IJCoL
Italian Journal of Computational Linguistics

9-2 | 2023
Italian Journal of Computational Linguistics vol. 9, n. 2 december 2023

#DEACTIVHATE: An Educational Experience for Recognizing and Counteracting Online Hate Speech

Alessandra Teresa Cignarella, Simona Frenda, Mirko Lai, Viviana Patti and Cristina Bosco

Electronic version
URL: https://journals.openedition.org/ijcol/1199
DOI: 10.4000/ijcol.1199
ISSN: 2499-4553

Publisher
Accademia University Press

Electronic reference
Alessandra Teresa Cignarella, Simona Frenda, Mirko Lai, Viviana Patti and Cristina Bosco, "#DEACTIVHATE: An Educational Experience for Recognizing and Counteracting Online Hate Speech ", *IJCoL* [Online], 9-2 | 2023, Online since 01 March 2024, connection on 10 October 2024. URL: http://journals.openedition.org/ijcol/1199 ; DOI: https://doi.org/10.4000/ijcol.1199

#DEACTIVHATE: An Educational Experience for Recognizing and Counteracting Online Hate Speech

Alessandra Teresa Cignarella*
Università di Torino

Simona Frenda**
Università di Torino

Mirko Lai†
Università di Torino

Viviana Patti‡
Università di Torino

Cristina Bosco§
Università di Torino

The possibility of raising awareness –especially in young generations– about misbehavior online such as hate speech, could help society to reduce its impact, and thus, its negative consequences. In the last years, the Computer Science Department of the University of Turin has designed various technologies that support educational projects and activities in this perspective. In this paper, we describe the creation of a laboratory called #DeactivHate*, specifically designed for secondary school students. The laboratory aims at countering hateful phenomena online and making students aware of technologies that they use on a daily basis. We describe the teaching experience of the first year of life of the laboratory held in different high school classes between April 2021 and March 2022 and the outcomes of the technologies and activities tested. In this extended version of the paper, some sections are especially devoted to observe and analyze the impact of the course on students and their perspective on the laboratory.*

Warning: This work contains words and expressions that could be considered vulgar or offensive to varying degrees. Some discriminatory expressions are included in our written examples (used in our teaching as samples of real online hate speech). We emphasize that all authors of this paper are deeply involved in activities to counter the spread of online hatred and do not condone the use of such expressions in any way.

1. Introduction

Recently, the presence of digital technologies has enormously grown, with a strong impact on our daily lives. Digital spaces and social media have become a privileged channel for communication, information exchange and socialization, simultaneously

* Computer Science Department. E-mail: alessandrateresa.cignarella@unito.it
** Computer Science Department. E-mail: simona.frenda@unito.it
† Computer Science Department. E-mail: mirko.lai@unito.it
‡ Computer Science Department. E-mail: viviana.patti@unito.it
§ Computer Science Department. E-mail: cristina.bosco@unito.it
Alessandra Teresa Cignarella and Simona Frenda equally contributed to this work.

frequented by millions of people. Along with the new relational opportunities to access knowledge, misbehavior, such as hate speech, have also gained increasing visibility and virality. In spite of a causal link between hate speech and crime is difficult to prove, the risk of offenses and effects on victim's psychological and physical well-being have been clearly demonstrated in psychological and social studies (Nadal et al. 2014; Fulper et al. 2014). Their extreme consequences might be suicide, especially among adolescents, as suggested by recent studies examining the connection between cyberbullying and suicidal behavior among adolescents in the U.S. (Nikolaou 2017). To prevent such scenarios, some awareness-raising projects in schools are being carried out by NGOs in Italy, such as Amnesty International[1] or Cifa ONLUS[2].

The *Commissione Orientamento e Informatica nelle scuole*[3] of the Department of Computer Science of the University of Turin supports a manifold of activities with the main goal of creating a link between schools and academia, also in the context of the national project *Piano Lauree Scientifiche* (PLS)[4].

The members of the CCC (Content-Centered Computing) group of the Computer Science Department of the University of Turin are especially active in the investigation and counteraction of online hate speech[5]. They have led and participated in several hate-speech-related projects, including "Contro l'odio"[6] (Capozzi et al. 2020), a joint effort of some non-profit entities and the University Aldo Moro of Bari, aiming at monitoring hate speech against minorities in Italy.

Within the current experience, as members of the CCC group, we adapted an annotation platform specifically to support educational activities and reflect on the importance of a conscientious communication. In this perspective, the idea of #DEACTIVHATE takes hold. The name of the laboratory is a portmanteau generated by combining the verb '*to deactivate*' and the substantive '*hate*' with the addition of the pound sign '#' in front to resemble a hashtag, and thus, to establish a link with social media. The activities, aimed at secondary school students, are articulated in three main modules with the purpose of:

1) raising awareness about the social problem of hate speech, encouraging the reflection also on microaggressions, stereotypes and prejudices;
2) stimulating the so-called *computational thinking* and the study of linguistic devices that social media users exploit to offend or to express hate against victims online (hashtags, emoticons, or figures of speech);
3) introducing high schoolers to tools based on NLP (Natural Language Processing) and their application for incentivizing a more conscious use of technology.

To achieve these goals, we designed a series of activities that include: the analysis of the online spread of the hate speech, by investigating personal profiles and social networks; the linguistic analysis and annotation of hateful messages on the "Contro l'odio" platform the manual identification of hate speech in Italian texts, asking students to behave as an automatic classifier; the creation of two types of automatic classifiers in Python. These activities, have been distributed in 5 meetings (lasting 2 hours each) for each class, between April 2021 and March 2022, for a total of 10 hours per class and

1 http://di.unito.it/silencehateitaly.
2 http://di.unito.it/iorispetto.
3 http://di.unito.it/orientamentoscuole.
4 https://www.pianolaureescientifiche.it/.
5 http://hatespeech.di.unito.it/.
6 https://controlodio.it/.

replicated in three different editions. In Table 1 we report the details of all the three editions of the laboratory.

Table 1
Details of the editions of the laboratory held during the last year. *(In the second edition, we taught the laboratory to two different classes of the same grade α and β; additionally y.o. stands for 'years old').*

edition	mode	period	type	grade and age	n° of students
1st	online	April-June 2021	humanities	III (15/16 y.o.)	21
				IV (16/17 y.o.)	14
2nd	in person	October-December 2021	humanities	IIIα (15/16 y.o.)	20
				IIIβ (15/16 y.o.)	26
3rd	online	February-March 2022	technical	III (15/16 y.o.)	25
				IV (16/17 y.o.)	20
				V (17/18 y.o.)	19

2. Related Work

A popular workshop series on the topic of "Teaching NLP" has been recently held on its fifth edition at NAACL-HLT 2021 (Jurgens et al. 2021), where the participants discussed and shared experiences on a variety of important issues such as: teaching guidelines, teaching strategies, adapting to different student audiences, resources for assignments, and course or program design. The most valuable lesson learned is that it is important to create materials that describe NLP, not only for learners at a university/college level, but also for younger learners with different educational backgrounds. In this respect, the experience of Sprugnoli et al. (2018) is a great inspiration for the work with schools in Italy, where the authors – although with a different goal than ours – started a project with NLP and pupils from Italian schools, aged 12-13.

That experience was chiefly dedicated to the study of cyberbullying among preteens and the creation of a corpus of WhatsApp threads in the context of the CybeR-bullying EffEcts Prevention activities (CREEP) project. Our idea of starting a project that could bring NLP to high schoolers and that, at the same time, could introduce the themes of hate speech, microaggressions, and discrimination by eliciting personal experiences and students' opinions, is somehow in continuity with that experience.

A second work, of great relevance for the creation of our experience, has been Pannitto et al. (2021), in which the authors point out for the first time the absence of *(computational) linguistics* in the curricula of Italian high schools, and the consequent need to choose computational linguistics only as part of a university degree. Furthermore, the authors highlight that NLP is, indeed, at the core of many tools young people use in their everyday life, and having almost zero knowledge of this field makes the use of such tools less responsible than it could be (the same purpose inspired also Bioglio et al. (2019)). The authors have also been the first to organize a dedicated workshop for Italian, aimed at raising awareness of Italian students aged between 13 and 18 years regarding the subject of NLP (Messina et al. 2021).

Additionally, the idea of creating some playful and meaningful activities regarding NLP and the themes of hate speech for high schoolers, are in line with the concept of '*gamification*', which lately has been applied to many linguistic annotation tasks, as an alternative to crowdsourcing platforms to collect annotated data in an inexpensive way (Bonetti and Tonelli 2020), such as "Contro l'odio" annotation platform.

