An Investigation of Hate Speech in Italian

Use, Identification, and Perception

Edited by
Silvio Cruschina
&
Chiara Gianollo

HUP HELSINKI UNIVERSITY PRESS

An Investigation of Hate Speech in Italian

Use, Identification, and Perception

Edited by

Silvio Cruschina
&
Chiara Gianollo

HUP HELSINKI UNIVERSITY PRESS

Published by
Helsinki University Press
www.hup.fi

© the authors 2024

First published in 2024

Cover design by Ville Karppanen
Cover image by Istock / Jamakosy
Print and digital versions typeset by Jukka Lauhalahti.

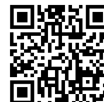
ISBN (Paperback): 978-952-369-115-5
ISBN (PDF): 978-952-369-116-2
ISBN (EPUB): 978-952-369-117-9

https://doi.org/10.33134/HUP-27

This work is licensed under the Creative Commons Attribution-Non-Commercial 4.0 International (CC BY-NC 4.0) License (unless stated otherwise within the content of the work). To view a copy of this license, visit https://creativecommons.org/licenses/by-nc/4.0/. This license allows sharing and copying any part of the work, providing author attribution is clearly stated. Under this license, the user of the material must indicate if they have modified the material and retain an indication of previous modifications. This license prohibits commercial use of the material.
The full text of this book has been peer reviewed to ensure high academic standards. For full review policies, see http://www.hup.fi/

Suggested citation:
Cruschina, Silvio, and Chiara Gianollo, eds. 2024. *An Investigation of Hate Speech in Italian: Use, Identification, and Perception*.
Helsinki: Helsinki University Press. https://doi.org/10.33134/HUP-27

To read the free, open access version of this book with your mobile device, scan this QR code:



Contents

Figures and Tables v
Contributors ix

I INTRODUCTION: CURRENT LINGUISTIC APPROACHES TO HATE SPEECH

Chapter 1: Words that matter: The use of language with hate purposes 1
Silvio Cruschina and Chiara Gianollo

Chapter 2: Call me by my name: Hate speech and identity 27
Claudia Bianchi

II THE INTERACTION OF LEXICAL STRATEGIES AND CONTEXT

Chapter 3: A typology of insults: A corpus-based study of Italian political debates on Twitter 57
Margarita Borreguero Zuloaga

Chapter 4: The words of hate speech: A lexical study of homotransphobia in an Italian Twitter corpus 111
Elena Sofia Safina

Chapter 5: Homophobic hate speech in Italian tweets: Contextual cues of offensiveness 159
Stefano De Pascale, Edoardo Cavirani and Stefania Marzo

Chapter 6: The impact of bilingualism on hate speech perception and slur appropriation: An initial study of Italian UK residents **193**
Mattia Zingaretti, Maria Garraffa and Antonella Sorace

III GRAMMATICAL AND PRAGMATIC STRATEGIES

Chapter 7: Decoding implicit hate speech: Italian political discourse on social media during the COVID-19 pandemic **233**
Mattia Retta

Chapter 8: A linguistic analysis of nationality-based hate speech on Facebook: The case of the Italian language **281**
Alicja Paleta and Anna Dyda

Chapter 9: Hate speech in non-cooperative contexts: Question types as a measure of cooperativity **315**
Valentina Bianchi and Silvio Cruschina

Index 347

Figures

Figure 3.1: Personal pronouns in Italian politicians' tweets 75
Figure 3.2: Injury vs defamation in politicians' tweets 78
Figure 3.3: Preliminary quantitative analysis of insulting mechanisms in our corpus 85
Figure 3.4: A quantitative analysis of insulting strategies in our corpus 86
Figure 3.5: Percentages of rhetorical figures as insulting strategies 100
Figure 4.1: Distribution of hate speech and intensity labels in TWEER 120
Figure 4.2: Distribution of aggressiveness and offensiveness labels in TWEER 120
Figure 4.3: Distribution of irony and stereotype labels in TWEER 121
Figure 4.4: Cluster of hate words in the 'distance' conceptual space 142
Figure 5.1: Qualitative classification of the near-synonyms, in terms of vulgarity (VULG.), regionality (REG.), archaism (ARC.), literary use (LIT.), and neutrality (NEUTR.) 168
Figure 5.2: Absolute frequencies of near-synonyms for the concept HOMOSEXUAL MAN 169
Figure 5.3: Average scores of offensiveness of five near-synonyms for homosexual uses in neutral or expressive contexts 171
Figure 5.4: Proportions of 'in-concept' (blue) vs 'out-of-concept' (orange) tokens, ordered in decreasing order of 'in-concept' tokens 173
Figure 5.5: Conditional inference tree plot of the impact of the features on the distribution of the nine most frequent near-synonyms 180
Figure 5.6: Conditional inference tree plot of the impact of the features on the distribution of the three most frequent near-synonyms 182

Figure 5.7: Conditional inference tree plot with offensiveness as a response variable 184
Figure 6.1: Scatter plots with regression lines and shaded standard errors, showing how length of residence predicts familiarity in Italian (a), personal use in Italian (b), imageability in Italian (c) and offensiveness in English (d) 217
Figure 6.2: Effects plots with 95 per cent confidence interval showing how sexuality significantly predicts the likelihood of perceiving slur appropriation as appropriate 220
Figure 6.3: Effects plots with 95 per cent confidence interval showing how (a) sexuality and (b) ethnicity significantly predict the likelihood of slur appropriation in response to hate speech 221
Figure 9.1: Partition of the context set 320

Tables

Table 3.1: Italian politicians and political parties represented in the corpus 72
Table 3.2: Number of tweets per political party in our corpus 72
Table 3.3: Examples of Italian slurs 79
Table 3.4: Examples of Italian insulting epithets 82
Table 4.1: Most frequent words in the TWEER subcorpus 125
Table 4.2: List of words in the Politics cluster 130
Table 4.3: List of words in the Nature cluster 134
Table 4.4: List of words in the Values and Customs cluster 136
Table 4.5: List of words in the Affectivity and Sexuality cluster 139
Table 4.6: List of offensive words 139
Table 4.7: Biphobic examples on Facebook page Gay.it 150
Table 5.1: Near-synonyms of the concept HOMOSEXUAL MAN 166
Table 6.1: Italian–English word pairs rated by the participants of the study 209
Table 6.2: Participants' background information variables with either Mean (M) and Standard Deviation (SD) or percentage (%) for each variable 211
Table 6.3: Student's t-tests and Wilcoxon signed-ranks tests for word pair ratings 213

Table 6.4: Correlation matrix for all variables in Italian (I) and English (E) with Pearson's r (r) and Spearman's rho (ρ) values 215
Table 7.1: Analytical schema of the study material according to DHA 241
Table 7.2: Discursive strategies 242
Table 7.3: The politicians and the political context of the corpus 245
Table 9.1: Hate speech in Corpus A 331
Table 9.2: Question types in Corpus A 332
Table 9.3: Hate speech in Corpus B 334
Table 9.4: Question types in Corpus B 334
Table 9.5: Hate speech in Corpus C 337
Table 9.6: Question types in Corpus C 337
Table 9.7: Summary of results for all three corpora 340
Table 9.8: Comparison of the three corpora: cooperative vs rhetorical questions 340
Table 9.9: Comparison of the three corpora: answered canonical questions 341

Contributors

Claudia Bianchi (PhD from École Polytechnique, Paris) is Full Professor of Philosophy of Language in the Philosophy Faculty at the University Vita-Salute San Raffaele, Milan, where she is currently the Coordinator of the PhD in Philosophy, the Director of the research centre Gender (Interfaculty centre for gender studies) and of the master's in Communication of Science and Health. Her main research interests are in philosophy of language, pragmatics, and feminist philosophy of language. In particular, she works on hate speech, slurs, and discursive injustice. She has published four books, edited a further one, and co-edited three more. Her latest authored work is *Hate speech: il lato oscuro del linguaggio* (2021).
https://orcid.org/0000-0001-8799-8755

