The Expression of Moral Values in the Twitter Debate: a Corpus of Conversations

Marco Stranisci[*]
Università degli Studi di Torino

Michele De Leonardis[**]
Università degli Studi di Torino

Cristina Bosco[†]
Università degli Studi di Torino

Viviana Patti[‡]
Università degli Studi di Torino

The present work introduces MoralConvITA, the first Italian corpus of conversations on Twitter about immigration whose annotation is focused on how moral beliefs shape users interactions. The corpus currently consists of a set of 1,724 tweets organized in adjacency pairs and annotated by referring to a pluralistic social psychology theory about moral values, i.e. the Moral Foundations Theory (MFT). According to the MFT, different configurations of moral values determines the stance of individuals on sensitive topics, such as immigration, civil rights, and gender equality. Moreover, an annotation of adjacency pairs' conversational dynamics is provided. Results of the analysis we applied on MoralConvITA shows that interesting patterns occur in the corpus, emerging from the intersection of moral studies and pragmatics that need to be generalized over larger corpora, and shedding some light on a novel promising perspective on the inter-user dynamics occurring in social media.

1. Introduction

The conversational nature of social media has been studied from several perspectives, among which, in the last years, community detection (Waseem and Hovy 2016; Lai et al. 2019; Vilella et al. 2020), and counter-speech analysis (Chung et al. 2019; Mathew et al. 2018; Fanton et al. 2021). Social media are indeed conversational environments where users and communities interact with each other, also producing conflictual situations, polarization, and sometimes toxic contents. That is the case of hate speech, that often affects the public online debate.

In particular, when people publicly debates about topics related to important societal challenges – those that trigger hatefulness – the conversation often takes the form of an exchange of moral values among social media users. In this context each single user can

[*] Dipartimento di Informatica - C.so Svizzera 185, 10149, Turin, Italy.
E-mail: marcoantonio.stranisci@unito.it

[**] Dipartimento di Studi Umanistici - via Sant Ottavio 20, 10124, Turin, Italy.
E-mail: michele.deleonard745@edu.unito.it

[†] Dipartimento di Informatica - C.so Svizzera 185, 10149, Turin, Italy.
E-mail: cristina.bosco@unito.it

[‡] Dipartimento di Informatica - C.so Svizzera 185, 10149, Turin, Italy.
E-mail: viviana.patti@unito.it

indeed provide his/her values for confirming or contrasting those expressed by other users.

Nevertheless, despite the large variety of computational linguistic resources developed in the last few years for detecting hate speech and a wide range of related phenomena (Poletto et al. 2021; Fortuna and Nunes 2018; Schmidt and Wiegand 2017), to the best of our knowledge, the joint observation of conversational aspects and moral values involved has not been the major focus of any of their annotations. The only exception is the *Moral Foundations Twitter Corpus*[1], the large corpus for English described in (Hoover et al. 2020).

In this paper, following the research line started in our previous work about hate speech detection and the development of corpora (Sanguinetti et al. 2018) and benchmarks for this task (Basile et al. 2019; Sanguinetti et al. 2020), we want to investigate the relationship between conversation and moral values in contexts where users debate about topics that can trigger hate. Inspired by the above mentioned corpus created for English and aiming at developing a resource currently missing for Italian, we introduce a novel Italian social media corpus, where conversation dynamics are modeled in the perspective of the involved moral foundations. We especially focus on this level for observing the conversational interaction between users, with the main aim to shed light on the possible influence that the moral concerns expressed by the first message of the adjacency pair can exert on the second one.

For this purpose, we selected a discourse domain related to an issue that we know as especially relevant to moral values and a topic with sufficient popularity among Twitter users. We focused in particular on three categories of people especially vulnerable to hate speech, namely Roma, ethnic and religious minorities, and we drawn 1,724 tweets from TWITA (Basile, Lai, and Sanguinetti 2018) by using a keyword-based filtering. In addition, we also collected and organized data so that they can keep a record of the conversation dynamic where they were originally generated by users. The dataset consists indeed of adjacency pairs (Schegloff and Sacks 1973; Simpson 2005) of tweets that form micro-conversations where a tweet and a reply are generated by Twitter users self labeling as against discrimination by using the hashtag *#facciamorete* on their screenname or user-description.

As far as the annotation of this corpus is concerned, the scheme mainly relies on the Moral Foundations Theory (MFT) categories. According to MFT, humans consistently rely on five moral concerns emerged as adaptive challenges: two individualizing foundations (Harm/Care and Fairness/Reciprocity) as they deal with the role of individuals within social groups, and three binding foundations (Ingroup/Loyalty, Authority/Respect, Purity/Sanctity), as they pertain to the formation and maintenance of group bonds (Weber and Federico 2013).

Going beyond the observation of moral concerns, we developed our annotation scheme also along two other directions that can better describe the inter-user interaction: the focus of the concern (on the violation or respect of the moral foundation expressed in the message) and the relation between the messages of the adjacency pair (whether the reply attacks, supports, or continues the conversation initiated by the tweet that is included in the same adjacency pair). An example annotated according to this scheme follows.

[1] The *Moral Foundations Twitter Corpus* is available at `https://osf.io/k5n7y/`

(1) **Tweet:** *Cara di #mineo sono 100mila euro al giorno per il business immigrazione, penso agli italiani in difficoltà... #portaaporta.*[2]
Concern expressed by the tweet: Ingroup-Loyalty/Betrayal
Focus: Prohibitive

Reply: *@user Paragonati ai 49 milioni di euro rubati dalla lega ancora pochi. A voglia di ospitare migranti..*[3]
Concern expressed by the reply: Fairness/Cheating
Focus: Prohibitive
Relation: Attack

As far as the suitability of the dataset within the context of applications, it has been observed (Kalimeri et al. 2019) that the detection of moral values together with other behavioral features of users might prove useful in general for designing more precise personalised services, communication strategies, and interventions, and can be used to sketch a portrait of people with similar worldview. Features based on moral concerns has been moreover proven to be useful in tasks related to sentiment analysis, see e.g. (Lai et al. 2021).

The paper is organized as follows. Section 2 briefly surveys related work, mostly focusing on MFT and its application in different contexts, and on pragmatics of conversation. Section 3 describes data collection and annotation, also discussing the inter-annotator agreement detected during the annotation process. Finally, Section 4 provides an analysis of moral and pragmatics features emerging from the gold standard corpus released. Section 5 concludes the paper and also addresses some future direction for the development of this research line.