3. #DeactivHate

The main goals of #DeactivHate are: 1) raising awareness about misbehavior online, such as hate speech, eliciting also personal experiences, 2) stimulating computational thinking and linguistic observation of hateful messages, and 3) promoting a more conscious use of technology through a greater awareness of how it works. To reach these objectives, we articulated three modules as described below.

3.1 Introducing Hate Speech

The first module aims to provide students with a definition of hate speech and to improve their awareness of the features and facets of this phenomenon, among which, e.g., virality and multi-modality. Hate speech is often mistaken for a general insult rather than a specific phenomenon "connected with hatred of members of groups or classes of persons identified by certain ascriptive characteristics (e.g., race, ethnicity, nationality)" (Brown 2015).

This session begins with an ice-breaking activity in which students present themselves through an image they selected on the Web, depicting an aspect of their identity (see Figure 1). We then asked them to tell whether they were ever attacked or stigmatized for this characteristic.

Figure 1
Example of a Jamboard used in the class.

In this way, we guided the class in drawing a distinction between **non-ascriptive** identity traits (e.g., political belief, style of dressing) and **ascriptive**[7] ones (e.g., ethnicity, sexual orientation, skin color) (Reskin 2005). The idea behind this activity is twofold: i) it connects issues, such as hate speech and racial microaggression (Sue 2010), to students' lives; ii) it helps to distinguish the spreading of discriminatory contents[8] from generic

7 Qualities beyond the control of an individual.
8 The definition of hate speech we referred to is the one codified by The Council of Europe: "the term 'hate speech' shall be understood as covering all forms of expression which spread, incite, promote or justify

insults. The module ends with an assignment: students had to find at least one public figure who had been a victim of online discrimination, providing one or more hateful messages as an example, and a counter-narrative response that makes the author reflect about their offensive message.

3.2 "If I Were a Classifier..."

The second module aims to introduce two main aspects: the importance of the detection of hate speech, and the logic underlying the classification of hateful content that can be manually done (as in the second section of this module) or automatically obtained by applying supervised approaches (as in the third module described in Section 3.3). This module has therefore been organized in two sessions respectively focused on the classification of hate speech and related phenomena, and on their annotation in linguistic corpora.

Within the first session of this module, we discuss the homework by asking each student to comment the messages they found to be hateful by looking at their own social media pages, and we define together the type of attack and the linguistic characteristics that makes them hateful. The variety of examples in which different forms of hate can be expressed, e.g., towards different targets, led to the introduction of a taxonomy of discrimination, in which e.g., misogyny, homophobia, sexism, etc... are organized. As expected, the group discussion, that concludes this first session, shows how subjective the perception of the observed phenomena is, and that it is necessary to find a common scheme that is shared by all and that allows mutual understanding when talking about these concepts. From the idea of a common schema, we move to the idea of defining a formal schema to be used for classifying examples, such as the examples contained in a corpus, which is the object of the following session of this module.

In the second session of this module, after a brief introduction on what corpora are and the role they play today used in language technologies, an annotation scheme is introduced which consists of four categories (described in detail in Section 3.2.1). Following, students are invited to perform an annotation task, asking each of them to express the judgment on at least 30 tweets on a dedicated annotation platform, as will be described in the next Section.

3.2.1 The Annotation Platform

For this purpose, we adapted the data annotation platform[9] within the "Contro l'odio" project for supporting these educational activities. This web application, built using PHP, MySQL, and JavaScript[10], preserves the student's annotation history by using a passwordless authentication based on the so called "magic link" sent to the user's email. This method has the twofold advantage of not requiring the student to register to the platform and of preventing ourselves to save the student's email or other personal data. It then ensures the annotation anonymity and satisfies the requirements of General Data Protection Regulation (GDPR), as a desired consequence.

The home page of the web application consists of a dashboard with the annotation guidelines and basic information about the student's activity.

racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance" (Recommendation No. R (97) 20).

9 https://didattica.controlodio.it/.

10 https://github.com/mirkolai/DEACTIVHATELab.

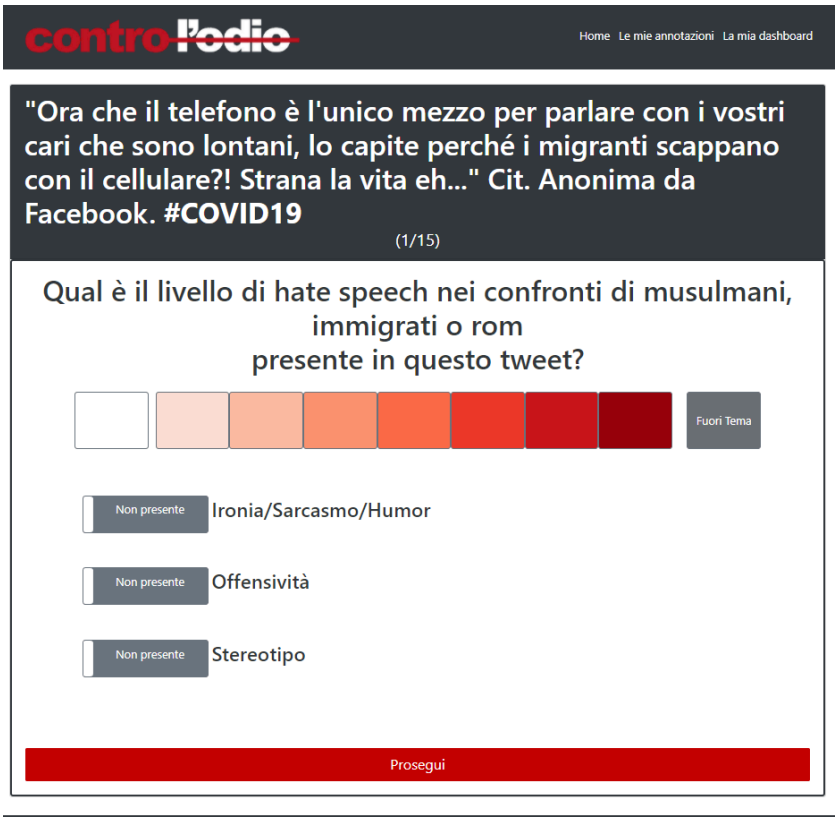


Figure 2
Data Annotation Platform

When a session starts, according to the scheme introduced above, the student could annotate the level of hatefulness of a tweet through a 7 square scale filled with a color scale from *Watusi* to *Sangria* as shown in Figure 2. Two additional squares, respectively filled with *White* and *Mid-Gray*, allow stating the absence of hate or to consider off-topic the content of the tweet. Finally, three toggle switches (on/off button) were added to check the presence of 'irony/sarcasm/humor', 'offensiveness', and 'stereotype', giving the students the possibility to reflect about the ways in which users spread hate online.

For each student, the platform shows the number of completed annotation sessions (each session consists of annotating 15 tweets) and the level of agreement achieved (expressed in percentage) with respect to the results generated by the automatic model realized in the "Contro l'odio" project (Capozzi et al. 2020). Through this comparison, we provide the basis for a discussion about the fallibility of automatic systems. The introduction of the measures of inter-annotator agreement is instead linked to the group discussion on subjectivity in the perception of hatred (in the first session of this module), based on the comparison of the annotation given by each student with that of their classmates, but also reinforced by another activity performed by students.

During the annotation task, students were indeed asked to fill in a common table with the tweets that most impressed them, whether for their offensiveness, humorous intent, or because they were the most difficult to annotate. By discussing the annotation results with the students, we introduce therefore the module's latest core concept,

namely the **agreement** among different annotators. We present some metrics typically used to calculate it and explain some of the best practices that recently emerged in the usage of annotated corpora inside the Natural Language Processing community, such as the inclusion of minorities and different perspectives in the development of corpora to avoid bias (Basile 2020).

3.3 My First Classifier

The third module aims at stimulating computational thinking, that is the transition from linguistic findings to more computational observations. The activity of annotation has, indeed, given the opportunity to reflect on how users tend to verbally express hate online, and on how minorities are represented through stereotypes. To incentivize this transition, in this module, we propose two activities:

A. To mark in each tweet the textual span that could make a classifier aware of the presence of hate speech creating a list of word n-grams;
B. To develop two automatic classifiers (supervised and unsupervised) exploiting this list of word n-grams.