Valentina Bianchi has been Professor of Linguistics at the University of Siena since 2007. She was previously a researcher at the Scuola Normale Superiore in Pisa, where she obtained her PhD in 1995. Her research, initially focussed on formal syntax, was gradually extended to consider the interface between syntax, semantics-pragmatics, and prosody. Her current research is mainly concerned with the analysis of discourse particles and questions from an interface perspective. Additionally, she participates in interdisciplinary work on hate speech and on neo-populist discourse on social media.
https://orcid.org/0000-0002-4441-063X

Margarita Borreguero Zuloaga studied Spanish and Italian Philology and Philosophy at the Complutense University of Madrid, where she is now Associate Professor in the Romance Studies department. She was awarded a Humboldt Fellowship at the University of Heidelberg (2010–2012) and has been a visiting scholar at the universities of Tübingen, Marburg, Bergamo, and Turin. Her research interests are contrastive linguistics and discourse analysis, with particular attention to information structure, discourse markers, anaphors, and discursive strategies in information manipulation. Among her publications are the books *La interfaz lengua texto: un modelo de estructura informativa* (2015) in cooperation with Angela Ferrari and L*as teorías textuales de János S. Petöfi: de la gramática del texto a la Textología Semiótica* (2024). She has been co-editor-in-chief of the journal *Cuadernos de Filología Italiana* (2016–2023) and has coordinated the book series *Sprache – Gesellschaft – Geschichte* (Peter Lang) since 2014.
https://orcid.org/0000-0003-1992-0766

Edoardo Cavirani studied Historical and Comparative Linguistics at the University of Pisa. In 2015, he obtained his PhD (University of Pisa and University of Leiden) with the thesis 'Modeling Phonologization: Vowel Epenthesis and Reduction in Lunigiana Dialects'. From 2014 to 2017, he held a postdoctoral researcher position at the Meertens Institute (KNAW) on the Maps & Grammar project, where he stayed until 2018 as a guest researcher. In 2019 he joined KU Leuven, Belgium, as a FWO Marie Skłodowska-Curie Actions – Seal of Excellence postdoctoral researcher with a project on the microvariation of gender and number spell-out. Throughout his career, he has mostly worked on phonological theory and microvariation, and in recent years has started working more on morphophonology and on morphosyntax, focusing on the formalisation of expressivity. He has published papers in peer-reviewed journals such as *Journal of Linguistics*, *Linguistic Review*, *Studi e Saggi Linguistici*, and *Glossa*.
https://orcid.org/0000-0001-7184-8226

Silvio Cruschina is Professor of Italian in the Department of Languages at the University of Helsinki. He specialises in Romance linguistics, in particular Italian, Sicilian, and other dialects of Italy. His area of expertise is syntax, but his research interests range over a variety of subjects and topics, including information structure, pragmatics, and diachronic variation. Together with Chiara Gianollo, he coordinated the project 'ALIHAS—A Linguistic Investigation of Hate Speech', funded by the university alliance UNA Europa (2020–2022), from which this volume originates. He is the author of *Discourse-Related Features and Functional Projections* (2012) and, together with Delia Bentley and Francesco Maria Ciconte, of *Existentials and Locatives in Romance Dialects of Italy* (2015). He has also co-edited several volumes and journal special issues, and has published in international peer-reviewed journals such as *Glossa*, *Natural Language & Linguistic Theory*, *Isogloss*, *Word Structure*, and *Probus*.
https://orcid.org/0000-0002-8082-8232

Stefano De Pascale obtained his PhD in Linguistics in 2019 at KU Leuven, Belgium. In his thesis he investigated the contribution of token-based semantic vector space models in the study of lexical variation. After working as a postdoctoral scholar on the project 'Nephological Semantics—Using token clouds for meaning detection in variationist linguistics', he obtained a junior FWO-postdoctoral fellowship in 2021 on diachronic semantics and started as a part-time assistant professor of Italian linguistics at Vrije Universiteit Brussel. His research is at the intersection of sociolinguistics, cognitive linguistics, and corpus linguistics, with a strong focus on computational methodology. He is the co-author of *Lexical Variation and Change: A Distributional Semantic Approach* (2023, with Dirk Geeraerts, Dirk Speelman, Kris Heylen, Mariana Montes, Karlien Franco, and Michael Lang) and 'Lexical coherence in contemporary Italian: a lectometric analysis' (2023, with Stefania Marzo).
https://orcid.org/0000-0003-2455-9004

Anna Dyda is an academic researcher at the Institute of Italian Philology at Jagiellonian University, Poland. She holds a PhD in Italian Linguistics (2019). In 2021 she published a monograph entitled *Leggibilità e comprensibilità del linguaggio medico attraverso i testi dei foglietti illustrativi in italiano e in polacco*. She has been a member of research teams in several scientific projects, including 'Romance culture in Poland—Jagiellonian Library Manuscripts'; ALIHAS (A Linguistic Investigation of Hate Speech: How to identify it and how to avoid it)—UNA Europa; and 'Romance languages for Slavic-speaking university students' under the Erasmus+ programme. She is the author and co-author of numerous scientific articles.
https://orcid.org/0000-0002-5116-7586

Maria Garraffa is an Associate Professor of Clinical Linguistics at the University of East Anglia, UK. Her research sits at the interface between medicine and linguistics. She specialises in profiling the language of individuals with language disorders using linguistically informed frameworks. Maria has a particular interest in public health initiatives for speakers of minoritised languages. Her research has received support from various funders, including the Academy of Medical Science and the British Council. She has published over 80 articles in international peer-reviewed journals, employing diverse methodologies, single-case and group studies, and linguistic and cognitive screenings. She has collaborated with international partners on studies involving several clinical populations, including stroke patients with aphasia, adults with ADHD, young offenders, children with developmental language disorder, and individuals with neurodegenerative disorders. Additionally, she has authored an introductory book on the benefits of bilingualism and numerous resources for teaching clinical linguistics in the health sciences.
https://orcid.org/0000-0003-1767-424X

Chiara Gianollo is Associate Professor of Linguistics at the University of Bologna. She is a historical linguist, with expertise in the

diachrony of Greek, Latin, and Old Romance, and has a special interest in formal approaches to historical syntax and semantics. Her research on the dimensions of linguistic variation extends to the study of multilingualism in migration settings, especially with respect to educational contexts. With Silvio Cruschina she coordinated the project 'ALIHAS—A Linguistic Investigation of Hate Speech', funded by the university alliance UNA Europa (2020–2022), from which this volume originates. She is the author of *Indefinites between Latin and Romance* (2018). She co-edited, with Agnes Jäger and Doris Penka, *Language Change at the Syntax-Semantics Interface* (2015); with Ilaria Fiorentini and Nicola Grandi, *La classe plurilingue* (2020); with Maria Napoli and Klaus von Heusinger, *Determiners and Quantifiers: Functions, Variation, and Change* (2022).
https://orcid.org/0000-0001-6380-6368

Stefania Marzo is Professor of Italian Linguistics and head of the Italian Department at KU Leuven, Belgium. She received her PhD in Linguistics from the University of Leuven in 2006. Prior to her appointment at KU Leuven, she worked at the Department of Translation, Interpreting and Communication at the University of Ghent, Belgium. Her research interests lie mainly in the fields of variationist sociolinguistics and contact linguistics. She focuses on the diffusion of urban vernaculars in Flanders, the evolution of Italian as a heritage language in Europe, and the dynamics of (re) standardisation in contemporary Italian. Methodologically, she combines corpus-based research with experimental methods to investigate the social meaning of language variation. Her work has been published in peer-reviewed journals, including *Language in Society*, *Italian Journal of Linguistics*, and *International Journal of Bilingualism*. She has also edited several international volumes.
http://orcid.org/0000-0001-5543-0882

Alicja Paleta is a researcher in Italian language and linguistics. Currently she is working in the Department of Romance Studies of Jagiellonian University, Poland, where she is the faculty advisor

for Italian Language and Culture. She teaches Italian language and translation between Italian and Polish, and is responsible for the training of future teachers of Italian as a foreign language. Her scientific interests concern the didactics of foreign languages (especially in the context of Italian as a foreign language taught to Polish students). She is an author of papers on Italian linguistics and didactics, including 'Multimediale e interattivo…: un'esigenza del metodo o del mercato?' (Multimedia and interactive activities: are they required by the method or by the market?), 'La formula di saluto "salve" nell'insegnamento della lingua italiana come lingua straniera (livello di competenza A1)' (The form of greeting 'salve' in teaching Italian as a foreign language at reference level A1).
https://orcid.org/0000-0002-5146-6244

