Implicit Stereotypes: A Corpus-Based Study for Italian

Wolfgang S. Schmeisser-Nieto[1,2,*], Giacomo Ricci[2], Simona Frenda[3,4], Mariona Taulé[1] and Cristina Bosco[2]

[1]Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, Barcelona, Spain
[2]University of Turin, Dipartimento di Informatica, Corso Svizzera 185, 10149 Torino, Italy
[3]Interaction Lab, Heriot-Watt University, The Avenue, Edinburgh, EH14 4AS, Scotland
[4]aequa-tech, Torino, Italy

Abstract

Detecting stereotypes is a challenging task, particularly when they are not expressed explicitly. In this study, we applied an annotation schema from the literature designed to formalize implicit stereotypes. We analyzed implicit stereotypes about immigrants in two datasets: StereoHoax-IT and SterheoSchool, which are created from different sources. StereoHoax-IT consists of reactions on Twitter to specific hoaxes aimed at discriminating against immigrants, while SterheoSchool includes comments from teenagers on fake news generated in psychological experiments. We describe the annotation process, annotator disagreements, and provide both quantitative and qualitative analyses to shed light on how implicitness characterizes stereotypes in different texts. Our findings suggest that implicit stereotypes are often conveyed through logical linguistic relations, such as entailment and behavioral evaluations of immigrants.

Keywords

Implicit stereotype, Corpora annotation, Corpora analysis, Italian language

1. Introduction and Background

Various recent NLP studies have focused on detecting stereotypes online, often in conjunction with forms of abusive language [1, 2, 3, 4, 5]. The importance of tackling this phenomenon is due to its impact on social structures and the power of individuals. Therefore, detecting stereotypes can prevent their emergence and spread, and thereby have a positive impact on our society.

In social psychology, a stereotype has been defined as a set of beliefs about others perceived as belonging to a different social group [6]. It oversimplifies the features of the group and generalizes a particular feature, applying it to all its members [6]. In contrast to the emotional component of prejudice and the behavioral component of discrimination, a stereotype is associated with the cognitive component of the triad [7]. In language, stereotypes can be expressed explicitly or implicitly [8]. Explicit stereotypes deliver a straightforward message, clearly revealing the associated traits, often using derogatory adjectives [9, 10]. In contrast, implicit stereotypes are more nuanced and indirect, requiring the reader to infer their meaning [11]. These implicit stereotypes can be communicated through linguistic devices such as metaphor and irony [9], negation [12], or entailments [13]. Recently, efforts have been made to formalize the strategies for expressing implicit stereotypes, with the goal of establishing standardized criteria for annotators [14]. An example of explicit stereotype is *"[Gli immigrati] buttano via il cibo che gli danno per poi andare a mangiare i poveri cani, dove finiremo!"* [1] (extracted from StereoHoax-IT corpus), in which the generalization of the target group and the association with an action is expressed in a present tense with a habitual aspect. On the other hand, in the example *"Come noi rispettiamo loro e il colore della loro pelle, così loro che abitano nei nostri paesi dovrebbero portare rispetto nei nostri confronti."* [2] (SterheoSchool corpus), the stereotype is not overtly manifested, but it must be inferred through the evaluation of the in-group and an exhortative sentence.

From a computational linguistics perspective, concerns have been raised about how to detect and process stereotypes, a task often considered closely related to the detection of abusive language or hate speech [15].
Alongside research on hate speech, the study of stereotype detection has increased, particularly within evaluation tasks [16, 4, 17, 18, 19]. However, the detection of implicit stereotypes remains a significant challenge [20]. There are several works that deal with stereotypes in more complex narratives, such as microportraits [21] and political debates [22]. The detection of implicitness has also been studied with reference to several other

CLiC-it 2024 - Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy
*Corresponding author.
wolfgang.schmeisser@ub.edu (W. S. Schmeisser-Nieto); giacomo.ricci@edu.unito.it (G. Ricci); s.frenda@hw.ac.uk (S. Frenda); mtaule@ub.edu (M. Taulé); cristina.bosco@unito.it (C. Bosco)
0000-0001-5663-6276 (W. S. Schmeisser-Nieto); 0000-0002-6215-3374 (S. Frenda); 0000-0003-0089-940X (M. Taulé); 0000-0002-8857-4484 (C. Bosco)

[1]Transl. *"They throw away the food they are given only to go eat the poor dogs. Where will we end up!"*
[2]Transl. *"Just as we respect them and the color of their skin, they, who live in our countries, should show respect toward us."*

phenomena, in particular those characterized by subjectivity, such as irony [23]. In this paper, we analyze the implicit manifestation of stereotypes targeting immigrants, using a well-defined annotation schema proposed by Schmeisser-Nieto et al. [14] and tested on a subset of comments from Spanish newspapers (DETESTS [5]). This schema represents different criteria for determining the implicitness of stereotypes in an attempt to formalize the concept. Disentangling strategies of implicitness presents a significant challenge, often resulting in the identification of multiple categories within the same text.