2. Related Work and Theoretical foundations

According to the Moral Foundation Theory (MFT) individuals' moral beliefs are not universal, but reside on a plurality of "irreducible basic elements" that gives rise to many and sometimes conflicting moral configurations (Graham et al. 2013).
This theory unifies in five moral dyads the set of values originally proposed by Shweder (Shweder et al. 1997), i.e. *community*, *autonomy* and *sanctity*, and those discussed by Fiske (Fiske 1991), i.e., *communal sharing*, *authority tanking*, *equality matching* and *market pricing*. The moral dyads can be resumed as follows.

1. Care/Harm. Prescriptive concerns related to caring for others and prohibitive concerns related to not harming others.
2. Fairness/Cheating. Prescriptive concerns related to fairness and equality and prohibitive concerns related to not cheating or exploiting others.
3. Ingroup Loyalty/Betrayal. Prescriptive concerns related to prioritizing one's ingroup and prohibitive concerns related to not betraying or abandoning one's ingroup.

[2] Translation: #mineo's reception center they are 100thousands euros a day for the immigration business, I think to the Italians in distress ... #portaaporta

[3] Translation: @user Compared with 49 millions euros stolen by the Lega party they are still a few. You can host a lot of immigrants ...

4. Authority/Subversion. Prescriptive concerns related to submitting to authority and tradition and prohibitive concerns related to not subverting authority or tradition.
5. Purity/Degradation. Prescriptive concerns related to maintaining the purity of sacred entities, such as the body or a relic, and prohibitive concerns focused on the contamination of such entities.

The morality of each individual is built upon a specific configuration of these concerns that are considered within the theoretical framework as partly innate, partly developed through experience and social relationships. This allows MFT's dyads to describe morality as organized in advance of experience, highly dependent on environmental influences collected during development within a particular culture, and to see moral judgments as intuitions that happen before the subject starts to reason.

Nevertheless, like, e.g., the list of basic emotions, whose definition and granularity meaningfully varies in different theories, also the MFT's list of basic foundations can be questioned and it cannot be in effect considered as the final list. MFT is a theory in motion, to be expanded but especially adequate for cross-disciplinary research, because it provides a common language for talking about the moral domain (Graham and Haidt 2012) also in different disciplinary contexts. For instance, several researches within this framework have been devoted to investigate relations between moral foundations and political ideology, referring in particular to the moral differences between liberals and conservatives (Graham, Haidt, and Nosek 2009), media studies (Winterich, Zhang, and Mittal 2012).

In recent years, MFT foundations in the online environment have been studied by some scholar together with its correlation with other topics, such as hate speech (Hoover et al. 2019), or political discourse (Johnson and Goldwasser 2018; Weber and Federico 2013). Concurrently, several resources to investigate this phenomenon have been released: corpora of annotated tweets (Hoover et al. 2020), dictionaries (Graham and Haidt 2012; Hopp et al. 2020), and knowledge graphs (Hulpus et al. 2020).

An especially interesting application of this theory is the Moral Foundations Twitter Corpus (MFTC) (Hoover et al. 2020). It is a large collection of English tweets annotated for moral sentiment built for advancing research at the intersection of psychology and Natural Language Processing. The collection focuses on seven distinct socially relevant discourse topics, among which that addressed in our dataset, i.e. hate speech and offensive language. The schema applied in the annotation separates the virtues from the vices of the moral dyads to consider the polarity of a message expressing a value.

Although our approach is inspired by the MFTC's major tenets, it addresses a different language, i.e. Italian, and adopts a revised version of the annotation schema which is multi-dimensional. Dyads are not splitted, and the user's focus on moral values is evaluated separately from the selected dyad. Moreover also conversational dynamics are evaluated, since the corpus consists of adjacency pairs of tweets, instead of single messages, for keeping a record of the conversation dynamic where they were originally generated by users.

Adjacency pairs are units of conversation consisting of sequences of two adjacent utterance length, produced by different speakers (Schegloff and Sacks 1973). The two messages are complementary: the first pair part assumes a specific kind of response (Levinson 1983). For instance, if the initial message contains a request, the reply will presumably express the function of an acceptance or a refusal.

Similarly, the Dialogue Act (DA) is a communicative activity with a certain communicative function, a semantic content, and an optional feedback dependence relation function (Bunt et al. 2010). A family of computational pragmatics models focuses on the identification of lexical, collocational, syntactic, or prosodic cues for DA detection in a message (Jurafsky 2004). Several annotation schemes derive from such models, among which the Dialog Act Markup in Several Layers (DAMSL) (Core and Allen 1997), implemented with some modification by (Stolcke et al. 2000), and ISO 24617-2 (Bunt et al. 2012). All of them list a set of function for DAs annotation. Recently, iLISTEN, a shared task for Italian consisting in automatically annotating dialogue turns with speech act labels, representing the communicative intention of the speaker, has been propose at the EVALITA evaluation campaign (Basile and Novielli 2018). The speech act taxonomy refines the DAMSL categories, based on two classes of functions (Cfr (Allen and Core 1997)): Forward Looking, the intended action expressed by the first pair part, and Backward Looking, which encodes how the reply is related with the original message. In our corpus, adjacency pairs internal structure often consists in a statement on immigration, accepted or rejected in the reply.

In our schema, replies are annotated with 'attack', that may imply rejection, 'support' and 'same topic', which can entail acceptance. However, these categories are not overlapping, since attacking, and supporting potentially fulfil other relevant functions to our work, such as outlining the moral or political stance of the speaker. Though, the two schema are mapped to support the qualitative analysis of conversational dynamics in Section 4.2.

3. MoralConvITA: A Corpus of Conversations with Annotated Moral Foundations

3.1 Data

In order to create the MoralConvITA corpus, a sample of 862 adjacency pairs of tweets were collected from January 2019 to June 2020. The data gathering process relied on the TWITA data set (Basile, Lai, and Sanguinetti 2018), and was structured as follows:

- all tweets generated by users self labeling as against discrimination with the hashtag *#facciamorete* on their screen-name or user-description were collected;
- the resulting selection was further filtered by using the Hate Speech corpus keywords (Sanguinetti et al. 2018);
- only reply messages were kept;
- first pair parts were retrieved through the Twitter Rest APIs.