Mattia Retta is a PhD researcher in Language Studies at the University of Helsinki. His research interests include Italian linguistics, language use in social media, pragmatics, critical discourse analysis, and language pedagogy. His publications include 'A pragmatic and discourse analysis of hate words on social media' (2023).
https://orcid.org/0000-0001-5160-5163

Elena Sofia Safina is a PhD student in the doctoral programme in Mind, Gender, and Language at the University of Naples Federico II, Italy. Her research interests include language, gender and sexuality, inspired by her MA studies, during which she produced a computational linguistics thesis on countering homotransphobia on Twitter. She participated in the UNA Europa seed funding project 'ALIHAS—A Linguistic Investigation of Hate Speech', analysing homotransphobic hate speech both lexically and semantically. Her current research focuses on psycholinguistics, particularly the effects of grammatical gender marking on mental representations of human referents in reading tasks. Additionally, she investigates gender-sensitive language strategies in Italian, analysing their practical usage through corpus-based studies and exploring speaker opinions on gender-sensitive language as future language policy via questionnaires. She published her experimental results

in the paper 'Effects of grammatical gender on gender inferences: experimental evidence from Italian common gender nouns' in the *International Journal of Linguistics* (2024).
https://orcid.org/0009-0000-3199-371X

Antonella Sorace is Professor of Developmental Linguistics at the University of Edinburgh and Honorary Professor at University College London. She is internationally known for her interdisciplinary research on bilingualism across the lifespan and for her contribution to language typology, especially for her work on constrained variation at the syntax–pragmatics interface and gradience in natural language. She is also committed to building bridges between research and people in different sectors of society. She is the founding director of the public engagement centre Bilingualism Matters, which currently has 34 branches across four continents.
https://orcid.org/0000-0002-8812-1979

Mattia Zingaretti is Lecturer in Applied Linguistics and TESOL at York St John University, UK. He completed his PhD at the University of Edinburgh in 2022, funded by the Arts and Humanities Research Council, investigating first language attrition in adult second language learners through behavioural psycholinguistic methods. As an active member of the Centre for Language and Social Justice Research at York St John, and Research and Knowledge Exchange Lead at York St John's Community Language School, his research focuses on second-language acquisition and teaching, and the impact of bi-/multilingualism from linguistic, cognitive, and emotional perspectives. His work has been presented at national and international conferences, as well as having been published in both academic and non-academic venues. He has also disseminated his research and promoted evidence-based insights on bi-/multilingualism and language learning through public engagement as an affiliate member of Bilingualism Matters in Edinburgh.
https://orcid.org/0000-0002-5565-2538

I

Introduction:
Current linguistic approaches to hate speech

CHAPTER 1

Words that matter

The use of language with hate purposes

Silvio Cruschina
University of Helsinki

Chiara Gianollo
University of Bologna

Abstract

This chapter introduces the rationale behind the volume and the main topics discussed in the various chapters. It surveys the difficulties surrounding the definition of hate speech and singles out the main issues that are relevant for its linguistic investigation: besides the lexical elements (slurs, insults, derogatory epithets), more hidden pragmatic and grammatical strategies are also argued to characterise hate speech and aggressive language. In this respect, a rigorous evaluation of the contextual conditions by means of the tools provided by linguistics helps towards establishing a more precise identification of types of hate speech in

How to cite this book chapter:
Cruschina, Silvio, and Chiara Gianollo. 2024. 'Words that matter: The use of language with hate purposes'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 1–25. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-1.

conversational dynamics (explicit and implicit hate speech, intensity and degree of offensiveness, intentions, and effects).

Keywords: hate speech, aggressive language, context-dependency, implicit meaning, speaker intentions

1.1 Introduction

Language is a key element in the construction and reinforcement of social identities, and, as a consequence, also in the creation and diffusion of stereotypes, discrimination, and social injustices. The use of language to attack an individual or group based on attributes such as race, ethnicity, religion, gender, nationality, political ideology, disability, or sexual orientation constitutes the basis of what is now known as *hate speech*. Hate speech is certainly not a new phenomenon. However, it is only since the 2000s that continuously evolving and fast-paced digital communication, combined with the amplifying function of social media, has made it a major topic of research in a variety of areas, including discourse analysis, psychology, sociology, philosophy of language, media, political, legal, and gender studies (see, among others, Leets 2002; Sternberg 2004; Van Blarcum 2005; Parekh 2006; Lillian 2007; Daniels 2008; Maitra and McGowan 2009; Bleich 2011; Herz and Molnar 2012; Waldron 2012; Foxman and Wolf 2013; Gagliardone et al. 2015; Gelber and McNamara 2016; Brown 2017; Richardson 2018; Knoblock 2022; Ermida 2023a; Guillén-Nieto 2023).

Social media channels have come to play an increasingly large role in our everyday lives and communication, creating novel discursive practices and technological affordances (see the recent overview in Esposito and KhosraviNik 2024). They provide a context in which people across the world can communicate, share knowledge, exchange messages, and interact with each other, irrespective of the distance and the social differences between them, thus allowing greater freedom of expression and empowering individual voices. At the same time, however, social media channels also enable anti-social behaviour, cyberbullying, online aggression, and hate speech. These manifestations are intensified

by the virtual nature of the interaction, which tends to remove socially imposed inhibitions, and can be particularly harmful because of the more persistent nature of the message, which is in written form and has the potential to reach a wide audience.

In this new culture of communication, which has been adopted and adapted in other contexts such as politics and other forms of public speech, appeals to emotions and personal beliefs are important persuasive devices. Verbal aggressions, offensive propaganda, and the construction of authority and subordination both in speech and in writing have become frequent features and instruments of communication in multiple spheres of society.

In the research on hate speech, attention has been focused on two main aspects: i) the individuals or groups who have overtly or covertly been victims of aggression or discrimination through hate speech, and ii) the legal and ethical controversy around the boundaries between the right to freedom of speech and the use of hate speech. This volume contributes to the investigation of hate speech by adopting the methodological and theoretical tools of linguistics.[1]

The focus of this volume is on the use and perception of hate speech, which can be produced either by lexical means (e.g. insults, derogatory terms or epithets) or via more subtle grammatical and pragmatic strategies related to implicit meanings or atypical conversational dynamics. The aim of this investigation is to identify the common linguistic characteristics and features of hate speech in different domains of communication and to establish a set of

1 The contributions in this volume stem from the work and collaboration of a research network sponsored and funded by Una Europa, an alliance of European universities (UNA Europa seed-funding scheme, funding number: SF2019003). This network brought together academic researchers to investigate the topic of hate speech in Italian within a project entitled 'A Linguistic Investigation of Hate Speech (ALIHAS): How to Identify It and How to Avoid It'. The preliminary results were presented in a digital workshop organised by the University of Helsinki on 17 May 2021. The papers were later developed and presented in a second workshop that took place at the University of Bologna on 12 November 2021.

criteria that can help distinguish between hate speech and freedom of expression. The studies collected in this volume all focus on Italian, with the aim of collecting data and drawing generalisations starting from relatively homogeneous conditions and allowing for immediate comparability. However, the analyses, methodologies, and findings of the individual chapters can easily be extended to other languages for comparative and contrastive purposes.

The emphasis is on the linguistic strategies and tools that are typically employed for hate purposes, and on the context and the communication situation that foster hate speech. The investigation is not limited to the more obvious and easier to recognise lexical elements of aggressive language and hate speech (e.g. slurs, derogatory epithets, metaphorical offences and insults; see Faloppa 2004, 2012; Croom 2013; Bianchi 2014, 2015; Capone 2014; Bolinger 2017; Cepollaro 2015, 2017; Retta 2023); the use of more hidden pragmatic and grammatical strategies, and the concomitant properties of the contexts in which hate speech proliferates will also be explored. Most of the studies contained in this collection address hate speech on social media, exploiting the potential of these communication channels as an invaluable source of linguistic data that would otherwise be difficult to collect and analyse in a systematic way. The focus on social media, however, should not obscure the fact that online discourse is inextricably connected to the offline settings in which hate speech and aggressive language emerge as a product of the same social context.