Our main contributions consist of expanding the annotation with topics of stereotypes about immigrants [5] and the strategies to implicitness [14], as well as testing this schema on two existing Italian datasets. These datasets share the same domain as those used for Spanish, stereotypes about immigrants, and include data extracted from Twitter (now X) as reactions to specific hoaxes (StereoHoax-IT) and comments written by high school students to two examples of fake news artificially created within psychological experiments (SterheoSchool) as described in [24, 25]. Analyzing the annotated texts, we noted that implicit stereotypes appear to be conveyed especially through logical linguistic relations like entailment and the behavioral evaluation of immigrants in both datasets. Moreover, in most cases, the annotators needed to use contextual information to determine the presence of stereotypes. For example, in this case *"Che centra lui e Italiano!, può essere massacrato!"* [3] (StereoHoax-IT) the author of the message expresses a stereotype complaining that foreigners enjoy better treatment than Italians, who can indeed be "macellati" (slaughtered).

The rest of the paper is organized as follows: Sections 2 and 3 describe the datasets and the annotation applied; Sections 4 and 5 present quantitative and qualitative analyses of the annotated data; and Section 6 summarizes the results and provides guidance regarding future work.

2. Datasets

In this work, we focus on two annotated corpora containing implicit stereotypes developed within the STERHEOTYPES project[4] and the SterotypHate project[5]. Their content is related to attitudes regarding immigrants and they share similar conversational structures and the same annotation scheme. Each message in these datasets is contextualized, i.e. collocated within a discourse thread or presented as a comment on a given news item. For the annotation scheme, each message is annotated for the presence or absence of anti-migrant stereotypes, and, if present, for other related categories such as whether the stereotype was expressed implicitly or explicitly and which forms of discredit the stereotype could be classified at. This category is inspired by the Stereotype Content Model (SCM) [7] and allowed us to observe the stereotype from a perspective that encompasses psychology and computational linguistics [26]. In section 3, we show how we extended this annotation to describe the dimension of implicitness[6]. **StereoHoax-IT** [27] is a contextualized multilingual dataset of tweets annotated primarily for the presence of anti-migrant stereotypes. The dataset consists of replies to tweets identified as containing racial hoaxes specifically targeting migrants and collected from debunking websites from French, Italian and Spanish Twitter, collected from 2019 to 2021. Each message is provided with its "conversation head" (the message containing the source racial hoax), and its direct parent message (if applicable). In this paper, we only use the Italian subset, which includes 3,123 instances. Due to the rarity of the phenomenon, there is a significant class imbalance: 472 instances (15%) contain a stereotype, 332 of which (70%) are implicit and 140 (30%) are explicit.

SterheoSchool [28] consists of a selection of data collected in Italian schools during experiments conducted by social psychologists [24, 25]. More precisely, it includes the reactions of teenagers, who read two hoaxes artificially created and presented as news articles, recorded via a cell phone interface. The hoaxes were designed to elicit reactions to stereotypes in readers. For each news item, readers were asked to comment on the news and on the main character of the articles. These comments are also associated with metadata, such as the age and declared gender of the author. By collecting data generated by teenagers, this corpus aims to fill a gap in the literature in which teenagers are an underrepresented category in data annotated for text classification tasks. We applied the annotation scheme mentioned above to the news and comments. This corpus consists of 1,147 comments, of which 337 (33.8%) are annotated as containing stereotypes, of which 152 (45%) are expressed in an implicit form.

3. Annotation

The annotation scheme we applied on the two corpora is based on two different layers, *topics of stereotypes* and *implicitness strategies*, as well as the need for *context*.

The **topics** of stereotypes were firstly introduced within an evaluation task, DETESTS [5], in which the participants had to train models to decide whether a text

[3] Transl. *"That's not the point, he is Italian! He can be slaughtered!"*

[4] STERHEOTYPES (Studying European Racial Hoaxes and sterEOTYPES) is an international project funded by Compagnia di San Paolo and VolksWagen Stiftung.

[5] StereotypHate is a project funded by Compagnia di San Paolo.