In order to collect a meaningful amount of data where moral sentiment occurs, we choose a discourse domain related to an issue that we know as especially relevant to moral values and a topic with sufficient popularity among Twitter users. Nevertheless, considering that expressions of moral sentiment in one domain and about a specific topic might not generalize to data extracted from another domain, in future work we want to address other domains also.

3.2 Annotation

The task of annotating a corpus according to the MFT shares similarities with sentiment classification, but it also introduces notable challenges, such as the co-occurrence of many moral values in a message, their implicitness and subjectivity (Hoover et al. 2020). For addressing these challenges we discussed the design of the schema within the research group and we performed annotation trials on a small subset of the data before starting with the actual annotation process. Finally, for validating the schema, we carefully observed the behavior of each annotator and the agreement among the annotators, as reported in Section 4.

The schema we provided for MoralConvITA is centered on the MFT and under this respect inspired by the one applied in the Moral Foundations Twitter corpus for English. Moreover, in order to take into account the pragmatics of conversation, we defined also some other issue to be annotated for better representing the conversation dynamics. Three are the dimensions along which we annotated the adjacency pairs.

1. the most relevant **Moral Foundation** dyad, among the five pointed out by the MFT (Section 2);
2. the **Concern Focus** of the message, which may be prescriptive, if it highlights a virtue, or prohibitive, if it blames a misbehavior;
3. the **Conversational Relation** within the adjacency pair, representing whether the reply attacks, support or deals with the same topic of the first pair part.

Table 1 resumes the list of labels used for the annotation of each of these three categories.

Table 1
The labels annotated in the MoralConvITA

category	label
Moral Foundation	Care/Harm
	Fairness/Cheating
	Ingroup-Loyalty/Betrayal
	Authority/Subversion
	Purity/Degradation
Concern Focus	prescriptive
	prohibitive
Conversational Relation	attack
	support
	same topic
	no relation

Conversational Relation and Concern Focus dimensions were elaborated to better fit the annotation schema to the analysis of Twitter conversations. The former provides information about how the pairs of tweets relate to each other. The Concern Focus, instead, was introduced to mitigate the dichotomy between moral vices and virtues. In existing schemas a text can either express the respect for a moral concern or the stigmatization of its violation, but this distinction seems not to capture expressions that deliberately violate a moral value and may have a pragmatic effect. On this respect toxic speech, that often affects the conversation about migration, can be interpreted as a blatant violation of the Care/Harm dyad. Hence, we considered the Concern Focus as an independent dimension to annotate. For instance, instead of considering 'care', and 'harm' two separated labels, we treated them as a whole, and later evaluate their focus, that is 'prescriptive' if the message dwells on the moral rule to comply with, 'prohibitive' when its violation is reported by the user. Examples of tweets expressing moral dyads and their Concern Focus are listed in Table 2, while Conversational Relations are exemplified in Table 3.

It is worth highlighting some strategy we applied in the annotation. First, in addition to the dyads of the MFT we also used for the category Moral Foundation the label 'no-moral' when any moral concerns occurs in the message. Second, as far as the concern, it is annotated only in the messages where a moral foundation has been previously recognized by the annotator. Finally, the conversational relation is only annotated in the reply message for showing its link with the tweet that started the micro-conversation.

The annotation process involved a team composed of two skilled researchers, a man and a woman, and nine undergraduate university students, among which 3 men, and 6 women, aged 22-27. The skilled annotators were especially involved in designing and testing the schema, in tutoring the rest of the annotation and in solving the disagreement. Each of the nine students annotated at least 250 adjacency pairs along the three dimensions for building the corpus we actually released[4] which includes 1,724 tweets, organized in 862 adjacency pairs.

The analysis of inter-annotator agreement (IAA), calculated using the the Fleiss' Kappa IAA metrics and considering each of the categories annotated, is described in Table 4. It confirms the subjectivity of the task, which also results from the observations reported in (Hoover et al. 2020) for the Moral Foundations Twitter corpus for English. Considering that our corpus is organized in micro-conversations, we can report also some findings about the agreement detected in the perspective of the annotation of the conversations that compose MoralConvITA. In particular, the results provided in Table 4 highlight that the annotation of replies of the adjacency pairs has been affected by an also lowest agreement (0.17 for the Moral Foundation, and 0.18 for its Focus) with respect to the annotation of the tweets that initiate the conversation, while for the others it shows a fair agreement. The issue has been already pointed out by (Hoover et al. 2020) with respect to the development of the Moral Foundation Twitter Corpus. According to this study, the interpretation of morality in a text is subjective both for the annotators' stance and for the lack of information about the author's intention. Moreover, this low agreement among the annotators can be also motivated by the fact that the moral concern expressed in a message is often ambiguous because many values potentially coexist within the text. See for instance the following example.

[4] https://github.com/marcostranisci/MoralConvITA

Table 2
Moral values annotated in the MoralConvITA corpus.

Moral Value	**Example**
Care/Harm	*@user Infatti lo dicevo perché entrambi erano cristiani! Concordo con lei che prima ci sono le Persone, che possono essere più o meno brave, cristiane o no, alte o basse...* (@user In fact I said it because both were Christians! I agree with you that first of all there are the individuals, who can be more or less good, Christian or not, high or low...)
Fairness/Cheating	*@user Ehm...e la crisi, la disoccupazione giovanile, sanità,strutture, l'istruzione. Queste non sono emergenze? No no.* (@user Ehm...and the crisis, youth unemployment, health, facilities, education. These are not emergencies? No no.)
Ingroup Loyalty/Betrayal	*Immagini esclusive di un gommone con 70 immigrati, scafista alla guida e motore potente, in acque maltesi. Qualcuno si degnerà di intervenire o li manderanno ancora una volta in direzione Italia???* (Exclusive images of a dinghy with 70 immigrants, a driver and powerful engine, in Maltese waters. Will someone deign to intervene or) will they send them once again to Italy?
Authority/Subversion	*A casa fanno la voce grossa e mostrano i muscoli con i disperati. A Bruxelles invece Salvini e company sono solo pecorelle di #Orban che è il primo nemico dell'Italia e che nega ogni giorno i nostri valori costituzionali.* (At home they speak louder and show their muscles with the desperates. In Brussels instead Salvini and company are only sheeps of #Orban who is the first enemy of Italy and who denies every day our constitutional values.)
Purity/Degradation	*#Iran, migliaia di prigionieri politici subiscono torture e maltrattamenti senza cure mediche. Libertà per #ArashSadeghi #FarhadMeysami #RajaeeShahr e per tutti i dissidenti che non si arrendono al regime khomeinista.* (#Iran, thousands of political prisoners suffer torture and ill-treatment without medical treatments. Freedom for #ArashSadeghi #FarhadMeysami #RajaeeShahr and for all dissidents who do not belong to the Khomeinist regime.)