1.2 Definition and identification

The first problem to be addressed in an investigation of hate speech is how to define it. Hate speech is a concept that is intuitively easy to grasp but difficult to rigorously define. Indeed, its definition both as a scientific and as a legal notion raises many complex questions ranging across several disciplines: what is hate speech? Who is the target of hate speech? What are the boundaries between hate speech and freedom of speech? (See, among many others, Assimakopoulos 2020; Baider, Millar, and Assimakopoulos

2020; Faloppa 2020: Ch. 1; Anderson and Barnes 2022; Ermida 2023a, Määttä 2023).

The need for a precise and binding definition of hate speech is felt particularly strongly by legislative institutions and international organisations looking to recommend specific measures or legislations against hate speech. In 1997 the European Council provided the governments of its member states with a set of principles and recommendations to combat hate speech, arriving at the following definition:

> the term 'hate speech' shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin. (Council of Europe 1997)

This definition makes reference to the purposes and targets of hate speech but remains quite *open* in its scope of application, using the verb 'including' before the list of potential cases of hate speech. Moreover, it puts a strong emphasis on hate speech motivated by racism and xenophobia. In this respect, the concise definition provided by the United Nations is at the same time more precise and more comprehensive:

> the term hate speech is understood as any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor. This is often rooted in, and generates intolerance and hatred and, in certain contexts, can be demeaning and divisive. (United Nations 2019)

This definition identifies both the operational means of hate speech and its targets—that is, it clarifies the type of speech ('any kind of communication in speech, writing or behaviour, that

attacks or uses pejorative or discriminatory language') and the type of target ('a person or a group on the basis of their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor'), expanding the scope to include all factors that contribute to personal or group identity, which can be targeted in a discriminatory way. According to this definition, hate speech is therefore not limited to extreme verbal aggressions characterised by abusive and offensive language (see Section 1.3 of this chapter), but includes all forms of communication that lead to discrimination or contempt.

Bianchi (2021) reaches a similar definition from the perspective of philosophy of language:

> quelle espressioni e quelle frasi che comunicano derisione, disprezzo e ostilità verso gruppi sociali, e verso individui in virtù della loro mera appartenenza a un certo gruppo; le categorie che sono bersaglio o target dei discorsi d'odio vengono anche in questo caso identificate sulla base di caratteristiche sociali (reali o percepite) come etnia, nazionalità, religione, genere, orientamento sessuale (dis)abilità, e così via. (Bianchi 2021: 5)
>
> (those expressions and sentences that communicate ridicule, contempt and hostility towards social groups, and towards individuals simply because they belong to a certain group; in this case too, the categories that are the target of hatred discourse are identified on the basis of social characteristics (real or perceived) like ethnicity, nationality, religion, gender, sexual orientation, (dis) ability, and so on.)[2]

Two important aspects emerge from this definition. First, there can be different degrees of intensity of hate speech, building a spectrum from ridicule to outright hostility, with different consequences for the communication dynamics and for the legal sanctioning of discourses of hatred (see Section 1.3 below for examples of implicit and covert forms of hate speech). Secondly, the

2 Our translation.

characteristics upon which discrimination is built can have an objective counterpart in reality, or can be represented by perceived social constructs. Moreover, it is clear from the typical social characteristics involved in producing hate speech that discriminatory discourses often capitalise on intersectionality—that is, the intertwining and overlapping of dimensions of identity.

From the definitions above, we clearly see that offensive communication can only be considered as hate speech if it is directed at groups or single individuals (usually as representatives of a group) and does not include verbal attacks on political institutions, public offices, or leaders. In fact, the target of hate speech is typically understood as a minority group—or an individual taken as representative of a minority group—which has been subject to persecution and discrimination. For instance, if a politician verbally attacks another politician with excessive and offensive language, this can hardly be viewed as hate speech because it does not give rise to the creation and diffusion of stereotypes, discrimination, and social prejudices against a (minority) group. From the legislative viewpoint, possible offences in political debate are regulated by specific laws against insults and defamation, which relate solely to the individual. By contrast, if a politician directs violent or offensive language towards a person or a group, for instance on the basis of their nationality (as immigrants) or religion, this *does* count as hate speech, because it reinforces stereotypes and incites or justifies discrimination.

On the basis of these considerations, we distinguish hate speech from more general aggressive language—that is, offensive, violent, and excessive language which is not directed at the groups or single individuals who can often be the targets of hate speech. The two notions are nevertheless not unrelated: since aggressive language can easily turn to hate speech as soon as the target of the abusive communication changes, understanding the linguistic properties of aggressive language may be pivotal to the study of hate speech (on this distinction, see also Ermida 2023b).

It is important to note, moreover, that the definitions provided by the European Council and the United Nations do not have

legal validity, insofar as they do not include strict guidelines for distinguishing hate speech from free speech. This issue is particularly relevant to legislators who can only promulgate and apply laws against hate speech if it is defined as a coherent and clearly demarcated concept. In fact, since the perception and the characteristics of the groups listed in definitions such as that provided by the United Nations can be historically and culturally conditioned, it is not surprising that the laws regarding hate speech, when they exist, vary significantly across countries (Fish 1994; Butler 1997; Perry 2001; Brown 2017; Määttä 2020).

Another argument that is often brought up as an obstacle to the definition of hate speech and to its study as a coherent concept is the role of context. We know that the interpretation of linguistic expressions depends on the context; for some scholars, therefore, hate speech should also be examined in the light of the contextual conditions that characterise the situation of communication, such as the (offensive vs non-offensive) intentions of the speaker, the relationship between the interlocutors, and the use of sarcasm. According to these scholars, in the presence of certain conditions the context would *not* make an utterance hate speech. Because of these complexities, several scholars have even rejected the idea that hate speech constitutes a coherent concept that can be investigated as such (see Boromisza-Habashi 2021).

In approaching the aims of the research project presented in this volume, we acknowledge the difficulties in establishing an unambiguous and universal definition of hate speech, both in theoretical and in applied terms; at the same time, we would like to emphasise two undeniable aspects of hate speech that call for an improved understanding of its linguistic determinants.

First of all, although hate speech is also difficult to define *a priori* from a linguistic viewpoint, the identification and perception of hate speech is much easier. As speakers of a language, we have clear intuitions regarding the presence of hate speech in verbal—or even non-verbal—communication, and we are also able to perceive different levels of intensity. From a scientific viewpoint, and independently of reaching a precise definition, the investigation

of hate speech is therefore crucial in understanding where such speech comes from, how it is generated, what its triggers are, and what contributes to its intensity.

Secondly, hate speech can be investigated independently of the context, on the basis of its content and its intrinsic properties. However, even if we admit that hate speech is strongly context-dependent, this should not be used as a reason to reject the systematic study of hate speech. On the contrary, linguists have the theoretical means to analyse and examine the contextual domain in which hate speech is produced, distinguishing between triggering and concomitant features, and between descriptive and performative functions of language (see, e.g., Bianchi 2018, 2021), and identifying the environments that legitimate the use of certain slurs, as in the case of metalinguistic uses or reappropriation (see Galinsky et al. 2003 and Brontsema 2004 on the reappropriation of terms such as 'bitch' and 'queer'; see also chapters 2 and 6 in this volume). In sum, the contextual variability of hate speech can also be subject to a principled account, allowing for a more nuanced understanding, with important empirical consequences.

The goal of the chapters in this volume is to contribute to the establishment of reliable criteria for the identification of the common linguistic characteristics of hate speech in different contexts of communication. In turn, this could be seen as a first essential step towards the far more complex issue of a scientific definition of hate speech and its practical application at the level of regulation and legislation.