[6] The datasets will be made available for research purposes after the acceptance of the paper in anonymized form.

contained stereotypes, and when they did, classify the stereotype into ten different categories:

- **Xenophobia victims** Immigrants are perceived as victims of xenophobia and discrimination. They enrich culture and diversity and should have the same rights as citizens.
- **Suffering victims** Immigrants are portrayed as victims of poverty and violence in their places of origin and as having to face difficult situations in their host countries.
- **Economic resources** Immigrants are seen as an economic resource. They do the jobs that locals do not want to do, pay taxes and solve the problems arising from low population growth.
- **Migration control** Immigrants present a threat due to massive influxes and a lack of control at the borders. Immigrants are illegal and should be expelled. It is seen as an invasion.
- **Culture and religion differences** Immigrants suppose a loss of the in-group's values and traditions and the replacement of the target group's customs and religions. They are also seen as uneducated and should adapt to their host country.
- **Benefits** Immigrants compete with the in-group for resources such as public subsidies, school places, jobs, health care and pensions. They are privileged over the in-group.
- **Public health** Immigrants are thought to be carriers of infections and diseases such as COVID-19, Ebola and HIV.
- **Security** Immigration brings security issues. Due to immigration, there is an increase in crime, domestic violence, robbery, drug use, sexual assault, murder, terrorist attacks and public disorders.
- **Dehumanization** Immigrants are seen as inferior beings and are compared with animals, parasites or scum. Their lives have less value than those of the in-group.
- **Other topics** Any other immigration stereotypes not covered in the previous categories.

Context and implicitness strategies were initially proposed as criteria that could help annotators to annotate implicitness, since their vagueness may decrease Inter-Annotator Agreement (IAA) [14]. By **context**, we refer to information contained in previous messages, which is considered necessary to understand the meaning of the message to be annotated, as in the following example: *"Sempre assolti...sempre misure e pesi differenti". Context: "Uccide anziana ebrea al grido di Allah Akbar. Assolto perché drogato."*[7] (StereoHoax-IT). Regarding the **strategies** and linguistic devices used to convey implicit stereotypes, we have revised the criteria proposed in [14] as follows:

- **World knowledge** World knowledge refers to the shared cultural, social and historical knowledge needed to interpret messages, e.g., *"La scuola si inchina all'islam: l'aceto è bandito dalle mense."* [8] (StereoHoax-IT)
- **Figures of speech** Every figure of speech except for irony and sarcasm, and humor and jokes. For instance, metaphor, rhetorical questions, euphemisms or reported speech, e.g., *"Chi è quel pazzo che si mette in casa uno di questi? Un suicidio"* [9] (StereoHoax-IT)
- **Irony/Sarcasm** The message expresses a meaning that is the opposite of what is said, e.g. in *"Che bella gente fanno arrivare.....che bello avere un paese pieno di risorse pronte a tutto.....ma proprio a tutto."* [10] (StereoHoax-IT)
- **Humor/Jokes** Jokes about a target group often use stereotypes and may or may not include irony, e.g. in *"Chissà se ha detto:"Cibo no buono"."* [11] (StereoHoax-IT)
- **Extrapolation** The target refers to an individual or specific members of a social group, not the group as a whole, e.g. in *"Classico del sud-italia Maleducata"* [12] (SterheoSchool)
- **Imperative/Exhortative** Calls to take certain actions related to the target group, e.g. *"Come in Cina FUCILATELO"* [13] (StereoHoax-IT)
- **Entailment/Evaluation** Logical relation between two sentences in which the condition of truth of sentence A implies the truth of sentence B. The implicit stereotype is implied in sentence A. An evaluation of the author's or in-group's thoughts, emotions and behaviors, rather than content about the out-group or target group, can be considered as a type of entailment, e.g. *"Saranno fuori o liberi presto"* [14](StereoHoax-IT) is the answer to a racial hoax in which a group of immigrants rape and murder a teenage girl. With the author's evaluation of the situation, it is entailed that immigrants are immune from punishment.
- **Other implicitness** Other types of implicitness not considered in the previous categories.
  e.g. *"al giorno d'oggi non ci si può fidare di nessuno una persona ripugnante"* [15](SterheoSchool)

[7]Transl. *"Always acquitted...always different measures and weights." Context: "Kills elderly Jewish woman while shouting 'Allah Akbar.' Acquitted because he was on drugs."*

[8]Transl. *"The school bows to Islam: vinegar is banned from canteens."*

[9]Transl. *"Who's that fool who takes one of these into his house? a suicide"*

[10]Transl. *"Such nice people they bring in... how nice it is to have a country full of resources ready for anything... anything at all"*

[11]Transl. *"I wonder if he said: «Food no good»"*

[12]Transl. *"Typical of Southern Italy"*

[13]Transl. *"SHOOT HIM like in China"*

[14]Transl. *"They will be out or free soon"*

[15]Transl. *"nowadays you can't trust anyone a repulsive person"*

Table 1
Inter-annotator agreement test using Fleiss' kappa (κ) coefficient on the categories of implicitness and stereotype topics of the StereoHoax-IT and the SterheoSchool corpora.