> (2) *Se io sono cittadino italiano non #Rom, allo Stato devo dire: dove abito, da quando ci abito, se sono sposata oppure no, quanti soldi ho in banca, devo pagare fino all'ultimo centesimo di tasse e se non faccio i vaccini mi denunciano. Scusate si può fare per tutti?*[5]

It could be intended as an instance of Ingroup-Loyalty/Betrayal, since it highlights a contrast between an ingroup (Italians) and an outgroup (Roma people). However, it

[5] If I am an Italian citizen, and not a #Roma person, I must declare to my country: my place of residence, since when I live there, If I am married or not, my account balance, I have to pay every cent of taxes, and if I don't take the vaccine I am reported. Excuse me, this can be done for everybody?

Table 3
Relations linking tweets and replies annotated in the MoralConvITA corpus.

Conversational pattern	**# Example**
Support	Tweet: *Oggi scopriamo dal Ministro Salvini che c'è una "questione rom" aperta. Ed io che pensavo che ci fosse invece una "questione MAFIA" aperta. O una "questione CORRUZIONE". E invece, dopo i migranti, si punta il dito contro un'altra minoranza. La miseria umana è tutta qui.* Reply: *@user è il suo standard, prima i meridionali, poi gli emigranti e adesso i Rom... chissà chi punterà prossimamente..* (Tweet: Today we discover from Minister Salvini that there is an open "Roma issue". And I thought that there was instead an open "MAFIA issue". Or a "question CORRUPTION". And instead, after the migrants, you point the finger at another minority. Human misery is all here. Reply: @user it is his standard, first the southerners, then the emigrants and now the Roma...who knows who will point soon..)
Attacks	Tweet: *Da oggi anche l'Italia comincia a dire NO al traffico di esseri umani, NO al business dell'immigrazione clandestina. Il mio obiettivo è garantire una vita serena a questi ragazzi in Africa e ai nostri figli in Italia.* Reply: *@matteosalvinimi NO alla propaganda fatta sulla pelle dei migranti! Più di seicento persone abbandonate in mare per scrivere un tweet? Vergognati per la tua disonestà.* (Tweet: From today Italy too begins to say NO to human trafficking, NO to the business of the illegal immigration. My goal is to ensure a peaceful life for these children in Africa and our children in Italy. Reply: @matteosalvinimi NO to the propaganda that negatively affect migrants! More than six hundred people banded at sea to write a tweet? Be ashamed of your dishonesty.)
Same topic	Tweet: *"Se si torna al voto come prima cosa dovremo costituire un fronte largo europeista da contrapporre al fronte anti-europeista di #Salvini e #DiMaio. L'Europa sarà la discriminante. #maratonamentana"* Reply: *@user Da nord a sud le elezioni le vincerà di nuovo chi farà propaganda anti-migranti.. è questo il nodo fondamentale purtroppo!!* (Tweet: "If we return to the vote as a first thing, we must form a broad pro-Europe front to join the anti-European front of #Salvini and #Dimaio. Europe will be the discriminating. #maratonamentana" Reply: @user From north to south the elections will be won again by those who make anti-migrant propaganda.. this is the fundamental issue, unfortunately!!)
No-relationship	Tweet: *C'ho i parenti fasci e razzisti, mi vergogno tantissimo.* Reply: *@user Fino a quando parli di diritti, migranti, accoglienza e amenità del genere nessuno cambierà idea. Se ai neosalviniani metteranno le mani in tasca, allora, potrai di nuovo discuterci. #SalviniDimettiti* (Tweet: I have family members fascists and racists, I'm so ashamed. Reply: @user As long as you talk about rights, migrants, reception and amenities of the generationnobody will change their mind. If the Neosalvinians get their hands in their pockets, then you can discuss it again. #Salvinigohome)

Table 4
Fleiss' Kappa for each label separately calculated for the tweet (which initiates the micro-conversation) and for the reply to the tweet.

label	Fleiss' Kappa
Moral Foundation (tweet)	0.32
Moral Foundation (reply)	0.17
Concern Focus (tweet)	0.26
Concern Focus (reply)	0.18
Conversational Relation (reply only)	0.30

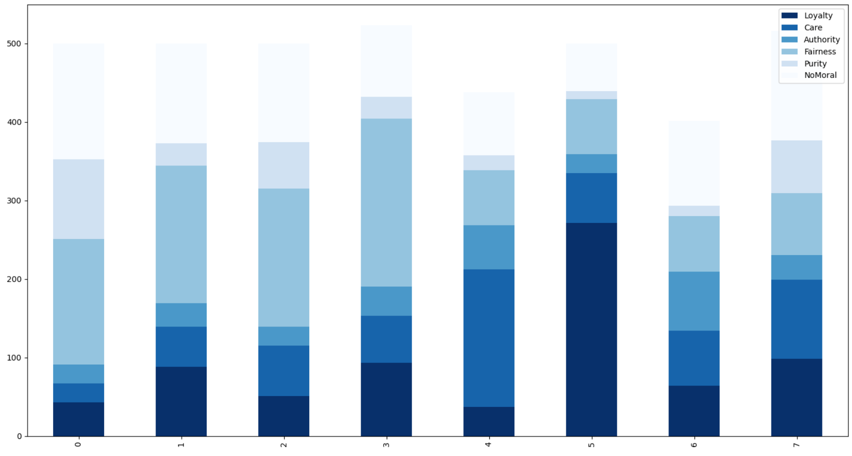

Figure 1
The distribution of moral foundations labels in the corpus (in tweets and in replies both) separately calculated for each of the 8 annotators (referred with numbers from 1 to 8).

could also express a concern on authority for its reference to the need of respecting country laws and therefore annotated with the label Authority/Subversion.
Separately calculating the distribution of the moral foundations for each annotator and putting together the tweets and the replies (as we did in Figure 1), we can see that some bias occurs and that some annotator used a very large amount of some label with respect of the average of the annotators. For instance annotator 9 used Ingroup-Loyalty/Betrayal more that twice that the other annotators.
A more general bias was moreover expected in our annotation, which depends on the involved annotators. Their age and skill is related in literature with a basically liberal vision, rather than to a conservative one, and to the exploitation of some specific moral foundation in the interpretation of messages. While conservative people tends to use all the moral foundations of the MFT spectrum, liberal people only rely its judgement

on the first ones, mostly Care/Harm, Fairness/Cheating and Ingroup-loyalty/Betrayal. This is confirmed by the analysis provided in the next section.