Before beginning the activity A, we asked students to justify their choice of tweets selected in the previous exercise as the most impressive, in the previous module. Some tweets triggered discussions on what should be considered hate speech or not, and the doubts have been resolved by resorting to the definitions previously provided and integrated in the annotation guidelines.

The most controversial tweets report aggressive events or racial propositions; and, for this reason, they were perceived as hurtful by the majority of the students:

(i) *Autobus per i bianchi e altri per i migranti. Non si parla dell'apertheid del Sudafrica né del periodo di segregazione negli Stati Uniti, ma di una proposta della Lega per la provincia di Bergamo. L'Italia non è un paese razzista ma nel 2020 questo è ciò di cui si discute. URL*[11]

Others triggered interesting linguistic reflections, such as:

(ii) *Peccato che non sbarcano povere famiglie africane, ma solo mafia nigeriana, ex galeotti tunisini, stupratori senegalesi, terroristi dell'Isis dalla libia, tutti criminali robusti 1.80 di altezza, pronti a spacciare droga, violentare le nostre donne, cannibali e assassini.*[12]

In these, the students retrieved specific figures of speech such as sarcasm, rhetorical questions and analogies, and also strong words that reflect the social biases towards the minorities. In activity A, all the words and expressions that could make the message hurtful have been collected in a list of n-grams of words called `our_lexicon` (Table 2). Following, the items of such list have been exploited by the classifiers to predict if a tweet contains hate speech or not.

11 Translation: *Buses for whites and others for migrants. There is no mention of South Africa's apartheid or the period of segregation in the United States, but of proposal by Lega for the province of Bergamo. Italy is not a racist country but in 2020 this is what we are discussing. URL.*

12 Translation: *Too bad that poor African families do not land, but only the Nigerian mafia, former Tunisian convicts, Senegalese rapists, ISIS terrorists from Libya, all heavy-weight criminals 1.80 tall, ready to sell drugs, rape our women, cannibals and murderers.*

13 Translation: Unigrams: resources, dirty, *godsend*, disgust, invasion, peddle. N-grams: closed harbours, send [them] away, defence of the fatherland.

Table 2
Examples from `our_lexicon`.

unigrams	risorse, sporchi, pacchia, schifo, invasione, spacciare
n-grams	porti chiusi, cacciarli via, difesa della patria[13]

For activity B, we created an interactive Python notebook using the *Colaboratory* platform provided by Google, as a similar initiative had successfully been carried out by Hiippala (2021) with a similar educational tool. To allow the students to use the notebook in spite of their computer skills, we elaborated some guidelines explaining even how to create a folder in Google Drive and how to import all the necessary materials inside of it. Among the required materials, we prepared the dataset using the tweets previously annotated by the students themselves. We proposed two types of classifiers:

1) unsupervised classifier based on the list `our_lexicon` for which if one of the selected grams are inside the text, the text is predicted as hateful;
2) supervised classifier based on Support Vector Machine algorithm using the list `our_lexicon` as main feature of the classification task.

The coding of the first classifier allowed students to gain confidence with some basics of Python; whereas the second one introduced them to the core of new technologies based on machine learning (see Figure 3). At the end of the activity, we observed together the performances of automatic systems and analyzed some of the tweets that were wrongly classified. This final step helped students to reflect on the limitations of machines and the important role of the linguistics in language-related technologies.

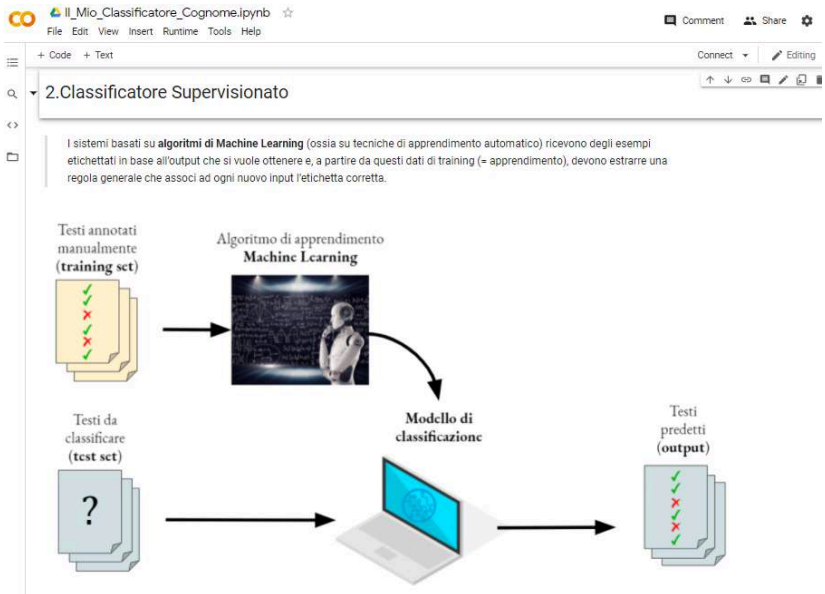


Figure 3
Screenshot of the Classifier Section on the Colaboratory Notebook.

4. Testing Students' Progress

After the first two editions of the #DEACTIVHATE laboratory (in April-June 2021 and October-December 2021 respectively) we published a paper describing these experiences (Frenda et al. 2021). The live presentation of such research at CLiC-it 2021[14] in Milan has generated many comments, some of which inspired the directions of development outlined in the current article. In particular, we decided to build tests for the assessment of the degree of assimilation of the main concepts covered in the course and the identification of the addressed topics that most interested the students.

Already in the first and second editions, we had planned to leave 10 minutes in the last hour of the laboratory for students to fill in a **survey** questionnaire. It contained mainly questions regarding the overall satisfaction towards the laboratory and the degree of interest in the topics of the course.

Within the third edition of the laboratory (February-March 2022), and after having received the feedback from other researchers engaged in teaching NLP, we decided to create the premises for an analysis of the outcomes of the laboratory to be spent also in a research perspective.

For this purpose, we have created a standard routine for the assessment that includes two steps: i) an assessment **test of prior knowledge** (PRE-TEST) to be administered before the beginning of the laboratory, and ii) a **test of final knowledge** (POST-TEST) delivered at the end of the 10-hour cycle of lessons. In the design of our questionnaires, we had three different types of questions:

1. True/false questions.
   These kinds of questions were easily evaluated as correct (1 point) or wrong (0 points).

2. Multiple choice questions and questions that require fairly short answers.
   The student did not answer, or the answer is wrong (0 points);
   The student tried to answer, or the answer partially ok, but it is incomplete (1 point);
   The answer is correct (2 points).

3. Open questions that require a more complex evaluation.
   We evaluated the brief textual productions on a correctness scale ranging from 0 to 5 points.

In Appendix A and Appendix B all the questions of both tests are reported in Italian, paired with their English translation. Below, we briefly show some questions from the PRE-TEST and POST-TEST, together with a statistical evaluation of the answers provided by students.

The PRE-TEST may be considered as an *ice-breaker activity* on the day of the first lesson, even before we introduced ourselves to the students. In this way, we were able to actively engage the participants right away, and ensure that the students remained unaffected by any new concepts we provided them before completing the PRE-TEST. Students were of course instructed to be honest and not to copy. They were also reassured that the test was meant for us –the instructors– to understand their level of

14 https://clic2021.disco.unimib.it/it/.

knowledge on certain topics, and not to grade them on the class register. Figure 4 shows a screenshot from the PRE-TEST.

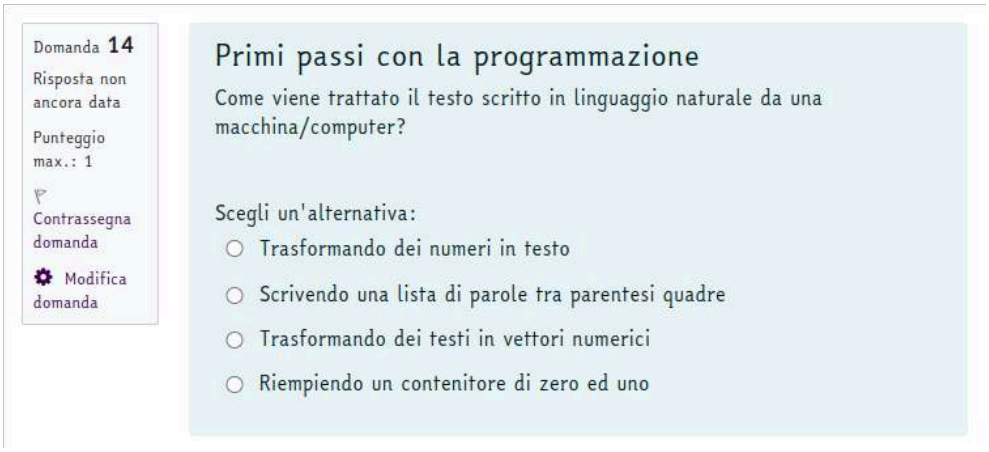


Figure 4
Screenshot of one question from the preliminary test.