Label	StereoHoax-IT	SterheoSchool
Xenophobia victims	0.57	0.50
Suffering victims	0.49	0.50
Economic resource	0.48	0.50
Migration control	0.77	0.55
Culture & religion	0.75	0.71
Benefits	0.75	0.62
Public health	0.86	0.50
Security	0.81	0.64
Dehumanization	0.71	0.71
Other topics	0.52	0.43
Context	0.72	0.50
World knowledge	0.52	0.51
Figures of speech	0.68	0.70
Irony/Sarcasm	0.70	0.50
Humor/Jokes	0.52	No cases
Extrapolation	0.51	0.53
Imperative/Exhortative	0.73	0.53
Entailment/Evaluation	0.45	0.49
Other implicitness	0.51	0.52

The annotation was carried out on the Label Studio platform by three native Italian speakers with a background in linguistics, some of whom specialized in NLP. They achieved an acceptable to good IAA in the majority of cases, as reported in Table 1, which varies across categories and corpora. By observing Table 2, we can see that only a few topics have been marked by the majority of annotators , while not all the implicit criteria have been identified in the texts (i.e., 'humor/jokes').

4. Quantitative Analysis

Table 2 shows the distribution of the disaggregated annotations across both datasets. Columns *0%*, *33%*, *67%* and *100%*, respectively, indicate the number of instances per label that were annotated by no annotator (0%), by one annotator (33%), by two annotators (67%) and by all three annotators (100%). Column *% positive class* shows the percentage of the label voted by the majority of annotators, and its total number of cases in parentheses.

Firstly, an inconsistency in the distribution of labels can be observed since SterheoSchool has a representation of labels of more than 10% on only four labels. This disparity is due to the extraction methods of each dataset: the topics of the racial hoaxes used to extract the dataset were more balanced in StereoHoax-IT than in SterheoSchool, with the latter focusing generally on security and cultural differences that are discussed in the two only contexts provided to the students for their comments. However, while in the former there is a representation of all the stereotypical topics that portray immigrants as threats, the security issue is highly prevalent in both datasets.

A common trend shows that the most frequent implicitness strategy in both datasets is 'entailment/evaluation', accounting for 64% in StereoHoax-IT and 80% in SterheoSchool. To a lesser degree, 'extrapolation' appears in both datasets, with 13% in the former and 19% in the latter, respectively. Other represented strategies that exceed 10% of instances are only found in StereoHoax-IT.

The label 'context' has a high prevalence in both datasets, accounting for 38% in StereoHoax-IT and 80% in SterheoSchool. This is expected, as it depends on the methodology to produce the comments—spontaneous versus controlled—and the variety of contexts: two fake news for StereoSchool and 50 racial hoaxes for StereoHoax-IT. The limited amount of data unfortunately does not allow us to reliably evaluate a correlation between 'context' and certain implicitness strategies, as shown in Table 3, except for the association between 'entailment/evaluation' and 'context' across both datasets. The correlation between 'implicitness' and 'context' is also shown in Bourgeade et al. [27], with significant associations of the aforementioned labels in three languages: French, Italian and Spanish. In StereoHoax-IT, the correlations between the 'context' and 'irony/sarcasm', 'extrapolation' and 'imperative/exhortative' are also significant, whereas the category of other implicitness strategies is also significantly correlated in SterheoSchool, which can be analyzed qualitatively to determine if there is a pattern among them. The other strategies do not have representative instances that allow for analyzing them comparatively, except for 'extrapolation', which is significantly correlated in StereoHoax-IT but not in SterheoSchool.

In terms of co-occurrences between topics and implicit strategies, we can observe from Table 4 that there is also a great disparity in both datasets. Focusing on the two topics with the highest representation in SterheoSchool (Culture & religion, 51%, and security, 35%), which account for the majority of the corpus, we can analyze some differences with StereoHoax-IT. Firstly, 'culture & religion' is expressed primarily through entailments or evaluations (65 co-occurrences) and secondarily through extrapolations in SterheoSchool. In contrast, the distribution of strategies used to represent 'culture & religion' stereotypes is more evenly spread in StereoHoax-IT. A similar pattern is observed with the topic of 'security', which, while concentrating strategies in 'entailment/evaluation,' also utilizes a range of other strategies, particularly 'extrapolation' and 'imperative/exhortative'. With these co-occurrences, we can reaffirm that the different methods to extract the data have an impact on the characteristics of it, and therefore, its distribution of labels. For instance, the messages were written in a non-controlled environment, which gives the authors the freedom to express themselves without constrains. Moreover, the

Table 2
Distribution of labels and percentages of positive class.

