX
2 parlarmi VERB_PRON
3 della ADP_A
4 tua DET
5 casa NOUN
```

Task	DEVSET	TESTSET	Team	Score
<i>Baseline (from POSTWITA)</i>				0.9319
Main	formal and informal	formal	UniBO	0.934880
			KLUMSy	0.875629
			UniBA	0.815819
		informal	UniBO	0.911316
			KLUMSy	0.882368
			UniBA	0.793684
Task A	formal	formal	KLUMSy	0.873672
			UniBA	0.787311
		informal	KLUMSy	0.875789
			UniBA	0.757895
Task B	informal	formal	KLUMSy	0.878144
			UniBA	0.771101
		informal	KLUMSy	0.881053
			UniBA	0.775000

Table 3: The official scores achieved by participants for the three subtasks (Main, Task A and Task B), by training systems on both or one of the datasets provided for development (DEVSET–formal and DEVSET–informal), on the TESTSET–formal and TESTSET–informal (best scores for each subtask in bold face).

In this example, the verb with clitic suffix "parlarmi" (speak to me) has been annotated as a compound on a single line, i.e. line 2.

4 Evaluation measures

For the KIPoS task a single measure has been used for the evaluation of participants' runs, i.e. accuracy, which is defined as the number of correct Part-of-Speech tags assignment divided by the total number of tokens in the gold TESTSET. The evaluation metric will be based on a token-by-token comparison and only a single tag is allowed for each token.

The evaluation is performed in a black box approach, where only the systems output is evaluated.

5 Participation and Results

As depicted in table 3, where the main task and the two subtasks results are presented at glance, three teams submitted their runs for KIPoS (see table 2 for their affiliation). Nevertheless, one team participated to the main task only, while the other two provided results for Task A and B too.

The three teams applied different approaches. UniBA team used a combination of two taggers implementing two different approaches, namely stochastic Hidden Markov Model and rule-based. UniBO applied a fine-tuning approach to Part of

Speech tagging that is based on a pre-trained neural language BERT-derived model (UmBERTo) and an adapted fine-tuning script.

KLUMSy used a tagger based on the averaged structured perceptron, which supports domain adaptation and can incorporate external resources for dealing with the limited availability of in-domain data.

The overall higher accuracy has been achieved in the main task by the UniBO team on the TESTSET-formal. The availability of a larger training corpus for the main task, which includes the DEVSET-formal and the DEVSET-informal both, and the results calculated on both the portions of the TESTSET allowed, as expected, the achievement of the KIPoS overall best score. This is confirmed also by the fact that all teams provided their best runs in it, for formal and informal register both. Even if the official submission of UniBO did not include the runs for Task A and B, the results it provided in its report (Tamburini, 2020) show indeed that also this team has ranked worst in Task A and B than in the main one. More precisely, for Task A, it achieved 0.8647 accuracy on TESTSET-formal and 0.8316 on TESTSET-informal, while in Task B it achieved 0.8974 on TESTSET-formal and 0.8952 on TESTSET-informal.

As far as the other teams are concerned, UniBA

provided in its report (Izzi and Ferilli, 2020) also the results achieved using a version of the TESTSET where a few errors detected after the official evaluation has been fixed. This allowed a small improvement in their scores (e.g. in the main task, +0.0078 for formal and +0.0056 for informal register).

The KLUMSy team provided the best runs for both registers in Task A and B, but in its runs, because of a misunderstanding of the guidelines about the annotation of contractions in the TESTSET (which is slightly different with respect to the DEVSET), a certain amount of mis-tagged tokens occurred. After they were fixed, also the scores of this team were improved (with an increase that varies from 0.0456 to 0.0187) with respect to the official ones reported in table 3, as described in the report of this team (Proisl and Lapesa, 2020).

Considered that the PoS tagging is a task mostly solved, it is not surprising that the participants' scores are quite high and close for all the tracks. The larger difference observed between the best and the worst score is indeed 0.126, and it is referred to Task B on TESTSET-formal.

Given the peculiarity of oral text on which KIPoS is focused, it seems not especially meaningful a comparison of our results with state-of-the-art Pos taggers results for the written standard language. A more interesting comparison can be instead developed with respect to the scores achieved within the PoSTWITA task (Bosco et al., 2016) on written texts extracted from social media. This genre is indeed often considered in between written and oral, sharing some feature with the former and some with the latter. Using the best PoSTWITA task accuracy score (0.9319) as our *baseline* (see table 3), we can observe that the best scores achieved in KIPoS are in line with this result. This confirms the hypothesis that oral text can be considered as almost equally hard to be morphologically tagged than social media.