The POST-TEST has been designed to be administered to students on the last lesson, or given the last day as homework with a hard deadline. It was presented to the students as a proper assessment test, the grade of which we report to the class council and in the class register, in order to encourage them to participate seriously and with commitment. Both questionnaires were held on the Moodle platform of our affiliation[15] and students were given a private access key for logging in.

4.1 Topics and Internal Guidelines for Grading

Both questionnaires are composed by questions related to the different topics and categories of concepts dealt with during the laboratory: (1) Sociology/Civics,[16] and Hate Speech (**C**); (2) Computational Linguistics (**CL**), and (3) Computer Science/Programming (**CS**). For the first topic, we posed questions such as: *What do you think Hate Speech is?* For the second topic a typical question was *Try to list a minimum of 2 applications of everyday life in which you think that the basis may be the use of technologies derived from computational linguistics*; and for the third topic one of the questions was: *Try to describe in your own words what an algorithm is*. We also asked the students whether they had some prior knowledge of programming with Python or with other programming languages.

As it can be seen from the Appendices A and B at the end of the paper, the questions in the PRE-TEST and in the POST-TEST are partially the same. Thanks to this overlapping design, comparing the answers provided in the first test with those given for the second one, we can see the progress made by the students during the activities. The evaluation of each answer depends on the type of question (true/false, multiple choice and open question) as described in Section 4.

15 `https://orientamento.educ.di.unito.it/`.
16 Subject typically thought in Italian schools: Educazione Civica.

Unfortunately, 10% of the students were absent either on the first lesson or on the last one and, therefore, did not take one of the two tests. We were not able to compute their progress according to the increment score and test their improvement. Their answers were partially taken into consideration for computing the aggregated results described in the next section, with some statistics drawn from the questionnaires outcomes.

5. Statistics: Evaluating Students' Progress

In this section, we focus firstly on the progress done by the students as detected by comparing their results in the PRE-TEST and POST-TEST; secondly we observe the acquisition of the three main categories of concepts on which the laboratory is centered, i.e., those related to Sociology/Civics and Hate Speech (C), Computational Linguistics (CL) and Computer Science/Programming (CS) (see also Section 4.1); and finally, we analyze the overall satisfaction of the students towards the laboratory.

In the computation of the statistics in the next three subsections, a normalization of students' replies' evaluation has been operated. Some adjustments were needed considering that some answers were evaluated using different strategies. More precisely, the statistics in Sections 5.1 and 5.2 are computed only regarding the 3rd edition of the laboratory, considering a total number of 64 students' replies. While, the results reported in Section 5.3 refer to all three editions, thus, resulting in 145 single answers.

5.1 Measuring the general improvement

The comparison between the answers given by the students in the overlapped questions of the PRE-TEST and POST-TEST allowed us to provide a general assessment of the students' progress determined by having attended the laboratory and carried out the activities we proposed.

In Figure 5, we show the statistic evaluation of the answers given by students of three different classes to the PRE-TEST and POST-TEST. Using box plots, in particular, we are able to observe the variance of the answers inside each class (represented by the length of the vertical lines) and the average of the grades obtained (the yellow line). By comparing the three box plots, it can be seen that only the answers of the students of the V class (Figure 5c) show an improvement in the final test (POST) with respect to the answers given in the prior one (PRE). On the other hand, almost all the students of the III class (Figure 5a) did some progress, but for 26% of them we reported negative grades.[17] In the IV class (Figure 5b), unfortunately, the improvement is lower than in other classes. Furthermore, the higher variance of results in the POST-TEST with respect to that shown in the PRE-TEST allows us to hypothesize that the key topics have been assimilated in very different degrees also by students of the same class. Differently, the performance of the students in the V class in the POST-TEST presents a very low variance (Figure 5c).

In order to better understand the meaning of the results of this analysis only based on a general assessment of the answers provided by the students for the PRE-TEST and POST-TEST, we provide more precise observations focused on the progress of students for what concerns the three different categories of topics, namely C, CL and CS in the following Section.

17 Having normalized the results, their values range from 0 to 1. With 'negative' we refer to values in the lower half of the considered range, thus below 0.5.

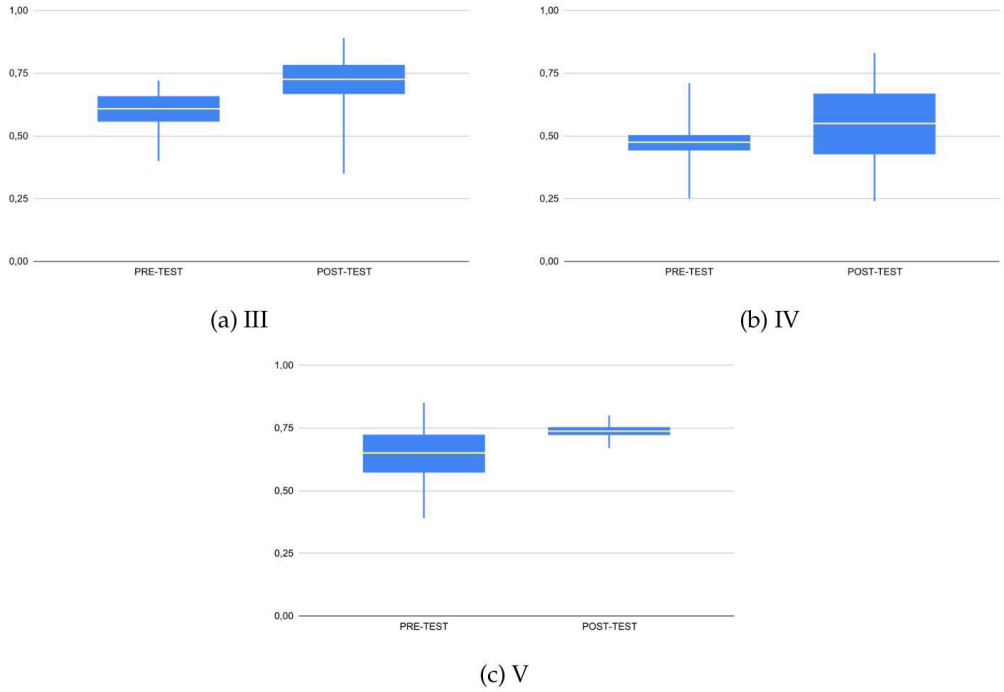


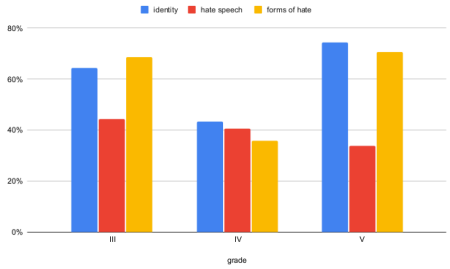
Figure 5
Box plots on the evaluation of PRE-TEST and POST-TEST.

5.2 Assimilation of the main concepts

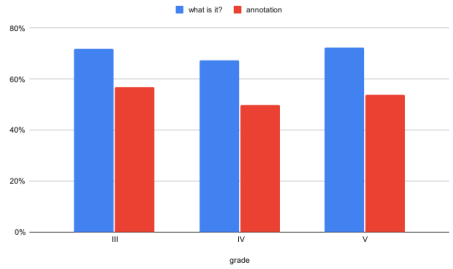
From Figure 5, we noticed that the IV class reported a lower progress and lower results than students in the III and the V class. In this section, we separately look at the assimilation of students regarding the three main topics of the #DEACTIVHATE laboratory: Sociology/Civics and Hate Speech (C), Computational Linguistics (CL) and Computer Science/Programming (CS), providing an explanation to such results.

In Figure 6 we report all the charts representing the results obtained evaluating the answers given by each class, considering the three topics. To better understand the progress of students with respect to these topics, we grouped the answers obtained by the students, taking into account the content of the questions in the preliminary and final tests. In particular, under the topic **C** we grouped the answers related to *identity*, *hate speech*, and *forms of hate*. In the topic of **CL**, are clustered we organized the answers in the categories *what is it?* and *annotation*. Finally, under **CS**, we collected the answers in *programming*, *algorithm* and *NLP*.