	StereoHoax-IT					SterheoSchool				
Labels	**0%**	**33%**	**67%**	**100%**	**% positive class**	**0%**	**33%**	**67%**	**100%**	**% positive class**
Xenophobia victims	265	54	12	1	4% (13)	149	3	0	0	%0 (0)
Suffering victims	313	19	0	0	0% (0)	148	4	0	0	0% (0)
Economic resource	299	33	0	0	0% (0)	151	1	0	0	0% (0)
Migration control	203	48	45	36	24% (81)	140	8	2	2	3% (4)
Culture & religion	254	43	15	20	11% (35)	37	38	49	28	51% (77)
Benefits	235	30	41	26	20% (67)	139	11	2	0	1% (2)
Public health	257	16	23	36	18% (59)	151	1	0	0	0% (0)
Security	128	42	48	114	49% (162)	48	50	29	25	36% (54)
Dehumanization	258	40	21	13	10% (34)	126	17	4	5	6% (9)
Other topics	316	15	1	0	0% (1)	66	76	10	0	7% (10)
Context	116	90	45	81	38% (126)	1	28	61	62	81% (123)
World knowledge	187	111	31	3	10% (34)	136	15	1	0	1% (1)
Figures of speech	257	40	27	8	11% (35)	142	8	0	2	1% (2)
Irony/Sarcasm	247	42	30	13	13% (43)	151	1	0	0	0% (0)
Humor/Jokes	300	29	3	0	1% (3)	152	0	0	0	0% (0)
Extrapolation	157	133	36	6	13% (42)	69	54	26	3	19% (29)
Entailment/Evaluation	20	100	167	46	64% (212)	1	30	63	58	80% (121)
Imperative/Exhortative	238	49	24	21	14% (45)	106	38	7	1	5% (8)
Other implicitness	301	29	2	0	1% (2)	100	41	11	0	7% (11)

Table 3
Association between contextuality and implicitness. The values where p is significant are shown in bold.

	StereoHoax-IT		SterheoSchool	
	Cramer's V	X^2 / p-value	Cramer's V	X^2 / p-value
World knowledge	0.074	1.8 / 0.18	0.064	0.623 / 0.43
Figures of speech	0.105	3.691 / 0.055	0.0	0.0 / 1.0
Irony/Sarcasm	0.188	11.759 / **0.001**	–	0.0 / 1.0
Humor/Jokes	0.089	2.648 / 0.104	–	0.0 / 1.0
Extrapolation	0.176	10.315 /**0.001**	0.041	0.258 / 0.611
Entailment/Evaluation	0.232	17.872 / **0.0**	0.232	8.189 / **0.004**
Imperative/Exhortative	0.116	4.502 / **0.034**	0.077	0.9 / 0.343
Other implicitness	0.059	1.173 / 0.279	0.22	7.344 / **0.007**

topics in StereoHoax-IT are more balanced, as seen in the distribution of 'entailment/evaluation', which is also used in 'migration control', 'benefits', 'public health' and 'dehumanization'. On the other hand, in SterheoSchool, both initial fake news have the same narrative features, such as describing an aggression and highlighting the origin of the aggressor, thus eliciting a reaction in the readers related to these topics. The example *"Siamo alla follia: ad Agrigento autobus gratis agli immigrati per evitare violenze e aggressioni."* [16] (StereoHoax-IT) is related to security expressed through extrapolation. The example *"Un cristiano che entrasse in una moschea in un paese arabo e sputasse per terra sopravviverebbe pochi secondi."* [17] (StereoHoax-IT) highlights cultural and religious differences by the evaluation of a hypothetical situation.

[16]Transl. *"It's crazy: in Agrigento, free buses for immigrants to prevent violence and aggressions."*

[17]Transl. *"A Christian entering a Mosque in an Arab country and spitting on the ground would survive a few seconds."*

5. Qualitative analysis

To deepen the analysis of implicitness strategies and their interaction with different topics, we explore some messages to uncover the linguistic structures that are characteristic of implicit communication.

Example 1 has been annotated with the topic 'public health' and 'figures of speech' and 'Irony/Sarcasm' for the strategy of implicitness; all labels achieved a 67% IAA.

1) *Governo di involtini primavera!!!* [18] (StereoHoax-IT)

In the context given for this message, the author complains that the government did not use more restrictive measures against Chinese children during the early stages of COVID-19. First, an ironic reading, i.e., as stating A to mean not-A, is triggered by the metonymy "spring rolls" [29], identifying Chinese citizens through a traditional Chinese dish. Second, disapproval is conveyed showing a kind of favorable attitude of the Italian

[18]Trasl. *"Spring rolls government."*

Table 4
Co-occurrence of implicitness strategies and topics of stereotypes. The numbers on the left correspond to StereoHoax-IT, whereas the numbers on the right correspond to SterheoSchool.