4. Analysis of the MoralConvITA Corpus

The final version of the MoralConvITA corpus consists of $1,724$ tweets arranged in adjacency pairs annotated by at least three annotators, but only in 487 cases annotators reached a partial or total agreement on all the five dimensions of the annotation schema. Excluding all the tweet labeled as 'no-moral', the corpus reduces to 253 adjacency pairs. We thus chose to separately analyze moral foundations, moral focus, and conversational patterns.

The distribution of Moral Foundations in the adjacency pairs is discussed in Section 4.1, while an analysis of how foundations are shaped by the conversation is provided in Section 4.2.

4.1 Moral Foundations

The distribution of the labels annotated will be analyzed in this section according two different perspectives, that is the moral dyads provided by the annotators and their occurrence in the tweets rather than in the replies, as shown in figure 2.

Two are the prevalent moral foundation dyads annotated in the corpus: **Fairness/Cheating** (408 occurrences), and **Loyalty/Betrayal** (255 occurrences). They both seem to be very specific to the topic of migration, since the latter draws a distinction between who is Italian and who is not, while the former is often used to report the hypocrisy of public players that deal with this topic.

In particular, the accuse of cheating follows two rhetorical patterns: the reception of asylum-seekers as a business, mainly occurring in original tweets, and the exploitation of migration for political propaganda, occurring in replies. This second case is more traditional in the Italian public debate, and most common in it. In fact, 67.7% of Fairness/Cheating labels occurs in replies, most of them focused on the anti-immigration proposals' inconsistency, and lack of actual effectiveness. For the same reason, the 9.6% of in-agreement adjacency pairs consists of a statement oriented to the Ingroup-loyalty/Betrayal value, and a response in which the Fairness/Cheating concern is present.

The 'immigration as a business' moral charge, more frequent in the first element of the pair, is a quite recent rhetorical argument, but its fast diffusion could be interpreted as a reshape of the traditional separation between liberals and conservatives (Haidt 2012). For instance, in the example (3), irregular migrants are depicted as victims of a foul game by pro-immigration organization, and the closing invective contains an exhortation to help them not only in words, but also in a concrete way.

> (3) *@user Tutto inutile, lei sarà arrestata, la nave sequestrata ed i clandestini usati per il vostro sporco giochino, sparsi come buste di spazzatura sulle strade. Aprite le vs porte di casa invece che cavarvela con 15 euro a testa. Maledetti.*[6]

[6] @user All for nothing, she will be arrested, the ship seized, and irregular migrants used for your foul game, scattered as trashbags on the streets. Instead of getting by 15 euros each, open your homes. You, damn.

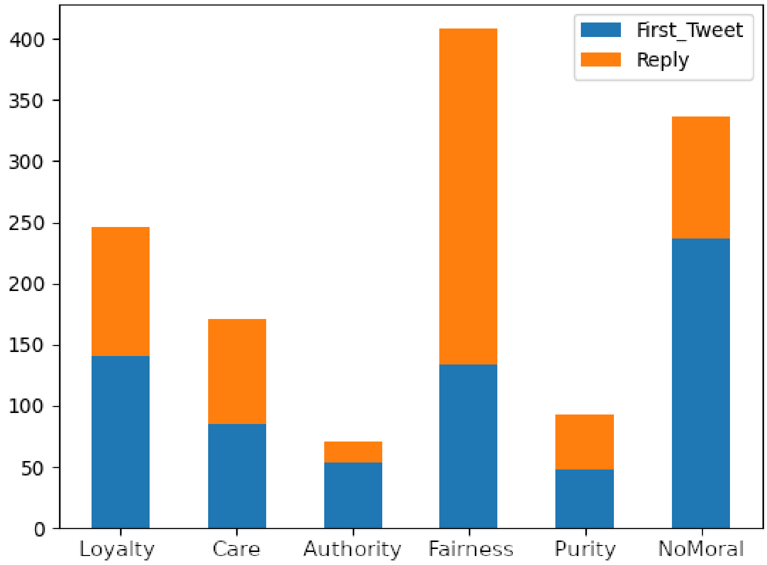

Figure 2
The distribution of moral concerns labels in the corpus (in tweets and in replies).

With 151 occurrences, **Care/Harm** is the third most prevalent moral concern in the corpus. Both when in the original tweet or in a reply, this value is almost always a pro-immigration stance signal as opposed to Fairness/Cheating or Loyalty/Betrayal. However, rare but interesting is the use of this moral concern to justify migrants rejections as a way to save their lives from human traffickers, as can be seen in (4).

(4) *@user i morti si moltiplicano per colpa di chi incoraggia il traffico di clandestini*[7]

In order to further understand whether there are linguistic clues signaling the correlation between the immigration topic and the three moral foundations more often annotated in the corpus, we calculated the **weirdness index** (Ahmad, Gillam, and Tostevin 1999; Florio et al. 2020), a technique that allows the retrieval of the most frequent and characterizing words within a specialized corpus of texts by contrasting it with a more general purpose dataset.

First, we calculated the relative frequency of each word in messages labeled with a given moral value, then we applied the same technique on the rest of the corpus. Finally, we computed the ratio between the two frequencies. This returned all the tokens that are frequent in messages annotated with a specific moral value, and occur less in other tweets of the corpus. In Table 5 a selection of the most specific words is listed.