1.3 Explicit and implicit hate speech

From a linguistic perspective it is generally recognised that there are two forms of hate speech (Gao, Kuppersmith, and Huang 2017; Caselli et al. 2020; Faloppa 2020; Brambilla and Crestani 2021): explicit and implicit. The explicit manifestations of hate speech are easier to identify, in that they are typically represented by lexical expressions that contain insults, derogative terms or epithets, threats, or overt references to stereotypes. With these words we

can offend, insult, discriminate, and impose our purported superiority or authority. The intensity of the offensiveness associated with these lexical items or expressions is commonly recognised and can also be measured (see chapters 4 and 5 in this volume).

However, stereotypes, prejudices, discrimination, intolerance, and hatred more generally can also be reinforced implicitly, in more subtle and hidden ways (see Ben-David and Fernández 2016; Bhat and Klein 2020; also Baider 2019, 2023, where the term 'covert hate speech' is used). These implicit manifestations of hate speech are more complex and difficult to detect and to investigate, in particular for natural language processing, which is an important regulatory resource used by social media (see Schmidt and Wiegand 2017; Fortuna and Nunes 2018; Chapter 5 in this volume).

If I utter 'You fucking faggot!', I am using a lexical device to attack, explicitly and directly, a specific target. I might argue that I was just joking or that my intention was just mockery or some kind of irony. This argument built on non-offensive intentions, however, cannot be used as an excuse or as an attempt to *normalise* hate speech. Words have their effect independently of what the speakers claim about their intentions: words can abuse others by ridiculing, hurting, and humiliating them, and are therefore powerful tools of oppression and aggression. Indeed, by uttering the sentence above, I am not simply expressing my opinion or an evaluation, I am producing a speech act that conveys offence, aggression, and denigration, not only towards an individual but also towards a group of people who can be identified on the basis of a specific sexual identity or orientation.[3]

3 On the context of the utterance as an essential element to understand the illocutionary dimension of the speech act, see Meza, Vincze, and Mogoș (2018), Baider (2020) and Culpeper (2021). On sarcasm and humour as forms of implicit or covert hate speech, see Bansal et al. (2020), Frenda et al. (2022). Note that, in this introductory chapter, we roughly use the term 'implicit hate speech' as a synonym of 'covert hate speech', but implicit hate speech could in fact be understood as limited to the subtype of covert hate speech that is associated with implicit meanings such as presuppositions and implicatures (see Chapter 7 in this volume).

But what happens if I use demeaning and divisive communication while carefully avoiding slurs or derogatory terms (for instance, in order not to incur sanctions)? An example could be political slogans that refer to nations or nationality such as 'Britain first!' or 'Prima gli italiani!' (Italians first). In fact, when uttering such expressions, superficially I am not attacking or insulting anybody. In reality, however, these are still messages of hate, which incite intolerance and prejudice towards other social groups by distinguishing and contrasting groups on the basis of nationality. Even though the social groups who are discriminated against are not explicitly mentioned, their identity can be gathered from the context, in the same way as it can be inferred that these *other* groups are to be placed in a position of subordination or inferiority with respect to the group that is prominently singled out in the slogan.

Implicit hate speech can be associated with various speech acts, beyond insults, and with a range of implicit meanings, such as presuppositions and implicatures, which may contribute to the construction or reinforcement of stereotypes (see Lombardi Vallauri 2019a, 2019b; see also Chapter 7 in this volume). The discussion of political slogans like 'Britain first!' has already provided an example of the use of items that trigger often unwarranted presuppositions, such as the existence of a ranking: the use of 'first' presupposes that there is going to be a second, third, and so on—that is, a hierarchy of social groups is necessarily superimposed in the interpretation.

As for implicatures, consider the examples in (1):

(1) a. Some Italians pay taxes.

b. Finns are honest, but Italians cook well.

Both examples are cases in which, in normal communication, the interpretation of the utterance is crucially enriched by implicit meaning, which is triggered in a very systematic way. In (1a) the use of the quantifier 'some' triggers an implicit comparison with a stronger quantifier, 'all', which the speaker could have used but

chose not to. From this scale of strength among possible alternatives of expression, the hearer will extract the non-literal meaning that not all Italians pay taxes. Note that, strictly logically, the utterance in (1a) is compatible with a scenario in which all Italians pay taxes, and does not deny it explicitly. However, well-studied mechanisms of human communication lead the hearer to interpret (1a) by enriching its literal meaning with a so-called scalar implicature ('if some, then not all'), with the effect of contributing to the stereotype that many Italians are tax evaders. This mechanism thus has the potential to generate implicit offensive speech.

Similarly, the conventional implicature evoked by the contrastive connective 'but' in (1b) conveys the implicit meaning that the opposite of what is predicated of one nationality holds for the other, although this is not stated literally. Finns and Italians are compared and a different quality is attributed to each nationality; the fact that a contrast is established by means of 'but' generates the implicit meaning that it cannot be said that Italian are honest, hence they are not, and it cannot be said that Finns cook well, hence they do not.

While the implicit meaning generated by conversational dynamics does not go beyond reinforcing and propagating stereotypes in the examples in (1), the effects of these strategies, in the appropriate context, can lead to hate speech proper. Consider:

(2) Calano fatturato e Pil ma aumentano i gay.

'Sales volume and GDP go down but the number of gays increases.'

The sentence in (2) is the headline of a front-page article published on 23 January 2019 in the right-wing Italian newspaper *Libero*. The headline brings together Italian economic trends in sales and gross domestic product, and the number of people who identify as gay. At first sight, the two facts strike the reader as completely unrelated. However, similarly to (1b), the contrastive connective *ma* (but), by contrasting the two facts, establishes a connection: it suggests that there is a relationship between the two trends, but

leaves it to the reader to figure out what that relationship is. The only way to make sense of the contrast in (2) is to extract from the two conjuncts a value judgement: the decrease in the economic trends is bad; 'but' leads to the expectation that the contrasted fact is, instead, good. This, in turn, forces the reader to recognise irony in the message: clearly, the author does not see the increase in the numbers of gay people as a positive thing that can counterbalance the negative trend. In fact, the opposite is true: what the author wants to imply is that it is a further sign of a decaying society. This means that not only is the result a message of hate, but the author of that message can hide behind its implicit nature, which emerges only because of the reader's pragmatic inferences.

The construction of stereotypes or other forms of generalised prejudices can also be expressed through various grammatical means used for referential purposes: for instance, the contrastive use of personal pronouns ('*we* Italians', '*you* immigrants'), or the use of generic pronouns to refer to an unspecified group (or sub-group) of individuals (e.g. '*they* are lazy and never work'), or the derogative use of demonstratives ('*this* immigrant', '*these* people') (see Fumagalli 2019; Chapter 8 in this volume). The following is an example from a Facebook post, discussed in Fumagalli (2019, 66), where repeated reference is made to *loro* (they) in opposition to *noi* (we) in the context of racist comments:

(3) Ma quale odio razziale … l'odio razziale è da parte loro verso noi. Sono loro che rifiutano i nostri usi e costumi. Sono loro che rifiutano di integrarsi. Sono ancora loro che pretendono, pretendono, pretendono, senza un minimo di riconoscenza.

'What racial hate … racial hate is by them towards us. It is they who reject our customs and traditions. It is they who reject integration. It is yet they who demand, demand, demand, without any gratitude whatsoever.'

The construction of authority and subordination (see Bianchi 2017; Langton 2018) may also arise from linguistic devices such as

the use of prohibition and obligation modals (e.g. 'foreigners must/cannot…'), the exclusive adverb 'only' (e.g. 'only Italians can…'), inclusive pronouns to refer to unspecified groups or parties (e.g. 'we all know this', 'we must act now'), high-degree/-intensity adverbs and adjectives (e.g. 'entire(ly)', 'total(ly)', 'absolute(ly)'), and evidential and epistemic adverbs and adjectives (e.g. 'evident(ly)', 'obvious(ly)', 'undoubted(ly)', 'undisputed(ly)'). All these authoritative strategies have the power to legitimise beliefs and behaviours and to influence perception of reality.[4]

The use of tools that express an opinion or an evaluation in a parenthetical form are also relevant here. These 'asides' are apparently not at the centre of the communicative exchange, but this is precisely why they are particularly insidious since they tend to be more passively accepted by the hearer together with the main core of the message (see Lombardi Vallauri 2000). This is true, for instance, of evaluative adverbs and adverbials (e.g. 'fortunately', 'unfortunately', 'unluckily', 'regrettably'; for instance, 'Regrettably, many [of a particular minority group] live in this neighbourhood') and of special evaluative constructions with comparatives (e.g. *Lei è più bella che intelligente*, 'She's more beautiful than intelligent': apparently a compliment, which becomes an offence when addressed to someone who is not characterised by outstanding beauty).