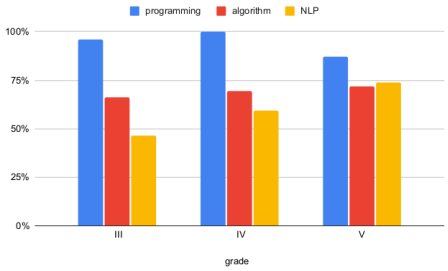
Observing in general all these charts we can notice that the IV class shows a low understanding in Civics and Computational Linguistics, and a good comprehension in Computer Science, mainly for the subtopic of *programming* in Python (Figure 6c). This is actually an expected result, due to the fact that Python is the main coding language studied by the students in that school year, so their programming skills are quite advanced with respect to the other two classes taken into account (who are studying C and Java instead).



(a) Sociology/Civics and Hate Speech (C)



(b) Computational Linguistics (CL)



(c) Computer Science (CS)

Figure 6
Graphics on the evaluation of comprehension on the three main topics of the laboratory.

Moreover, looking at each chart, it can be noticed that there are some specific subtopics that appear harder to learn for all the students of the three grades. In Civics (Figure 6a), for instance, the question regarding the presence or absence of hate speech in this social media text "Tomorrow if you have time, make me a list of the jobs that migrants have stolen from you and your children. I care about it." created much confusion. This is, indeed, a tricky question, because the post reports the negative stereotype that depicts immigrants as invaders who steal jobs from Italian people, but it is not explicitly expressing or inciting hate against immigrants.

In the Computational Linguistics part (Figure 6b), a question regarding the good practices of data annotation and inter-annotator agreement was perceived as the most difficult. The question is:

> If everyone in your class annotates the same text thinking about the binary classification with labels 0 and 1 (for example: this tweet contains a hate message (1), or it does not contain it (0)). Is it reasonable to think that there is a part of the class in agreement and a part of the class in disagreement. How does a computational linguist go about solving this problem? (POST-TEST, 7)

One of the answers given to this question, in the test took after the laboratory, was: "Using algorithms developed through supervised learning". This shows that some students were confusing the step of the creation of dataset (to reach a good agreement), with the step for the development of a system able to identify the hate in the text (using a supervised ML approach). Indeed, among the various answers given, something we have considered as valid is, for instance: "You look at the majority [of the annotations]

to [decide] if a message is hate speech or not.", in which the student, although quite concise, has gotten the concept of majority voting correctly.

Differently from previous topics, in Figure 6c, it can be observed how the quality of answers about *algorithm* and *NLP* increases gradually from the III class to the V class. We believe that this increment in correctness is linked to the scholarly experience of the students in the subject of Informatics. Therefore, the students that are familiar with the concepts of algorithms since more years, are able to understand better also the basic principles of AI's techniques and, thus, of NLP. About the *programming* subtask, the results reported by the students of the III and V classes are unexpected. On the one hand, the answers of the students of the III class proved an impressive increment in programming with Python, even if it was the first time that the students programmed in this language. On the other hand, the results of the V class are lower, although they studied Python for the entire school year before.

5.3 Students' satisfaction

Now we present the results of the survey regarding the degree of appreciation of the laboratory, mentioned in Section 4, that was dispensed to the students at the end of the laboratory. In particular, we analyzed the degree of satisfaction manifested by the students and the evolution of the main comments expressed throughout the three editions.

If the degree of satisfaction is computed exploiting the scale from 1 to 5 used in the survey, the evolution of the main comments have been analyzed extracting specific keywords for each of them (*technical problems*, *complain about subjects*, *lack of entertainment*, *complain about DAD*[18], *complain about manage of time*, and *positive comments*)[19]. To draw this evolution in Figure 8, we show the frequency of the group of keywords of each comment in every edition.

In Figure 7, the increment of satisfaction of the entire laboratory and especially of the contents provided to the students is displayed. A notable decrease of the two curves appears at the end of the first edition of #DEACTIVHATE. We speculate that the lower attention of the students could be due to the tiredness at the end of the school year, and the recurrent technical problems.

This aspect clearly emerges from Figure 8, where the *technical problems* are marked as one of the negative aspects experienced in the first edition. In particular, the low proficiency of the students of the Humanities schools about basic functioning of computers (as opposed to mobile devices) proved to be a relevant problem, that was also worsened by the hybrid mode used in the meetings of the first edition of the laboratory. This is further confirmed by the fact that, even if the laboratory in the third edition was also provided in a hybrid modality, the students from technical school –who are proficient with the use of computers– did not complain about such problems, but rather complained only about *DAD*. This justifies the slight decrease of the blue curve in Figure 7 at the beginning of the third edition (February 2022).

Looking in detail at Figure 8, it can be easily noticed that all the curves regarding some degree of complaint, present lower values during the second edition of the laboratory. In fact, that edition of #DEACTIVHATE was held *in praesentia*, and it is clear

18 Didattica A Distanza.

19 For instance, for *technical problems* we searched keywords like "technical problem" and "informatic room".

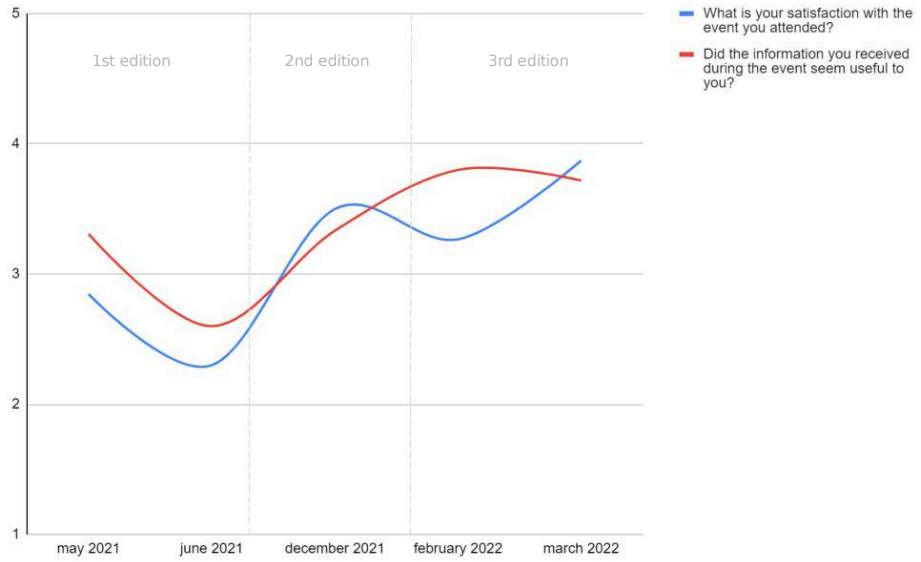


Figure 7
Degree of satisfaction expressed by the students in the three different editions.

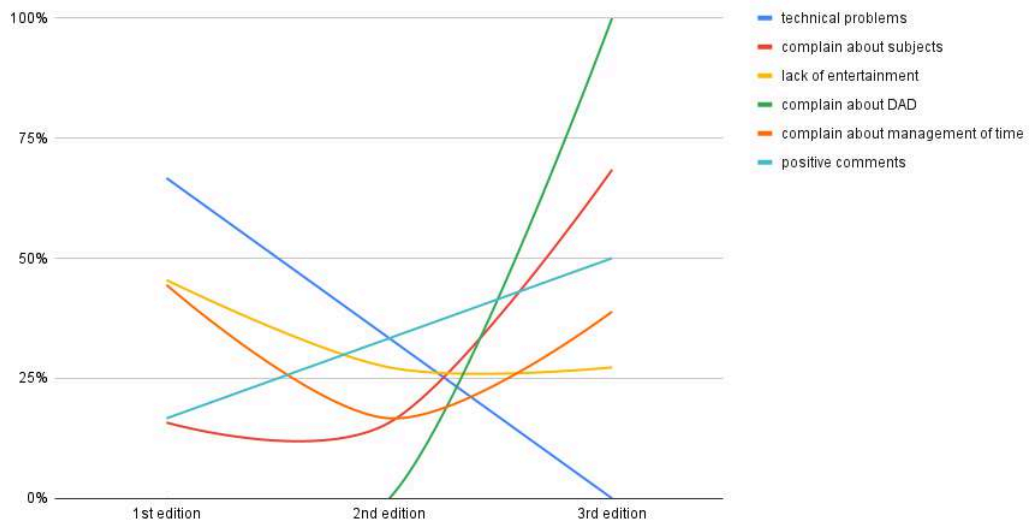


Figure 8
Evolution of comments thought the three editions.

from this chart that the majority of the issues such as *technical problems* (blue line), *lack of entertainment* (yellow line), *management of time* (orange line) tend to almost disappear. The same cannot be said about the *subjects* (red line) taught, that are perceived as less stimulating by the students from the technical school, who wrote comments complaining especially for the lack of more practical activities. For example:

> In my opinion, it would be useful to concentrate the introductory part in a single lesson (maybe also in oral form without inserting images on the Jamboard) and the civics part with sentences containing or not HS, in order to subsequently focus with more tranquility and more time on the programming parts.