	StereoHoax-IT / SterheoSchool							
	World knowledge	Figures of speech	Irony/ Sarcasm	Humor/ Jokes	Extrapolation	Imperative/ Exhortative	Entailment/ Evaluation	Other implicitness
Xenophobia victims	4 / 0	3 / 0	2 / 0	1 / 0	0 / 0	2 / 0	5 / 0	0 / 0
Suffering victims	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
Economic resource	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
Migration control	7 / 0	**13** / 0	**10** / 0	0 / 0	4 / 1	**13** / 0	**55** / 4	1 / 0
Culture & religion	**11** / 0	0 / 1	6 / 0	2 / 0	5 / **17**	3 / 7	**22** / **65**	0 / 1
Benefits	**12** / 0	8 / 0	**11** / 0	0 / 0	1 / 1	7 / 0	**51** / 2	0 / 0
Public health	2 / 0	**17** / 0	8 / 0	1 / 0	3 / 0	4 / 0	**43** / 0	0 / 0
Security	**7** / 0	**12** / 1	**17** / 0	0 / 0	**35** / 6	**29** / 2	**103** / **45**	0 / 4
Dehumanization	3 / 0	5 / 0	3 / 0	2 / 0	7 / 1	**13** / 1	**14** / 8	1 / 0
Other topics	0 / 0	0 / 0	0 / 0	0 / 0	0 / 4	0 / 0	0 / 5	1 / 4

government toward Chinese children.

Example 2 was annotated as 'culture & religion' by all three annotators. In terms of the implicitness strategies, it was labeled as both 'extrapolation' and 'entailment/evaluation' by two out of the three annotators.

2) *Venezia, donne velate sputano al crocifisso.* [19] (StereoHoax-IT)

In this case, the noun phrase "veiled women" is a case of lexical narrowing, i.e., a lexical item conveys a meaning that is more specific than the item's encoded meaning. The reader selects a more specific meaning on the basis of stereotypes and world knowledge [30] of the meaning of "veiled women", which denotes a set of women who wear a veil, narrowed to mean Muslim women. This equalization arises from the stereotype that posits that if a woman wears a veil, she is a Muslim. Furthermore, the absence of the determiner in the noun phrase, that usually indicates a generic reference, combined with the imperfective aspect and present tense of the verb, may suggest a habitual interpretation of the predicate "spit on the crucifix" [31]. 'Extrapolation' strategy here refers to the attribution of this action to the entire category.

Among the more frequently agreed implicitness strategies, there are 'imperative/exhortative' and 'figures of speech', which have linguistic and punctuation features closer to explicitness: the former is associated with a specific grammatical mood and the exclamation mark, while the latter is associated with a question mark (considering that rhetorical questions are frequently annotated as a figure of speech), see e.g.:

3) *Se non fate niente Fra 10 anni l'italia sarà tutta musulmana!*[20] (StereoHoax-IT)

4) *Come ci si può sentir sicuri in una società che permette questo? meschina*[21] (SterheoSchool)

The high IAA for the category of 'irony/sarcasm' is also interesting, and has been studied especially in social media [32, 33], as a means to lower the negative social cost of what has been said. The two categories that most frequently co-occur with 'irony/sarcasm' in StereoHoax-IT are 'figures of speech' (out of 35 instances, six are also ironic) and 'humor/jokes' (out of three cases, two are ironic), as in the next example:

5) *@Belle facce intelligenti! Viva Lombroso!*[22] (67% Humor/Jokes, 67% Irony/Sarcasm, StereoHoax-IT)

We found messages in which 'entailment/evaluation' co-occurs with 'irony/sarcasm', but this correlation should be analyzed in depth to be considered relevant, as 64% of instances were annotated as 'entailment/evaluation.'

6. Conclusions

In this paper, we applied an annotation scheme for analyzing the implicitness of stereotypes against immigrants according to two main dimensions (i.e., topics and strategies for making the content implicit) to the Italian StereoHoax-IT and SterheoSchool corpora. Adding these two layers of annotation allowed us to observe that annotators need to use contextual information to determine the presence of stereotypes especially, when specific strategies have been used by the author of the message (irony/sarcasm, extrapolation, entailment/evaluation, and imperative/exhortative). Moreover, implicit stereotypes appear to be conveyed mainly through logical linguistic relations such as the entailment and behavioral evaluation of immigrants and, in fewer cases, via 'imperative/exhortative', 'irony/sarcasm' and 'extrapolation.'

As future work, we plan to perform a comparative analysis with the datasets in Spanish, which have already been annotated with this schema, in order to understand cultural analogies and differences in portraying immigrants as threats, enemies or victims.

[19] Trasl. *"Venice, veiled women spit on the crucifix."*
[20] Trasl. *"If you do nothing In 10 years Italy will be completely Muslim"*
[21] Trasl. *"How can one feel secure in a society that allows this? mean"*
[22] Trasl. *"Nice smart faces! Long life Lombroso!"*

Acknowledgments

The work of Wolfgang Schmeisser-Nieto is funded by the project StereotypHate (Compagnia di San Paolo for the call 'Progetti di Ateneo - Compagnia di San Paolo 2019/2021 - Mission 1.1 - Finanziamento ex-post').
The work of Cristina Bosco is partially funded by the same project.