As far as the distinction between formal and informal conversation drawn in the KIPoS datasets is concerned, a general trend of better scoring in formal data tagging can be observed, but some meaningful difference among participant systems occurs. For all subtasks UniBO best scored in formal text, while KLUMSy did the same in informal data. UniBA achieved instead its best scores on TESTSET-formal with the exception of Task B where its score for the informal test set is a little

bit (0.0038) higher than that for the formal one. Focusing on the cross subtasks A and B, we can moreover notice that systems were not equally influenced by the type of data exploited for training: UniBO provided best scores against TESTSET-formal also when trained on DEVSET-informal (Task B), while KLUMSy provided best scores against TESTSET-informal also when trained on DEVSET-formal (Task A). UniBA seems instead slightly more influenced by the features of data used in training.

6 Discussion and Conclusion

The results described in this report can be only considered as preliminary. First of all, KIPoS is the first edition of a task about PoS tagging of spontaneous speech for Italian and there aren't other results about this kind of task for the same language to be compared with. Second, the corpus used for KIPoS has been newly released for the purpose of the task and never used before. Participants provided some useful feedback about errors occurring in the DEVSET and TESTSET, but some further check should be applied for improving the quality of data. Finally, only three participants submitted their runs (and only two provided official runs for cross-genre tasks). Even if PoS tagging is among the tasks which are considered as mostly solved in literature, only a larger participation may allow a meaningful comparison among different approaches and results.

Nevertheless, the KIPoS task produced the valuable result of making available a novel resource for the study of spoken Italian and for the advancement of NLP in this area. It can be of great relevance for the investigation of both spontaneous speech phenomena and sociolinguistic variation, but also e.g. in the development of chatbots and vocal recognition systems.

In particular, the insights gained within the context of this Evalita evaluation campaign for PoS tagging can pave the way for further investigating actual speech data. They provide a solid foundation for our future research also in the direction of more detailed morphological analysis and syntactic parsing, especially within the framework of Universal Dependencies where we would like to release the KIPoS dataset in the near future.

7 Acknowledgments

The construction of part of the corpus has been possible thanks to the financing of the *Fondazione CRT* under the *Erogazioni ordinarie 2018* program. The KIParla corpus has been made possible thanks to SIR Project 'LEAdHOC' (n. RBSI14IIG0), funded by MIUR. We would like to thank also the students from our BA and MA courses at the Universities of Bologna and Torino, who participated in collecting and transcribing the data.

References

- Giovanna Alfonzetti. 2002. *La relativa non-standard. Italiano popolare o italiano parlato?* Centro di Studi Filologici e Linguistici Siciliani, Palermo.
- Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro. 2020. EVALITA 2020: Overview of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.
- Cristina Bosco, Fabio Tamburini, Andrea Bolioli, and Alessandro Mazzei. 2016. Overview of the EVALITA 2016 Part Of Speech on TWitter for ITALian task. In *Proceedings of Evalita 2016*.
- Cristina Bosco, Silvia Ballarè, Massimo Cerruti, Eugenio Gorla, and Caterina Mauri. 2020. KIPoS @ EVALITA2020: Overview of the Task on KIParla Part of Speech tagging. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.
- Giovanni Luca Izzi and Stefano Ferilli. 2020. UniBA@KIPoS: A Hybrid Approach for Part-of-Speech Tagging. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online. CEUR.org.
- Per Linell. 2005. *The written language bias in linguistics: its nature, origins and transformations*. Routledge, London – New York.
- Caterina Mauri, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti, and Francesco Suriano. 2019. KIParla Corpus: A New Resource for Spoken Italian. In Raffaella Bernardi, Roberto Navigli, and Giovanni Semeraro, editors, *Proceedings of the 6th Italian Conference on Computational Linguistics (CLiC-it 2019)*, Online. CEUR.org.
- Daniela Mereu. 2019. *Il sardo parlato a Cagliari*. Franco Angeli, Milano.
- Thomas Proisl and Gabriella Lapesa. 2020. KLUMSy@KIPoS: Experiments on Part-of-Speech Tagging of Spoken Italian. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online. CEUR.org.
- Manuela Sanguinetti, Cristina Bosco, Alessandro Mazzei, Alberto Lavelli, and Fabio Tamburini. 2017. Annotating Italian social media texts in Universal Dependencies. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 229–239.
- Manuela Sanguinetti, Cristina Bosco, Alberto Lavelli, Alessandro Mazzei, and Fabio Tamburini. 2018. PoSTWITA-UD: an Italian Twitter Treebank in Universal Dependencies. In *Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018)*, pages 1768–1775.
- Fabio Tamburini. 2020. UniBO@KIPoS: Fine-tuning the Italian “BERTology” for PoS-tagging Spoken Data. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online. CEUR.org.