In terms of intensity and degree of offensiveness, implicit hate speech is certainly weaker than explicit hate speech. Consider the examples in (1), for instance, particularly compared to ethnic slurs or offensive terms related to gender or sexual orientation. Many of us would agree that the statements in (1), especially (1b), can be considered only slightly offensive. However, the role of context is fundamental: if sentences similar to (1a) were uttered in a context (say, a country) in which Italians constituted an immigrant minority that has historically been attacked and accused of

4 See also the contributions collected in Knoblock (2022) for the exploitation of morphosyntactic features, such as word formation strategies, to convey offensive messages.

criminal practices and behaviour, they could have stronger repercussions on the audience, even more so if uttered by a public person with some sort of authority.

Because of its intrinsic nature as a hidden message, implicit hate speech can also be subject to different interpretations and perceptions. It would therefore be difficult, if not totally impossible, to subject it to national or international regulations—this might not even be a desideratum from a legislative viewpoint. However, it is precisely because of its invisible and disguised character that implicit hate speech is employed and exploited in propaganda and with persuasive purposes, both in mass media and in political debate. Since the addressees of these messages very often process and accept the implicit meanings automatically and unconsciously, these communicative expedients have the power to justify and legitimise beliefs and behaviours and to influence our perception of reality. A linguistic analysis of both the explicit and implicit strategies involved in hatred discourse can therefore help us to identify hate speech in all its manifestations, including its most subtle and hidden forms and expressions.

The concepts of explicit and implicit hate speech come very close to, but are not interchangeable with, the notion of hard and soft hate speech, as applied for instance in Baider, Assimakopoulos, and Millar (2017) (see also Assimakopoulos 2020). Under hard hate speech we find all those manifestations that are prosecutable by law because they explicitly incite discriminatory hatred. Soft hate speech, instead, is not prosecutable in legal terms, because it does not explicitly manifest incitement to hatred, although it expresses prejudice and intolerance, and, as such, is capable of considerable harmful impact. Here, implicitness and explicitness refer to the intention to incite hatred, and are therefore tied to a differential legal treatment. In our linguistically minded investigation, instead, implicitness and explicitness refer to the linguistic means that are adopted in conveying discriminatory messages. This linguistic notion of explicitness and implicitness can also more generally be applied to aggressive language, and is not univocally linked to a differential legal status. For instance, there can

be cases of implicit hate speech that qualify as hard hate speech under the above definition: imagine a situation in which a term like 'daisy' assumes a derogatory meaning in the discourse of an in-group (see further Section 1.4); in such a situation, an utterance like 'We must eradicate the daisies' becomes an explicit incitement to violence against a whole group, although the target of hatred will only be recognisable within the in-group. The linguistic means are implicit, but within the in-group the intention to incite hatred is explicit.

1.4 Haters and hated, intentions and effects

Since Paul Grice's ([1975] 1989) distinction between two levels of meaning—*what is said* and *what is meant*—speaker intentions have been assumed to play an essential role in the interpretation of a linguistic utterance. Indeed, the pragmatic level of what is meant involves aspects of meaning that are drawn from the context in which the utterance occurred, so the process of interpretation that leads from *what is said* to *what is meant* can be viewed as an inferential process that is based on principled, pragmatic mechanisms, and that also relies on reasoning about speaker intentions.

In the debate on the boundaries between hate speech and freedom of speech, context-dependency—and especially, speaker intentions—are often put forward as a reason to deny the existence of hate speech in particular situations. Comedians may justify potentially offensive expressions as irony, in the same way as politicians may sanction the use of historically and culturally *loaded* words as a way to demonstrate their aversion towards the silencing that is allegedly imposed by the censorship of the *politically correct* or their rights to freely express their opinion and judgements. They may simply claim that they have no intention of verbally attacking or offending any group or individual, so there is no hate speech. According to this view, hate speech is a slippery concept that depends on the context, so potentially offensive hate expressions change their meaning and their impact depending on who uses them and how they are used.

On the one hand, it is true that insults and other derogatory terms are context-dependent. One example of this is the reappropriation of slurs through which a group reclaims words that have been used in an offensive and discriminatory way against that group (see chapters 2 and 6 in this volume and references therein). Another interesting case is represented by so-called *dog-whistles*—that is, apparently neutral terms (e.g. 'inner-city') that assume an ideologically loaded interpretation for an in-group (in the case of 'inner-city', the stereotypical, negatively connotated reference to African American neighbourhoods; see Henderson and McCready 2019). On the other hand, the effects of a word or a linguistic expression are not entirely tied to the speaker's intentions. Independently of the claimed intentions and of the appeal to free speech, in choosing to use a slur instead of a neutral equivalent, the speakers signal—even unconsciously—that they endorse the term, its connotations, and its associations. It might be a joke for the speaker, but it will still vilify the target. Indeed, it has long been acknowledged that language has a performative function that may be independent from speaker intentions: it does not simply describe the world but is also able to effect changes in the world.

The distinction between constative language, which describes the world and can be evaluated in terms of true or false statements, and performative language, which *does things with words*, was first described by the philosopher John L. Austin (1962). For Austin, the performative function of language includes speech acts such as swearing, promising, betting, and officiating a wedding ceremony. Performative language is not completely independent of the context and the situation of the utterance; indeed, Austin identifies a number of felicity conditions that must be met in order for an utterance to be used performatively. However, the effects and consequences that a word or an utterance can have on the interlocutor or on a listener—the so-called perlocutive effects— may or may not coincide with the intentions of the speaker: *'Saying something will often, or even normally, produce certain consequential effects upon the feelings, thoughts, or actions of the audience, or of the speaker, or of other persons: and it may be done with the design,*

intention, or purpose of producing them' (Austin 1962: 101). But, as implied in this quotation ('often', 'normally'), it may also be done without that design, intention, or purpose.

Performativity of language and its perlocutive effects have clearly played a prominent role in the scholarly debates on the pragmatic and social functions of languages (see Culpeper 2021 and references therein). Language is not only the mirror of society in its descriptive function, reflecting hierarchies, social injustices, conflicts, discriminations, classifications, and divisions; in its performative power, it is also able to generate and transform social identities, connections, and practices by creating, reinforcing, or removing classifications, hierarchies, and conflicts. In this sense, language performativity plays a key role in the construction of human and social identities, such as gender (Butler 1990, 1993, 1997; Bianchi 2021; Chapter 2 in this volume).

1.5 The volume and its structure

This volume undertakes an investigation of hate speech that tackles the linguistic strategies adopted by speakers when they employ language as a means of aggression. The authors apply established methods of data collection and analysis to a novel body of evidence, specifically collected for their studies, primarily from social media and other forms of public speech, with the aim of identifying the common linguistic characteristics of aggressive language and hate speech in different domains of communication. Most of the chapters deal directly with hate speech, while some address issues related to aggressive language, either with respect to their differing offensive potential (e.g. Chapter 3, which analyses different types of insults) or in comparison with hate speech (e.g. Chapter 9, which examines the contextual conditions that favour the emergence of excessive language and hate speech).[5]

5 On the distinction between aggressive language and hate speech, see Section 1.2.

The focus on Italian as an empirical testbed by our international group of authors offers both practical and methodological advantages in terms of comparability of data from different domains of communication and of cohesion in the research questions addressed by the various chapters. The resulting picture will pave the way for future research on other languages, with the aim of identifying cross-linguistic and language-specific strategies and constructions. Although our investigation has a primarily linguistic perspective, the authors contribute their expertise and strengths in different linguistic areas to arrive at a multifaceted and interdisciplinary approach, with insights from language variation, dialectology, language and migration, multilingualism, semiotics, language education, bilingualism, language development and cognition, and computational linguistics. Observations from or about other disciplines are also often present, especially philosophy of language, political studies, and sociology.