References

[1] M. Anzovino, E. Fersini, P. Rosso, Automatic identification and classification of misogynistic language on Twitter, in: M. Silberztein, F. Atigui, E. Kornyshova, E. Métais, F. Meziane (Eds.), Natural Language Processing and Information Systems - 23rd International Conference on Applications of Natural Language to Information Systems, NLDB 2018, Paris, France, June 13-15, 2018, Proceedings, volume 10859 of *Lecture Notes in Computer Science*, Springer, 2018, pp. 57–64. URL: https://doi.org/10.1007/978-3-319-91947-8_6.

[2] E. Lavergne, R. Saini, G. Kovács, K. Murphy, TheNorth @ HaSpeeDe 2: BERT-based language model fine-tuning for Italian hate speech detection, in: Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian Final Workshop, EVALITA - December 17th, 2020, volume 2765, CEUR-WS, 2020, pp. 142–147. URL: http://ceur-ws.org/Vol-2765/paper135.pdf.

[3] M. Sanguinetti, G. Comandini, E. D. Nuovo, S. Frenda, M. Stranisci, C. Bosco, T. Caselli, V. Patti, I. Russo, Haspeede 2 @ EVALITA2020: Overview of the EVALITA 2020 hate speech detection task, in: V. Basile, D. Croce, M. D. Maro, L. C. Passaro (Eds.), Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online event, December 17th, 2020, volume 2765 of *CEUR Workshop Proceedings*, CEUR-WS, 2020. URL: http://ceur-ws.org/Vol-2765/paper162.pdf.

[4] M. Taulé, A. Ariza, M. Nofre, E. Amigó, P. Rosso, Overview of DETOXIS at IberLEF 2021: DEtection of TOXicity in comments In Spanish, Procesamiento del Lenguaje Natural 67 (2021) 209–221. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6390.

[5] A. Ariza-Casabona, W. S. Schmeisser-Nieto, M. Nofre, M. Taulé, E. Amigó, B. Chulvi, P. Rosso, Overview of DETESTS at IberLEF 2022: DETEction and classification of racial STereotypes in Spanish, Procesamiento del Lenguaje Natural 69 (2022) 217–228.

[6] G. W. Allport, K. Clark, T. Pettigrew, The nature of prejudice, Addison-wesley Reading, MA, 1954.

[7] S. T. Fiske, Stereotyping, prejudice, and discrimination, in: The Handbook of Social Psychology, Vols. 1-2, 4th Ed, McGraw-Hill, New York, NY, US, 1998, pp. 357–411.

[8] A. G. Greenwald, M. R. Banaji, Implicit social cognition: Attitudes, self-esteem, and stereotypes, Psychological review 102 (1995) 4—27. URL: http://faculty.washington.edu/agg/pdf/Greenwald_Banaji_PsychRev_1995.OCR.pdf. doi:10.1037/0033-295x.102.1.4.

[9] K. A. Collins, R. Clément, Language and prejudice: direct and moderated effects, Journal of Language and Social Psychology 31 (2012) 376–396. URL: http://journals.sagepub.com/doi/10.1177/0261927X12446611. doi:10.1177/0261927X12446611.

[10] F. D'Errico, M. Paciello, Online moral disengagement and hostile emotions in discussions on hosting immigrants, Internet Research 28 (2018) 1313–1335. URL: https://www.emerald.com/insight/content/doi/10.1108/IntR-03-2017-0119/full/html. doi:10.1108/IntR-03-2017-0119.

[11] U. Quasthoff, The uses of stereotype in everyday argument, Journal of pragmatics 2 (1978) 1–48.

[12] C. J. Beukeboom, C. Finkenauer, D. H. J. Wigboldus, The negation bias: When negations signal stereotypic expectancies., Journal of Personality and Social Psychology 99 (2010) 978–992. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0020861. doi:10.1037/a0020861.

[13] T. F. Pettigrew, R. W. Meertens, Subtle and blatant prejudice in western Europe, European Journal of Social Psychology 25 (1995) 57–75. URL: https://onlinelibrary.wiley.com/doi/10.1002/ejsp.2420250106. doi:10.1002/ejsp.2420250106.

[14] W. S. Schmeisser-Nieto, M. Nofre, M. Taulé, Criteria for the annotation of implicit stereotypes, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022), 2022, pp. 753–762.

[15] C. Bosco, F. Dell'Orletta, F. Poletto, M. Sanguinetti, M. Tesconi, Overview of the evalita 2018 hate speech detection task, in: EVALITA 2018-Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, volume 2263, CEUR, 2018, pp. 1–9.

[16] M. Sanguinetti, G. Comandini, E. di Nuovo, S. Frenda, M. Stranisci, C. Bosco, T. Caselli, V. Patti, I. Russo, Haspeede 2 @ EVALITA2020: Overview of the EVALITA 2020 hate speech detection task, in: V. Basile, D. Croce, M. Di Maro, L. Passaro (Eds.), Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools

for Italian. Final Workshop (EVALITA 2020), volume 2765, CEUR Workshop Proceedings (CEUR-WS.org), 2020. Conference date: 17-12-2020.