Finally, observing the *positive comments* (cyan line), we can see that they present a constant increase from the first edition, thus show how our efforts in the amelioration of the laboratory's activities through the various editions have been fruitful.

6. What We Learned

Due to pandemic restrictions, we taught the 1st and 3rd edition of the laboratory through remote modality to two classes belonging to a secondary school of Humanities, and three classes from a Technical school (see details in Table 1). As described above, various resources and tools have been used (and created *ex novo*) to bring forward the educational activities in distant teaching mode, and later adapted for the 2nd edition for teaching *in praesentia*, also exploiting the computer rooms of the schools.

For each class, we organized the activities of the three modules (see Section 3) in five meetings of about two hours each. From the teaching viewpoint, the idea of organizing specific activities for each session helped to manage efficiently the available time. However, despite a smooth planning, due to the insurgence of technical problems, the activities could not be conducted as foreseen, lowering the general satisfaction of the students in some occasion (as seen in Figure 8). One of the most important lesson learned, is that the general perception of the students is much more positive in the edition held *in praesentia*, showing how –despite the good results that can be obtained online– **a live setting allows achieving better, quicker and more satisfactory results**.

We resorted to web applications to make up for the different devices and operating systems used by the students in their homes and in the computer rooms of their schools. And, in particular, for the online editions we used Google Meet, as it offers interactive tools such as virtual Jamboards, while in the live edition we just used the blackboard in the class, having adapted the activity in an *unplugged* modality. In the three editions, we used the Moodle learning platform provided by the University of Turin, that gave us the possibility to organize our activities, making available the necessary materials to students. Moreover, each meeting was supported by the use of slides for having visual and descriptive support.

In total, we taught the laboratory to 145 adolescents, divided into 7 classes coming from different schools (see Table 1), and with a different ethnical background. In general, as instructors, we noticed that **the students showed interest in the addressed subjects**, and we were positively surprised especially by the profoundness of some observations raised during the discussions. The students, indeed, were encouraged to share their opinions, doubts, and perspectives. These discussions made it clear that the students face problems related to technology and communication every day, sometimes even suffering the consequences. Hate speech is, indeed, a very sensitive issue and the perception of what is abusive or not, depends on the cultural background of each student. This fact, on the one side stimulated the debates, however, on the other side, it made it **difficult for us to find the *ideal* way to share complex concepts** and manage specific situations. Many students seemed to struggle in understanding the specific nuances of Hate Speech and its intensity or dimensions, and its correlation with stereotype, prejudice and offensive language. A feasible solution for addressing this point could be

receiving support in teaching from someone who has a more sociological/civics background, or that has a deeper knowledge of communication of such complex concepts.

Moreover, from our perspective, we noticed a **difference regarding commitment and participation** between students from the **Humanities** and those from the **Technical** school. The first were more actively engaged during the activities and discussions about the topics of Civics, Sociology and Hate Speech, and seemed very proficient in expressing their opinions and hold their stance during the oral debates. The latter were highly skilled with technical topics and programming, and showed more interest in the Computational Linguistics, Computer Science topics and hands-on coding on Google Colaboratory (as already seen in Section 5.3). Regarding this point, and paring it with also the results of Figure 6, where the assimilation of concepts is shown, a lesson learned is that it is **fundamental to know beforehand** the technical background and **the programming skills** of the students, in order to plan adequately the level of the coding part and better exploit the available time.

Finally, due to the **complexity of topics** such as annotation, inter-annotator agreement, creation of datasets for training systems, and machine learning algorithms, **paired with scarcity of time**, the students have often misunderstood the differences among the various steps of the pipeline. As a solution, in the next editions, it would be more productive to assess their comprehension after every step of the teaching and make sure all the students are learning in parallel.

7. Conclusion and Future Work

#DEACTIVHATE represents for Italian high schoolers a first step towards the introduction to subjects such as Linguistics and NLP, that are, for the most part, unknown in Italian high schools, in spite of their relevance in everyday technology. Indeed, this kind of laboratory reveals what are the possible hybrid and multidisciplinary applications of Computer Science and Linguistics related degrees, far from the *conventional* employment opportunities. Looking at the future, we would like to enhance the proposed activities in order to make them more interactive even in an online context (such as the DAD) following the example of Hiippala (2021).

The laboratory is always evolving and will take place in new editions in the current school year (2022-2023) and the following one (2023-2024). Furthermore, after having received feedback at the CLiC-it conference 2021 held in Milan-Bicocca last June, we will start new collaboration also with other experts of CL and NLP in other Italian cities, thus exporting the laboratory experience out of Turin. To validate the impact of #DEACTIVHATE in the society and, in particular, in the city context we think to measure the detection of the amount of hateful message online by means of monitoring platforms, such as the "Contro l'odio" map.

Acknowledgments

The work of S. Frenda, A. T. Cignarella and M. Lai has been funded under the national project *Piano Lauree Scientifiche (PLS) 2019/20* as part of the activities of *Computer Science Department, School of Science of Nature, University of Turin*. Partially, the work of S. Frenda, A. T. Cignarella, C. Bosco and V. Patti was also supported by the International project 'STERHEOTYPES - Studying European Racial Hoaxes and sterEOTYPES' funded by the Compagnia di San Paolo and Volkswagen Stiftung under the 'Challenges for Europe' call for Project (CUP: B99C20000640007).

The authors would like to extend a special thanks to the school 'Convitto Nazionale Umberto I', and in particular, to Professor Simona Ventura for her availability and her collaboration in this adventure with #DEACTIVHATE. We also would like to thank Professors Manuela Dalbesio and Pierangelo Verga of the 'I.I.S Rivoira' for their enthusiasm and engagement in this activity and for supporting us in disseminating the laboratory in a broader area than the province of Turin.

References

Basile, Valerio. 2020. It's the End of the Gold Standard as We Know It. In *Proceedings of International Conference of the Italian Association for Artificial Intelligence*, pages 441–453, November 25–27, 2020, Virtual Event. Springer.

Bioglio, Livio, Sara Capecchi, Valentina Di Noi, Gian Marino, Ruggero G. Pensa, and Giulia Venturini. 2019. Social4School: Educare alla consapevolezza nei social network. In *Selected Papers – Conferenza GARR 2019 – Connecting the future*, pages 24–28, 4–6 June, 2019, Turin. Consortium GARR.

Bonetti, Federico and Sara Tonelli. 2020. A 3D Role-Playing Game for Abusive Language Annotation. In *Workshop on Games and Natural Language Processing*, pages 39–43, May 11, 2020, Virtual Event. European Language Resources Association.

Brown, Alexander. 2015. *Hate speech law: A philosophical examination*. Routledge.

Capozzi, Arthur Thomas Edward, Mirko Lai, Valerio Basile, Fabio Poletto, Manuela Sanguinetti, Cristina Bosco, Viviana Patti, Giancarlo Ruffo, Cataldo Musto, and Marco Polignano. 2020. "Contro L'Odio": A Platform for Detecting, Monitoring and Visualizing Hate Speech against Immigrants in Italian Social Media. *IJCoL. Italian Journal of Computational Linguistics*, 6(6-1):77–97.

Frenda, Simona, Alessandra Teresa Cignarella, Marco Antonio Stranisci, Mirko Lai, Cristina Bosco, Viviana Patti, et al. 2021. Recognizing Hate with NLP: The Teaching Experience of the #DeactivHate Lab in Italian High Schools. In *Eighth Italian Conference on Computational Linguistics (CLiC-it 2021)*, volume 3033, pages 1–7, January 26-28, 2022, Milan, Italy. CEUR-WS.org.

Fulper, Rachael, Giovanni Luca Ciampaglia, Emilio Ferrara, Y. Ahn, Alessandro Flammini, Filippo Menczer, Bryce Lewis, and Kehontas Rowe. 2014. Misogynistic language on Twitter and sexual violence. In *Proceedings of the ACM Web Science Workshop on Computational Approaches to Social Modeling (ChASM 2014)*, June 23–26, 2014, Bloomington, IN, USA.