Without forgetting the effect that the limited size of the corpus can have on the

[7] @user deaths are multiplying due to people who encourage human traffick

validity of the index, some interesting signals can be drawn from this quantitative analysis. As expected, the words *reato, cattivo contribuente, corruzione* seem to correlate with the Fairness/Cheating domain, as well as *connazionali, sicure, tolleranzazero* with Loyalty/Betrayal, and *valori, restiamoumani, shoah* with Care/Harm.

We think that future work based on a larger dataset can provide the further and necessary evidence to these results, confirming the relationships between textual expressions and moral concerns.

Table 5
The most relevant words of MoralConvITA according to the weirdness-index calculation.

Fairness/Cheating	Loyalty/Betrayal	Care/Harm
reato	vivono	riace
fornero	bar	valori
dimaio	crimine	restiamoumani
cattivo	cambiato	raccontare
contribuente	stabile	shoah
inps	deliranti	passato
speso	connazionali	sanremo2019
corruzione	sicure	mediterraneo
stupro	buonisti	ponte
redditodicittadinanza	tolleranzazero	emigranti

4.2 Moral Foundations in Twitter Conversations

The application of the MFT framework for analyzing Twitter conversations resulted in the introduction of two additional dimensions to the annotation schema. The Concern Focus supports a more thorough investigation of how a message expresses the position of a user about a given moral foundation; the Conversational Relation allows to explore the conversational dynamic within the adjacency pair.

Concern Focus. For each tweet in the corpus expressing a moral dyad, the Concern Focus was annotated by choosing among the 'prescriptive' or 'prohibiting' label. Examples (5) and (6) were both annotated with the authority/subversion dyad, but the first with a 'prescriptive' focus, since it highlights respect for the law, while the second with a 'prohibitive' focus, as it is a critique to the government.

> (5) *Chiunque sfrutta l'immigrazione clandestina per riempirsi le tasche va PUNITO in maniera esemplare, senza se e senza ma.*
> *Complimenti a Carabinieri e Guardia di Finanza per l'operazione.*
> *Anche per gestori e cooperative in malafede, è finita la pacchia!*[8]

[8] Anyone who uses illegal immigration to line their own pockets should be PUNISHED in an exemplary manner, no ifs or buts.
Congratulations to Carabinieri and Guardia di Finanza for the operation.
Even for managers and cooperatives in bad faith, the free ride is over!

> (6) *A casa fanno la voce grossa e mostrano i muscoli con i disperati. A Bruxelles invece Salvini e company sono solo pecorelle di #Orban che è il primo nemico dell'Italia e che nega ogni giorno i nostri valori costituzionali.*[9]

Unlike existing resources, which provide data-driven support for studying the psychological aspects of morality, Moral ConvITA mainly focuses on how this phenomenon is expressed in texts. The approaches are complementary and may lead to different interpretations of a message. For instance, (7) can be interpreted as a violation of the Loyalty/Betrayal principle since it highlights a conflict between Christians and Muslims. Conversely, in our corpus the tweet was annotated as conveying the Care/Harm dyad with a prescriptive focus, because the violation of the principle of care is not only present but also suggested, as it happens in many examples of HS. Similarly, (8) expresses Loyalty/Betrayal with a prohibitive focus. However, instead of being a stigmatization of somebody betraying her/his group, it reports the mediatic emphasis on crimes committed by migrants.

> (7) - *@matteosalvinimi Pena di morte per i musulmani, TUTTI.*[10]

> (8) *@user Aspetta che sia un immigrato preferibilmente di colore ad ammazzare la prossima donna e si scatena il #Capitonedatastiera. L'omicida è un italiano? Quattro righe in cronaca, taglio basso e via la notizia dopo il primo lancio. Funziona così...*[11]

The distribution of the focus is generally skewed on prohibition. According to the annotation, only 273 out of $1,179$ focuses on the moral rule observance, which corresponds to 23%. The disproportion is more accentuated in replies, among which 81% of messages dwells on the violation of a moral rule.

The distribution differs when the intersection of the focus and the moral dyad is considered. While 86% of messages expressing Fairness/Cheating is also prohibitive (91% in replies), the annotated focus for the Care/Harm dyad is balanced. Finally, the presence of a prohibitive focus together with Loyalty/Betrayal values occurs in 76% of data, on average with the overall distribution, even if it is more rare in replies (71%). A deeper analysis should be performed in order to understand whether these numbers are the product of the topic, the contextual constraints of the social media where the conversations take place, or both. Moreover, a fine-grained annotation schema for this dimension is needed to capture a richer set of morally oriented communication functions.

Conversational Relation. The relation between two tweets in an adjacency pair could be either annotated as 'attack', 'support', 'continue' or 'no relation' (see Table 1). This dimension supports the analysis of the acceptance or rejection of messages expressing moral values (Section 2).

The quantitative analysis of these conversational patterns show a high prevalence of rejections to the original statement. 378 out of 786 in-agreement conversational pattern

[9] At home they make a show of force and flex their muscles with desperate people. In Bruxelles, instead, Salvini and company are sheeps of #Orban, who he is Italy's first enemy and who negates constitutional rights every day.

[10] - @matteosalvinimi Death penalty for Muslims, ALL.

[11] @user Wait for a preferably black immigrant to kill the next woman and unleashes the #Snakefromthekeyboard. The murderer is an Italian? Four lines in the news, low profile and off the news after the first launch. It works like this...

was marked as an attack to the first element of the pair, while labels 'support', and 'same topic' collected together 360 annotations. The disproportion may also be larger because 33% of first pair elements in the corpus are replies themselves. Hence the adjacency pair may consist of two rejection responses to an original tweet which was not collected in the corpus (9).

> (9) - *@user1 @user2 Impressionante superficialità. Più che Ministro....uno sceriffo. Caspita che cambiamento. - @user3 @user4 @user5 È diventato il #ministrodellimmigrazione. Altro non gli interessa...vedi #camorra #Ndrangheta #sacracoronaunita etc etc...*[12]

When the conversational relation and the moral dyad expressed by a reply are considered together, the number of adjacency pairs that can be usefully exploited for our analysis is reduced from 862 to 468, due to the low inter-annotator agreement (Table 6). In this subset the number of attacks increases by 8%, while the percentage of supporting replies is stable. As for the analysis of the concern focus, the distribution of conversational relations differs according to the moral dyad. More specifically, there are less messages annotated as expressing Loyalty/Betrayal and an attack at the same time.