[17] F. Rodríguez-Sánchez, J. C. de Albornoz, L. Plaza, J. Gonzalo, P. Rosso, M. Comet, T. Donoso, Overview of EXIST 2021: sexism identification in social networks, Procesamiento del Lenguaje Natural 67 (2021) 195–207. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6389.

[18] F. Rodríguez-Sánchez, J. C. de Albornoz, L. Plaza, A. Mendieta-Aragón, G. Marco-Remón, M. Makeienko, M. Plaza, J. Gonzalo, D. Spina, P. Rosso, Overview of EXIST 2022: sexism identification in social networks, Procesamiento del Lenguaje Natural 69 (2022) 229–240. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6443.

[19] L. Plaza, J. Carrillo-de Albornoz, R. Morante, E. Amigó, J. Gonzalo, D. Spina, P. Rosso, Overview of exist 2023–learning with disagreement for sexism identification and characterization, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2023, pp. 316–342.

[20] W. S. Schmeisser-Nieto, P. Pastells, S. Frenda, M. Taule, Human vs. machine perceptions on immigration stereotypes, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 8453–8463. URL: https://aclanthology.org/2024.lrec-main.741.

[21] A. Fokkens, N. Ruigrok, C. Beukeboom, G. Sarah, W. Van Atteveldt, Studying muslim stereotyping through microportrait extraction, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018, pp. 3734–3741.

[22] J. J. Sánchez-Junquera, B. Chulvi, P. Rosso, S. P. Ponzetto, How do you speak about immigrants? taxonomy and stereoimmigrants dataset for identifying stereotypes about immigrants, Applied Sciences 11 (2021). URL: https://www.mdpi.com/2076-3417/11/8/3610. doi:10.3390/app11083610.

[23] J. Karoui, F. Benamara, V. Moriceau, V. Patti, C. Bosco, N. Aussenac-Gilles, Exploring the impact of pragmatic phenomena on irony detection in tweets: A multilingual corpus study, in: M. Lapata, P. Blunsom, A. Koller (Eds.), Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 262–272. URL: https://aclanthology.org/E17-1025.

[24] G. Corbelli, P. G. Cicirelli, F. D'Errico, M. Paciello, Preventing prejudice emerging from misleading news among adolescents: The role of implicit activation and regulatory self-efficacy in dealing with online misinformation, Social Sciences 12 (2023).

[25] F. D'Errico, P. G. Cicirelli, G. Corbelli, M. Paciello, Addressing racial misinformation at school: A psycho-social intervention aimed at reducing ethnic moral disengagement in adolescents, Social Psychology of Education (2023).

[26] C. Bosco, V. Patti, S. Frenda, A. T. Cignarella, M. Paciello, F. D'Errico, Detecting racial stereotypes: An italian social media corpus where psychology meets nlp, Information Processing & Management 60 (2023) 103118. URL: https://linkinghub.elsevier.com/retrieve/pii/S0306457322002199. doi:10.1016/j.ipm.2022.103118.

[27] T. Bourgeade, A. T. Cignarella, S. Frenda, M. Laurent, W. Schmeisser-Nieto, F. Benamara, C. Bosco, V. Moriceau, V. Patti, M. Taulé, A Multilingual Dataset of Racial Stereotypes in Social Media Conversational Threads, in: Findings of the Association for Computational Linguistics: EACL 2023, Association for Computational Linguistics, Dubrovnik, Croatia, 2023, pp. 686–696.

[28] E. Chierchiello, T. Bourgeade, G. Ricci, C. Bosco, F. D'Errico, Studying reactions to stereotypes in teenagers: an annotated italian dataset, in: Proceedings of the Fourth Workshop on Threat, Aggression and Cyberbullying (TRAC-2024), 2014.

[29] G. Lakoff, Women, fire, and dangerous things: What categories reveal about the mind, University of Chicago Press, Chicago, 1987.

[30] Y. Huang, Implicitness in the lexis, in: P. Cap, M. Dynel (Eds.), Implicitness: From lexis to discourse, John Benjamins, Amsterdam/ Philadelphia, 2017, pp. 67–94.

[31] C. Lyons, Definiteness, Cambridge University Press, Cambridge, 1999.

[32] S. Frenda, V. Patti, P. Rosso, Killing me softly: Creative and cognitive aspects of implicitness in abusive language online, Natural Language Engineering 29 (2023) 1516–1537. doi:10.1017/S1351324922000316.

[33] S. Frenda, V. Patti, P. Rosso, When sarcasm hurts: Irony-aware models for abusive language detection, in: A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, M. Aliannejadi, M. Vlachos, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer Nature Switzerland, Cham, 2023, pp. 34–47.