The volume is divided into two parts, preceded by this introductory chapter written by the editors and by an invited position paper in Chapter 2, 'Call me by my name: hate speech and identity', written by Claudia Bianchi, a leading figure in the international debate on hate speech at the crossroads between linguistics and philosophy of language. The first of the two main parts of the book concerns the interaction of lexical strategies and context in hate speech, while the second part enlarges the scope by including grammatical and pragmatic analyses. Because of the fluidity and interconnectedness of the themes discussed, however, this categorisation and subdivision of the chapters should not be understood in a rigid way. Indeed, some chapters within each part often overlap in their examination of lexical, contextual, and pragmatic strategies, blurring the delineation between sections.

Three contributions in the first part discuss data from social media in two particularly polarising domains: political debate and homotransphobic discourse. Chapter 3 by Borreguero Zuloaga focuses on contextual insults, which require more than lexical items to be performed and rely on the relevance of cultural knowledge and stereotypes. Meanwhile, Chapter 4 by Safina

and Chapter 5 by De Pascale, Cavirani, and Marzo present data from different social media that show the mechanisms by which homotransphobia emerges from lexical and discursive tools, identifying clusters of meanings that, although not necessarily connected to sexual aspects, contribute to the propagation of linguistic aggressions in this domain. A fourth contribution, Chapter 6 by Zingaretti, Garraffa, and Sorace, addresses the dimension of bilingualism in connection with the perception of hate speech, in particular with respect to slur appropriation, by employing a specifically designed questionnaire.

The second part begins with Chapter 7 by Retta, which takes advantage of the new discourse context provided by the COVID-19 pandemic to analyse the emergence of xenophobic and racist discourse in social media, identifying specific argumentative strategies. The study by Paleta and Dyda (Chapter 8) focuses on linguistic aggressions performed by lexical and pragmatic means, especially pronouns, in a Facebook group comprising expatriates. The last chapter in this part, Chapter 9 by Bianchi and Cruschina, investigates the use of questions in online political debate, showing how they can help to distinguish between cooperative and non-cooperative contexts.

References

Anderson, Luvell, and Michael Barnes. 2022. 'Hate speech'. *The Stanford Encyclopedia of Philosophy* (Spring 2022 edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2022/entries/hate-speech/.

Assimakopoulos, Stavros. 2020. 'Incitement to discriminatory hatred, illocution and perlocution'. *Pragmatics and Society* 11 (2): 177–195. https://doi.org/10.1075/ps.18071.ass.

Austin, John Langshaw. 1962. *How to Do Things with Words*. Cambridge, MA: Harvard University Press.

Baider, Fabienne. 2023. 'Accountability issues, online covert hate speech, and the efficacy of counter-speech'. *Politics and Governance* 11 (2): 249–260. https://doi.org/10.17645/pag.v11i2.6465.

Baider, Fabienne. 2020. 'Pragmatics lost? Overview, synthesis and proposition in defining online hate speech'. *Pragmatics and Society* 11 (2): 196–218. https://doi.org/10.1075/ps.20004.bai.

Baider, Fabienne. 2019. 'Le Discours de haine dissimulée: le mépris pour humilier'. *Déviance et Société* 43 (3): 359–387. https://doi.org/10.3917/ds.433.0359.

Baider, Fabienne, Stavros Assimakopoulos, and Sharon Millar. 2017. 'Hate speech in the EU and the C.O.N.T.A.C.T. project'. In *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*, edited by Stavros Assimakopoulos, Fabienne Baider, and Sharon Millar, 1–6. Cham: Springer.

Baider, Fabienne, Sharon Millar, and Stavros Assimakopoulos, eds. 2020. 'Hate speech: definitions, interpretations and practices'. Special issue, *Pragmatics and Society* 11 (2). https://doi.org/10.1075/ps.11.2.

Bansal, Srijan, Vishal Garimella, Ayush Suhane, Jasabanta Patro, and Animesh Mukherjee. 2020. 'Code-switching patterns can be an effective route to improve performance of downstream NLP applications: a case study of humour, sarcasm and hate speech detection'. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 1018–1023. Kerrville, TX: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.96.

Ben-David, Anat, and Ariadna Matamoros Fernández. 2016. 'Hate speech and covert discrimination on social media: monitoring the Facebook pages of extreme-right political parties in Spain'. *International Journal of Communication* 10: 1167–1193.

Bhat, Prashanth, and Ofra Klein. 2020. 'Covert hate speech: white nationalists and dog whistle communication on Twitter'. In *Twitter, the Public Sphere, and the Chaos of Online Deliberation*, edited by Gwen Bouvier and Judith E. Rosenbaum, 151–172. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-41421-4_7.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.

Bianchi, Claudia. 2018. 'Perspectives and slurs'. In *Beyond Semantics and Pragmatics*, edited by Gerhard Preyer, 187–198. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198791492.003.0011.

Bianchi, Claudia. 2017. 'Linguaggio d'odio, autorità e ingiustizia discorsiva'. *Rivista di Estetica* 64: 18–34. https://doi.org/10.4000/estetica.2059.

Bianchi, Claudia. 2015. 'Parole come pietre: atti linguistici e subordinazione'. *Esercizi Filosofici* 10 (2): 115–135.

Bianchi, Claudia. 2014. 'Slurs and appropriation: an echoic account'. *Journal of Pragmatics* 66 (May): 35–44. https://doi.org/10.1016/j.pragma.2014.02.009.

Bleich, Erik. 2011. *The Freedom to be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism*. Oxford: Oxford University Press.

Bolinger, Renée Jorgensen. 2017. 'The pragmatics of slurs'. *Noûs* 51 (3): 439–462. https://doi.org/10.1111/nous.12090.

Boromisza-Habashi, David. 2021. 'There's no such thing as hate speech and it's a good thing, too'. In *Digital Hate: The Global Conjuncture of Extreme Speech*, edited by Sahana Udupa, Iginio Gagliardone, and Peter Hervik, 23–33. Bloomington, IN: Indiana University Press.

Brambilla, Marina, and Valentina Crestani. 2021. '"Der kriminelle …", "I soliti …": parole dell'odio in tedesco e in italiano'. In *Il linguaggio dell'odio: fra memoria e attualità*, edited by Marilisa D'Amico, Marilisa Brambilla, Valentina Crestani, and Nannerel Fiano, 85–105. Milan: Franco Angeli.

Brontsema, Robin. 2004. 'A queer revolution: reconceptualizing the debate over linguistic reclamation'. *Colorado Research in Linguistics*, 17. https://doi.org/10.25810/dky3-zq57.

Brown, Alexander. 2017. 'What is hate speech? Part 1: The myth of hate'. *Law and Philosophy* 36 (4): 419–468. https://doi.org/10.1007/s10982-017-9297-1.

Butler, Judith. 1997. *Excitable Speech: A Politics of the Performative*. London and New York: Routledge.

Butler, Judith. 1993. *Bodies That Matter: On the Discursive Limits of "Sex"*. London and New York: Routledge.

Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. London and New York: Routledge.

Capone, Alessandro. 2014. 'Indirect reports, slurs, and the polyphonic speaker'. *Reti, Saperi, Linguaggi* 2: 301–318.

Caselli, Tommaso, Valerio Basile, Jelena Mitrović, Inga Kartoziya, and Michael Granitzer. 2020. 'I feel offended, don't be abusive! Implicit/explicit messages in offensive and abusive language'. Paper presented at *LREC 2020 – 12th International Conference on Language Resources and Evaluation*, Conference Proceedings, 6193–6202. Marseille: European Language Resources Association.

Cepollaro, Bianca. 2017. 'The Semantics and Pragmatics of Slurs and Thick Terms'. PhD thesis, Paris Sciences et Lettres Research University.

Cepollaro, Bianca. 2015. 'In defense of a presuppositional account of slurs'. *Language Sciences* 52: 36–45. https://doi.org/10.1016/j.langsci.2014.11.004.