The joint presence of a moral dyad in the first element of the pair and the conversational relation leads to a more important reduction of adjacency pairs that can be analyzed, since they are reduced to 417. In this subset it is worth mentioning the 78% of first elements expressing Loyalty/Betrayal and being attacked, that is more than 30% above messages conveying Care/Harm or Fairness/Cheating foundations.

Table 6
The joint distribution of moral dyads and conversational relations in the corpus.

	Care	Fairness	Authority	Loyalty	Purity	Total
1st tweet & attack	29	54	34	103	22	242
1st tweet & support	16	33	9	13	12	83
1st tweet & continue	27	33	8	15	9	92
reply & attack	43	144	5	44	28	264
reply & support	16	40	5	12	6	79
reply & continue	17	63	4	35	6	125

Questions as forms of Moral Rejection. Prosodic cues seem also to correlate with the presence of an attack in replies: 205 out of 378 rejection messages contain indeed a questions, while in 360 supporting responses there are only 138 questions.

Many of them appear to convey ironic statements, such as *quindi adesso i migranti possono*

[12] - @user1 @user2 Impressive superficiality. More than Minister.... a sheriff. Wow that change.- @user3 @user4 @user5 He has become the #MinisterofImmigration. He doesn't care about anything else...see #camorra #Ndrangheta #sacracoronaunita etc etc...

affogare in pace senza che nessuno li soccorra?[13]. Others may be considered pragmatic rejections (Schlöder and Fernández 2015), namely utterances whose interpretation relies on information to be drawn from the context. For instance, the foundation expressed by the question *@giorgiameloni difendere da chi?*[14] is recognizable only along with the exhortation in the first element of the pair: *Avanti insieme per difendere l'Italia!*[15]. Hence, the question conveys a stigmatization of the Loyalty/Betrayal dyad.
The interpretation of some message can be more problematic, like for instance *matteosalvinimi user con 49 milioni di euro sai quanti migranti ospito, matteo?*[16], since external knowledge is needed to infer the Fairness/Cheating dyad from this question.
Finally, the detection of moral values expressed in a question may be supported by dialogical repetition (Bazzanella 2017). In *a lei il passato cosa ha insegnato?*, the repetition of the word 'passato/past' from the first message of the pair - *Il Governo sostiene tutte le iniziative in memoria della #Shoah, perché il passato ci insegni a combattere ogni forma di discriminazione e di odio*[17] - is a cue of rejection. The first element of the pair, focused on the Care/Harm foundation, is challenged by a reply expressing Fairness/Cheating, since it seems to highlight the interlocutor's inconsistency.

The analysis of MFT in Twitter conversations shows some promising results. Considering the Concern Focus a separated dimension from foundations brought out a richer taxonomy of moral expressions that may be useful in understanding how specific moral stances interact with the spreading of toxic contents, as it emerges in the example (7). The conversational relation in adjacency pairs, especially when jointly investigated with dyads, appeared to show that some foundation are most likely to be rejected by the interlocutor, while others are more adopted to communicate disagreement. A preliminary analysis of questions as device for conveying a moral conflict emphasised the need of providing a fine-grained analysis of dyads are shaped within the conversation.

5. Conclusion and Future Work

This paper describes a novel Italian resource which is a collection of micro-conversations drawn from Twitter (adjacency pairs of messages, i.e. a tweet and its reply) and annotated for making explicit the occurrence of moral values and the conversational dynamics. The annotation scheme includes indeed moral concerns as categorized within the Moral Foundations Theory, the focus of each of the annotated moral concern and the relation that links the reply to a tweet in the conversation. As far the topic on which the corpus is focused, we selected a discourse domain related to an issue that we know as especially relevant to moral values and a topic with sufficient popularity among Twitter users, i.e. immigrants.

The main aim of making available this resource to the computational linguistics research community is at providing a missing dataset for Italian and at discussing some currently underrepresented phenomena that collocate at the intersection of social psychology, linguistics and conversational analysis.

[13] So now migrants can drown in peace without anyone helping them?

[14] @giorgiameloni, defend from whom?

[15] Forward together to defend Italy!

[16] @matteosalvinimi @user with 49 million euros do you know how many migrants I host, matteo?

[16] What has the past taught you?

[17] The Government supports all the initiatives in memory of the #Shoah, so that the past teaches us to fight all forms of discrimination and hatred

Nevertheless, considering that the expressions of moral sentiment in one domain and about a specific topic hardly generalize to data extracted from another domain, in future work we want to address other domains, e.g., misogyny, by collecting more data and by testing on them the scheme we propose in this paper.

References

Ahmad, Khurshid, Lee Gillam, and Lena Tostevin. 1999. University of Surrey Participation in TREC8: Weirdness Indexing for Logical Document Extrapolation and Retrieval (WILDER). In *The Eighth Text REtrieval Conference (TREC-8)*, Gaithersburg, MD, US, January. National Institute of Standards and Technology (NIST).

Allen, James and Mark Core. 1997. Draft of DAMSL: Dialog act markup in several layers.

Basile, Pierpaolo and Nicole Novielli. 2018. Overview of the Evalita 2018 itaLIan Speech acT labEliNg (iLISTEN) Task. In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018)*, volume 2263 of *CEUR Workshop Proceedings*, Turin, Italy, December. CEUR-WS.org.

Basile, Valerio, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco Manuel Rangel Pardo, Paolo Rosso, and Manuela Sanguinetti. 2019. SemEval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter. In *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*, pages 54–63, Minneapolis, Minnesota, US, June. "Association for Computational Linguistics".

Basile, Valerio, Mirko Lai, and Manuela Sanguinetti. 2018. Long-term social media data collection at the University of Turin. In *5th Italian Conference on Computational Linguistics (CLiC-it 2018)*, volume 2263 of *CEUR Workshop Proceedings*, pages 1–6, Turin, Italy, December. CEUR-WS.

Bazzanella, Carla. 2017. Dialogic repetition. In *Dialoganalyse IV, Teil 1*. Max Niemeyer Verlag, pages 285–294.

Bunt, Harry, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, et al. 2010. Towards an ISO standard for dialogue act annotation. In *11th International conference on Language Resources and Evaluation (LREC 2010)*, pages 1787–1794, Valletta, Malta, May. European Language Resources Association (ELRA).