Hiippala, Tuomo. 2021. Applied Language Technology: NLP for the Humanities. In *Proceedings of the Fifth Workshop on Teaching NLP*, pages 46–48, June 10-11, 2021, Virtual Event. Association for Computational Linguistics.

Jurgens, David, Varada Kolhatkar, Lucy Li, Margot Mieskes, and Ted Pedersen, editors. 2021. *Proceedings of the Fifth Workshop on Teaching NLP*, June 10-11, 2021, Virtual Event. Association for Computational Linguistics.

Messina, Lucio, Lucia Busso, Claudia Roberta Combei, Alessio Miaschi, Ludovica Pannitto, Gabriele Sarti, and Malvina Nissim. 2021. A Dissemination Workshop for Introducing Young Italian Students to NLP. In *Proceedings of the Fifth Workshop on Teaching NLP*, pages 52–54, June 10-11, 2021, Virtual Event. Association for Computational Linguistics.

Nadal, Kevin L., Katie E. Griffin, Yinglee Wong, Sahran Hamit, and Morgan Rasmus. 2014. The impact of racial microaggressions on mental health: Counseling implications for clients of color. *Journal of Counseling & Development*, 92(1):57–66.

Nikolaou, Dimitrios. 2017. Does Cyberbullying Impact Youth Suicidal Behaviors? *Journal of Health Economics*, 56:30–46.

Pannitto, Ludovica, Lucia Busso, Claudia Roberta Combei, Lucio Messina, Alessio Miaschi, Gabriele Sarti, and Malvina Nissim. 2021. Teaching NLP with Bracelets and Restaurant Menus: An Interactive Workshop for Italian Students. In *Proceedings of the Fifth Workshop on Teaching NLP*, June 10-11, 2021, Virtual Event. Association for Computational Linguistics.

Reskin, Barbara F. 2005. Including mechanisms in our models of ascriptive inequality. *Handbook of employment discrimination research*, pages 75–97.

Sprugnoli, Rachele, Stefano Menini, Sara Tonelli, Filippo Oncini, and Enrico Piras. 2018. Creating a WhatsApp Dataset to Study Pre-teen Cyberbullying. In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, pages 51–59, October 31, 2018. Brussels, Belgium. Association for Computational Linguistics.

Sue, Derald Wing. 2010. *Microaggressions in everyday life: Race, gender, and sexual orientation*. John Wiley & Sons.

Appendix A

List of questions of the preliminary test (PRE-TEST):

1. **[C]** Anche se non hai mai sentito questi due termini... Prova a spiegare a parole tue cosa sono l'identità sociale e l'identità personale e, secondo te, in cosa differiscono tra loro?
   (Even if you've never heard of these two terms ... Try to explain in your own words what social identity and personal identity are and, in your opinion, how do they differ from each other?)

2. **[CL]** Secondo te, la linguistica computazionale si occupa di:
   (In your opinion, computational linguistics deals with:)
   (a) studiare i linguaggi naturali e di creare algoritmi che codifichino i testi e li rendano comprensibili a dei computer
   (studying natural languages and creating algorithms that encode texts and making them understandable to computers)
   (b) leggere tweet e passare il tempo sui social network e commentare ciò che leggono
   (reading tweets and spending time on social networks and commenting on what they read)
   (c) studiare lingue inventate come l'elfico, il Dothraki, il Klingon e guardare serie TV
   (studying invented languages such as elf, Dothraki, Klingon and watching TV series)
   (d) decidere in modo arbitrario cos'è sarcastico e cosa non lo è per poi spiegarlo a tutti i loro conoscenti
   (arbitrarily decide what is sarcastic and what is not and then explaining it to all their acquaintances)

3. **[CL]** Prova ad elencare almeno 2 (o più) applicazioni della vita di tutti i giorni in cui pensi che alla base ci possa essere l'uso di tecnologie derivate dalla linguistica computazionale. (Cerca di specificare il più possibile).
   (Try to list at least 2 (or more) applications of everyday life in which you think that the basis may be the use of technologies derived from computational linguistics. (Try to specify as much as possible).)

4. **[C]** Cos'è secondo te lo Hate Speech?
   (What do you think Hate Speech is?)

5. **[C]** Leggendo il seguente testo decidi se il testo contiene discorsi di odio (HS) o non ne contiene (NON-HS).
   (By reading the following text you decide whether the text contains hate speech (HS) or does not contain any (NO-HS).)
   (a) "Naturalmente meglio omettere che il 27 era una risorsa migrante del Gambia. Tranquillo sindaco il centro di #Cagliari è ancora sicuro...come no."
   ("Of course it is better to omit that the 27 was a migrant resource of the Gambia. Keep calm mayor, the center of #Cagliari is still safe...yeah sure.")
   (b) "Domani se avete tempo, fatemi una lista dei lavori che i migranti hanno rubato a voi e ai vostri figli. Ci tengo."

("Tomorrow if you have time, make me a list of the jobs that migrants have stolen from you and your children. I care about it.")

6. **[C]** Il seguente testo contiene almeno una forma di discorsi d'odio. Scegli il fenomeno discriminatorio che ritieni più adatto tra le opzioni qui sotto e spiega il perché. [Razzismo, Misoginia, Sessismo, Ageismo, Omofobia, Abilismo]
   (The following text contains at least one form of hate speech. Choose the discriminatory phenomenon you think is most suitable from the options below and explain why. [Racism, Misogyny, Sexism, Ageism, Homophobia, Ableism])
   (a) "A me ha sempre fatto schifo, pure conciata bene, è proprio la base che mi disgusta"
   ("To me it has always sucked, even if she was well-dressed, it is precisely the core that disgusts me")
   (b) "Ma, hai sentito Manlio di Stefano cosa ha dichiarato riguardo ai SORCI DI FOGNA COSÌ DETTI ROM? DARGLI LA CASA!! Cioè noi paghiamo le tasse e una casa ce la sogniamo. Mentre i ROM truffano rubano NN PAGANO TASSE. anzi si allacciano all'enel abusivamente"
   ("But, did you hear what Manlio di Stefano said about the RATS SO-CALLED ROMAS? GIVE HIM THE HOUSE!! That is, we pay taxes and we can just dream of a house. While ROMAS cheat they steal DON'T PAY TAXES. indeed they are connected to electricity cables illegally")
   (c) "Come si fa a mettere una foto cosi volgare, vergognati non sei all'altezza di farti seguire da dei bambini, non dovresti dare un esempio così ad un pubblico di ragazzini/e che ti seguono"
   ("How can you put a photo so vulgar, shame you are not up to being followed by children, you should not give such an example to an audience of kids who follow you")
   (d) "che schifo fai negro? almeno studialo l'inno. Mi fai pena, hai sputato su una nazione piena di storia! non ci sono negri italiani!"
   ("how disgusting are you nigga? at least study the hymn. I feel sorry for you, you spit on a nation full of history! there are no Italian niggers!")
   (e) "hai 40 anni, sei un padre di famiglia e vai in giro con le unghia colorate?!"
   ("You are 40 years old, you are a family man and go around with colored nails?!")
7. **[CL]** Secondo te, considerando le domande precedenti che ti abbiamo fatto sulle forme di odio [misoginia, sessismo, omofobia, abilismo, ecc...], i tuoi compagni hanno dato le risposte uguali alle tue per ciascun testo? Spiega il perché.
   (In your opinion, considering the previous questions we asked you about forms of hatred [misogyny, sexism, homophobia, ableism, etc ...], did your classmates give the same answers as yours for each text? Explain why.)
8. **[CS]** Prova a descrivere a parole tue che cos'è un algoritmo.
   (Try to describe in your own words what an algorithm is.)
9. **[CS]** Come viene trattato il testo scritto in linguaggio naturale da una macchina/computer?
   (How does a machine/computer deal with written text in natural language?)
   (a) Trasformando dei numeri in testo
   (By turning numbers into text)

(b) Scrivendo una lista di parole tra parentesi quadre
(Writing a list of words in square brackets)
(c) Trasformando dei testi in vettori numerici
(By transforming texts into numerical vectors)
(d) Riempiendo un contenitore di zero ed uno
(Filling a container with zeros and ones)

10. Che linguaggio di programmazione conosci già?
(What programming language do you already know?)
(a) Java
(b) Javascript
(c) C
(d) Python

11. **[CS]** "che schifo fai negro? almeno studialo l'inno. Mi fai pena, hai sputato su una nazione piena di storia! non ci sono negri italiani!"
("how disgusting are you nigga? at least study the hymn. I feel sorry for you, you spit on a nation full of history! there are no Italian niggers!")
—
Il classificatore automatico ha 'etichettato' il testo qui sopra come contenente hate speech. Perché secondo te? Sulla base di quali parametri? Spiegalo qui sotto.
(The automatic classifier has 'tagged' the text above as containing hate speech. Why do you think? On the basis of what parameters? Explain it below.)