Bunt, Harry, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David R. Traum. 2012. Iso 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, pages 430–437, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Chung, Yi-Ling, Elizaveta Kuzmenko, Serra Sinem Tekiroglu, and Marco Guerini. 2019. Conan-counter narratives through nichesourcing: a multilingual dataset of responses to fight online hate speech. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2819–2829, Florence, Italy, July-August.

Core, Mark G. and James Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, volume 56, pages 28–35, Boston, MA, US, November.

Fanton, Margherita, Helena Bonaldi, Serra Sinem Tekiroğlu, and Marco Guerini. 2021. Human-in-the-loop for data collection: a multi-target counter narrative dataset to fight online hate speech. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3226–3240, Online, August. Association for Computational Linguistics.

Fiske, Alan P. 1991. *Structures of social life. The four elementary forms of human relations: communal sharing, authority ranking, equality matching, market pricing*. Free Press.

Florio, Komal, Valerio Basile, Marco Polignano, Pierpaolo Basile, and Viviana Patti. 2020. Time of your hate: The challenge of time in hate speech detection on social media. *Applied Sciences*, 10(12).

Fortuna, Paula and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, 51(4):1–30.

Graham, Jesse and Jonathan Haidt. 2012. The moral foundations dictionary.

Graham, Jesse, Jonathan Haidt, Matt Motyl, Sena Koleva, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto. 2013. Chapter two - moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47:55–130.

Graham, Jesse, Jonathan Haidt, and Brian A. Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5):1029.

Haidt, Jonathan. 2012. *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Hoover, Joe, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, et al. 2020. Moral Foundations Twitter Corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071.

Hoover, Joseph, Mohammad Atari, Aida M Davani, Brendan Kennedy, Gwenyth Portillo-Wightman, Leigh Yeh, Drew Kogon, and Morteza Dehghani. 2019. Bound in hatred: The role of group-based morality in acts of hate, Jul. 10.31234/osf.io/359me.

Hopp, Frederic R., Jacob T. Fisher, Devin Cornell, Richard Huskey, and René Weber. 2020. The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, pages 1–15.

Hulpus, Ioana, Jonathan Kobbe, Heiner Stuckenschmidt, and Graeme Hirst. 2020. Knowledge graphs meet moral values. In *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*, pages 71–80, Barcelona, Spain (Online), September. Association for Computational Linguistics (ACL).

Johnson, Kristen and Dan Goldwasser. 2018. Classification of moral foundations in microblog political discourse. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 720–730, Melbourne, Australia, July. Association for Computational Linguistics (ACL).

Jurafsky, Daniel. 2004. Chapter 26: Pragmatics and computational linguistics. In *The handbook of pragmatics*. Wiley Online Library, pages 578–604.

Kalimeri, Kyriaki, Mariano G. Beiró, Matteo Delfino, Robert Raleigh, and Ciro Cattuto. 2019. Predicting demographics, moral foundations, and human values from digital behaviours. *Computers in Human Behavior*, 92:428–445, March.

Lai, Mirko, Marco Antonio Stranisci, Cristina Bosco, Rossana Damiano, and Viviana Patti. 2021. HaMor at the Profiling Hate Speech Spreaders on Twitter. In *Notebook for PAN at CLEF 2021*, volume 2936 of *CEUR Workshop Proceedings*, Online, September. CEUR-WS.

Lai, Mirko, Marcella Tambuscio, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso. 2019. Stance polarity in political debates: A diachronic perspective of network homophily and conversations on Twitter. *Data Knowledge Engineering*, 124.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Mathew, Binny, Navish Kumar, Pawan Goyal, Animesh Mukherjee, et al. 2018. Analyzing the hate and counter speech accounts on Twitter. *arXiv preprint arXiv:1812.02712*.

Poletto, Fabio, Valerio Basile, Manuela Sanguinetti, Cristina Bosco, and Viviana Patti. 2021. Resources and benchmark corpora for hate speech detection: a systematic review. *Language Resources and Evaluation*, 55:477–523.

Sanguinetti, Manuela, Gloria Comandini, Elisa Di Nuovo, Simona Frenda, Marco Stranisci, Cristina Bosco, Tommaso Caselli, Viviana Patti, and Irene Russo. 2020. Haspeede 2@evalita2020: Overview of the evalita 2020 hate speech detection task. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings o,f the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online, December. CEUR.org.

Sanguinetti, Manuela, Fabio Poletto, Cristina Bosco, Viviana Patti, and Marco Stranisci. 2018. An Italian Twitter Corpus of Hate Speech against Immigrants. In Nicoletta Calzolari (Conference chair), Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Schegloff, Emanuel A. and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 8(4):289–327.

Schlöder, Julian J. and Raquel Fernández. 2015. Pragmatic rejection. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 250–260, London, UK, April. Association for Computational Linguistics (ACL).

Schmidt, Anna and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the fifth international workshop on natural language processing for social media*, pages 1–10, Valencia, Spain, April. Association for Computational Linguistics (ACL).

Shweder, Richard A., Nancy C. Much, Manamohan Mahapatra, and Lawrence Park. 1997. The "big three" of morality (autonomy, community and divinity), and the "big three" explanations of suffering. In A. Brandt and P. Rozin, editors, *Morality and health*. Routledge, pages 119–169.

Simpson, James. 2005. Conversational floors in synchronous text-based CMC discourse. *Discourse studies*, 7(3):337–361.

Stolcke, Andreas, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.

Vilella, Salvatore, Mirko Lai, Daniela Paolotti, and Giancarlo Ruffo. 2020. Immigration as a Divisive Topic: Clusters and Content Diffusion in the Italian Twitter Debate. *Future Internet*, 12(10):173.

Waseem, Zeerak and Dirk Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on Twitter. In *Proceedings of the NAACL student research workshop*, pages 88–93, San Diego, California, US, June. Association for Computational Linguistics (ACL).

Weber, Christopher R. and Christopher M. Federico. 2013. Moral foundations and heterogeneity in ideological preferences. *Political Psychology*, 34(1):107–126.

Winterich, Karen Page, Yinlong Zhang, and Vikas Mittal. 2012. How political identity and charity positioning increase donations: Insights from moral foundations theory. *International Journal of Research in Marketing*, 29(4):346–354.