Appendix B

List of questions of the final test (POST-TEST):

1. **[C]** Nella prima lezione per conoscerci un po' meglio abbiamo visto cosa sono l'identità sociale e l'identità personale... Qual è la differenza tra identità sociale e identità personale?
   (In the first lesson, to get to know each other a little better, we saw what social identity and personal identity are... What is the difference between social identity and personal identity?)
2. **[C]** Nella prima lezione per conoscerci un po' meglio abbiamo visto cosa sono l'identità sociale e l'identità personale... Riporta alcune conclusioni tra quelle che sono emerse dal dibattito in classe.
   (In the first lesson, to get to know each other a little better, we saw what social identity and personal identity are... Please write here some of the conclusions among those that emerged from the debate in class.)
3. **[CL]** I linguisti computazionali si occupano principalmente di:
   (In your opinion, computational linguists mainly deal with:)
   (a) studiare i linguaggi naturali e di creare algoritmi che codifichino i testi e li rendano comprensibili a dei computer
   (studying natural languages and creating algorithms that encode texts and making them understandable to computers)
   (b) leggere tweet e passare il tempo sui social network e commentare ciò che leggono
   (reading tweets and spending time on social networks and commenting on what they read)
   (c) studiare lingue inventate come l'elfico, il Dothraki, il Klingon e guardare serie TV
   (studying invented languages such as elf, Dothraki, Klingon and watching TV series)
   (d) decidere in modo arbitrario cos'è sarcastico e cosa non lo è per poi spiegarlo a tutti i loro conoscenti
   (arbitrarily decide what is sarcastic and what is not and then explaining it to all their acquaintances)
4. **[CL]** All'inizio del laboratorio, abbiamo raccontato un po' di cosa si occupa la linguistica computazionale... Prova ad elencare almeno 2 (o più) applicazioni della vita di tutti i giorni in cui si vede l'uso della linguistica computazionale.
   (At the beginning of the lab, we spoke a little about what computational linguistics deals with... Try to list at least 2 (or more) applications of everyday life in which you think that the basis may be the use of technologies derived from computational linguistics. (Try to specify as much as possible).)
5. **[C]** Leggendo il seguente testo decidi se il testo contiene discorsi di odio (HS) o non ne contiene (NON-HS).
   (By reading the following text you decide whether the text contains hate speech (HS) or does not contain any (NO-HS).)
   (a) "Naturalmente meglio omettere che il 27 era una risorsa migrante del Gambia. Tranquillo sindaco il centro di #Cagliari è ancora sicuro...come no."

("Of course it is better to omit that the 27 was a migrant resource of the Gambia. Keep calm mayor, the center of #Cagliari is still safe...yeah sure.")

(b) "Domani se avete tempo, fatemi una lista dei lavori che i migranti hanno rubato a voi e ai vostri figli. Ci tengo."
("Tomorrow if you have time, make me a list of the jobs that migrants have stolen from you and your children. I care about it.")

6. **[C]** Il seguente testo contiene almeno una forma di discorsi d'odio. Scegli il fenomeno discriminatorio che ritieni più adatto tra le opzioni qui sotto e spiega il perché. [Razzismo, Misoginia, Sessismo, Ageismo, Omofobia, Abilismo]
(The following text contains at least one form of hate speech. Choose the discriminatory phenomenon you think is most suitable from the options below and explain why. [Racism, Misogyny, Sexism, Ageism, Homophobia, Ableism])

   (a) "A me ha sempre fatto schifo, pure conciata bene, è proprio la base che mi disgusta"
   ("To me it has always sucked, even if she was well-dressed, it is precisely the core that disgusts me")

   (b) "Ma, hai sentito Manlio di Stefano cosa ha dichiarato riguardo ai SORCI DI FOGNA COSÌ DETTI ROM? DARGLI LA CASA!! Cioè noi paghiamo le tasse e una casa ce la sogniamo. Mentre i ROM truffano rubano NN PAGANO TASSE. anzi si allacciano all'enel abusivamente"
   ("But, did you hear what Manlio di Stefano said about the RATS SO-CALLED ROMAS? GIVE HIM THE HOUSE!! That is, we pay taxes and we can just dream of a house. While ROMAS cheat they steal DON'T PAY TAXES. indeed they are connected to electricity cables illegally")

   (c) "Come si fa a mettere una foto cosi volgare, vergognati non sei all'altezza di farti seguire da dei bambini, non dovresti dare un esempio così ad un pubblico di ragazzini/e che ti seguono"
   ("How can you put a photo so vulgar, shame you are not up to being followed by children, you should not give such an example to an audience of kids who follow you")

   (d) "che schifo fai negro? almeno studialo l'inno. Mi fai pena, hai sputato su una nazione piena di storia! non ci sono negri italiani!"
   ("how disgusting are you nigga? at least study the hymn. I feel sorry for you, you spit on a nation full of history! there are no Italian niggers!")

   (e) "hai 40 anni, sei un padre di famiglia e vai in giro con le unghia colorate?!"
   ("You are 40 years old, you are a family man and go around with colored nails?!")

7. **[CL]** Se tutte le persone della tua classe annotano lo stesso tweet pensando alla classificazione binaria con le etichette 0 e 1 (ad esempio: questo tweet contiene un messaggio di odio (1), oppure non lo contiene (0)), è ragionevole pensare che ci sia una parte della classe in accordo e una parte della classe in disaccordo. Come fa un linguista computazionale per risolvere questa problematica?
(If everyone in your class annotates the same tweet thinking about the binary classification with labels 0 and 1 (for example: this tweet contains a hate message (1), or does not contain it (0)), it is reasonable to think that there is a part of the class in agreement and a part of the class in disagreement. How does a computational linguist go about solving this problem?)

8. **[CS]** Prova a descrivere a parole tue che cos'è un algoritmo.
   (Try to describe in your own words what an algorithm is.)
9. **[CS]** La vettorizzazione del testo è...
   (Text vectorization is ...)
   (a) Trasformando dei numeri in testo
   (By turning numbers into text)
   (b) Scrivendo una lista di parole tra parentesi quadre
   (Writing a list of words in square brackets)
   (c) Trasformando dei testi in vettori numerici
   (By transforming texts into numerical vectors)
   (d) Riempiendo un contenitore di zero ed uno
   (Filling a container with zeros and ones)
10. Che linguaggio di programmazione abbiamo usato in classe durante il laboratorio?
    (What programming language did we use in the classroom during the laboratory?)
    (a) Java
    (b) Javascript
    (c) C
    (d) Python
11. **[CS]** Un esempio pratico di algoritmo nella vita di tutti i giorni è...
    (A practical example of an algorithm in everyday life is ...)
12. **[CS]** Che output ti aspetteresti se dovessi eseguire il seguente codice:
    (What output would you expect if you were to run the following code:)

    ```
    print("Ciao a tutti!"*3)
    ```

13. **[CS]** Che output ti aspetteresti se dovessi eseguire il seguente codice:
    (What output would you expect if you were to run the following code:)

    ```
    for n in range(5):
        print(n)
    ```

14. **[CS]** "che schifo fai negro? almeno studialo l'inno. Mi fai pena, hai sputato su una nazione piena di storia! non ci sono negri italiani!"
    ("how disgusting are you nigga? at least study the hymn. I feel sorry for you, you spit on a nation full of history! there are no Italian niggers!")
    —
    Il classificatore automatico ha 'etichettato' il testo qui sopra come contenente hate speech. Perché secondo te? Sulla base di quali parametri? Spiegalo qui sotto.
    (The automatic classifier has 'tagged' the text above as containing hate speech. Why do you think? On the basis of what parameters? Explain it below.)
15. **[CS]** Immagina di dover costruire un classificatore automatico di testi che sia capace di distinguere i testi che contengono discorsi di odio da quelli che non ne contengono. Che strategia adotteresti?
    (Imagine having to build an automatic text classifier that is capable of distinguishing texts that contain hate speech from those that do not. What strategy would you adopt?)