Council of Europe. 1997. 'Recommendation No. R (97) 20 of the Committee of Ministers to Member States on "Hate Speech"', 30 October 1997. Accessed 19 July 2024. https://rm.coe.int/1680505d5b.

Croom, Adam. 2013. 'How to do things with slurs: studies in the way of derogatory words'. *Language and Communication* 33: 177–204. https://doi.org/10.1016/j.langcom.2013.03.008.

Culpeper, Jonathan. 2021. 'Impoliteness and hate speech: compare and contrast'. *Journal of Pragmatics* 179: 4–11. https://doi.org/10.1016/j.pragma.2021.04.019.

Daniels, Jessie. 2008. 'Race, civil rights, and hate speech in the digital era'. In *Learning Race and Ethnicity: Youth and Digital Media*, edited by Anna Everett, 129–154. Cambridge, MA: MIT Press.
Ermida, Isabel, ed. 2023a. *Hate Speech in Social Media: Linguistic Approaches*. Cham: Palgrave Macmillan.
Ermida, Isabel. 2023b. 'Distinguishing online hate speech from aggressive speech: a five-factor annotation model'. In Ermida, *Hate Speech in Social Media*, 35–75. https://doi.org/10.1007/978-3-031-38248-2_2.
Esposito, Eleonora and Majid KhosraviNik, eds. 2024. *Discourse in the Digital Age: Social Media, Power, and Society*. London: Routledge.
Faloppa, Federico. 2020. *#ODIO: manuale di resistenza alla violenza delle parole*. Turin: UTET.
Faloppa, Federico. 2012. *Razzisti a parole (per tacer dei fatti)*. Bari: Laterza.
Faloppa, Federico. 2004. *Parole contro: la rappresentazione del 'diverso' nella lingua italiana e nei dialetti*. Milan: Garzanti.
Fish, Stanley. 1994. *There's No Such Thing as Free Speech and It's a Good Thing, Too*. New York: Oxford University Press.
Fortuna, Paula, and Sérgio Nunes. 2018. 'A survey on automatic detection of hate speech in text'. *ACM Computing Surveys* 51 (4): Article no. 85. https://doi.org/10.1145/3232676.
Foxman, Abraham H., and Christopher Wolf. 2013. *Viral Hate: Containing Its Spread on the Internet*. New York: St. Martin's Press.
Frenda, Simona, Alessandra Teresa Cignarella, Valerio Basile, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2022. 'The unbearable hurtfulness of sarcasm'. *Expert Systems with Applications* 193: 116398. https://doi.org/10.1016/j.eswa.2021.116398.
Fumagalli, Corrado. 2019. 'Discorsi d'odio come pratiche ordinarie'. *Biblioteca della libertà* 224: 55–75. https://doi.org/10.23827/BDL_2019_1_3.
Gagliardone, Iginio, Danit Gal, Thiago Alves, and Gabriela Martinez. 2015. *Countering Online Hate Speech*. Paris: UNESCO Publishing.
Galinsky, Adam D., Kurt Hugenberg, Carla Groom, and Galen V. Bodenhausen. 2003. 'The reappropriation of stigmatizing labels: implications for social identity'. In *Identity Issues in Groups*, edited by Jeffrey Polzer, 221–256. Bingley: Emerald Group Publishing Limited. https://doi.org/10.1016/S1534-0856(02)05009-0.
Gao, Lei, Alexis Kuppersmith, and Ruihong Huang. 2017. 'Recognizing explicit and implicit hate speech using a weakly supervised two-path bootstrapping approach'. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, retrieved from *Computation and Language* (cs.CL). https://doi.org/10.48550/arXiv.1710.07394.
Gelber, Katharine, and Luke McNamara. 2016. 'Evidencing the harms of hate speech'. *Social Identities* 22 (3): 324–341. https://doi.org/10.1080/13504630.2015.1128810.

Grice, H. Paul. 1989. 'Logic and conversation'. In *Studies in the Way of Words*, 22–40. Cambridge, MA, and London: Harvard University Press. Originally published in 1975 in *Syntax and Semantics*, vol. 3: *Speech Acts*, edited by Peter Cole and Jerry L. Morgan, 41–48. New York: Academic Press.

Guillén-Nieto, Victoria. 2023. *Hate Speech: Linguistic Perspectives*. Berlin: De Gruyter.

Henderson, Robert, and Elin McCready. 2019. 'Dogwhistles and the at-issue/non-at-issue distinction'. In *Secondary Content: The Semantics and Pragmatics of Side Issues*, edited by Daniel Gutzmann and Katharina Turgay, 222–245. Leiden: Brill. https://doi.org/10.1163/9789004393127_010.

Herz, Michael, and Peter Molnar, eds. 2012. *The Content and Context of Hate Speech: Regulation and Responses*. Cambridge: Cambridge University Press.

Knoblock, Natalia, ed. 2022. *The Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive, and Dehumanizing Discourse*. Cambridge and New York: Cambridge University Press.

Langton, Rae. 2018. 'The authority of hate speech'. In *Oxford Studies in Philosophy of Law*, vol. 3, edited by John Gardner, Leslie Green, and Brian Leiter, 123–152. Oxford: Oxford University Press.

Leets, Laura. 2002. 'Experiencing hate speech: perceptions and responses to anti-semitism and antigay speech'. *Journal of Social Issues* 58 (2): 341–361. https://doi.org/10.1111/1540-4560.00264.

Lillian, Donna L. 2007. 'A thorn by any other name: sexist discourse as hate speech'. D*iscourse and Society* 18 (6): 719–740.

Lombardi Vallauri, Edoardo. 2019a. 'Implicitation and power of choice'. *Rivista italiana di filosofia del linguaggio* 13 (2): 73–93.

Lombardi Vallauri, Edoardo. 2019b. *La lingua disonesta: contenuti impliciti e strategie di persuasione*. Bologna: Il Mulino.

Lombardi Vallauri, Edoardo. 2000. *Grammatica funzionale delle avverbiali italiane*. Rome: Carocci.

Määttä, Simo. 2023. 'Linguistic and discursive properties of hate speech and speech facilitating the expression of hatred: evidence from Finnish and French online discussion boards'. *Internet Pragmatics*, online-first article. https://doi.org/10.1075/ip.00094.maa.

Määttä, Simo. 2020. 'Syrjivän kielenkäytön rajoilla'. *Niin&Näin* 3: 71–77.

Maitra, Ishani, and Mary Kate McGowan. 2009. 'On racist hate speech and the scope of a free speech principle'. *Canadian Journal of Law and Jurisprudence* 23 (2): 343–372. https://doi.org/10.1017/S0841820900004963.

Meza, Radu, Hanna-Orsolya Vincze, and Andreea Mogoş. 2018. 'Targets of online hate speech in context: a comparative digital social science analysis of comments on public Facebook pages from Romania and Hungary'. *Intersections. East European Journal of Society and Politics* 4: 26–50. https://doi.org/10.17356/ieejsp.v4i4.503.

Parekh, Bhikhu. 2006. 'Hate speech: is there a case for banning?' *Public Policy Research* 12 (4): 213–223. https://doi.org/10.1111/j.1070-3535.2005.00405.x.

Perry, Barbara. 2001. *In the Name of Hate: Understanding Hate Crimes*. London and New York: Routledge.

Retta, Mattia. 2023. 'A pragmatic and discourse analysis of hate words on social media'. *Internet Pragmatics* 6 (2): 197–218. https://doi.org/10.1075/ip.00096.ret.

Richardson, Louise. 2018. 'Woman-hating: on misogyny, sexism, and hate speech'. *Hypatia: A Journal of Feminist Philosophy* 33 (2): 256–272. https://doi.org/10.1111/hypa.12398.

Schmidt, Anna, and Michael Wiegand. 2017. 'A survey on hate speech detection using natural language processing'. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. Valencia: Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-1101.

Sternberg, Robert J. 2004. *The Psychology of Hate*. Washington, DC: American Psychological Association.

United Nations. 2019. 'United Nations strategy and plan of action on hate speech'. Accessed 19 July 2024. https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf.

Van Blarcum, Christopher D. 2005. 'Internet hate speech: the European framework and the emerging American haven'. *Washington and Lee Law Review* 62 (2): 781–830.

Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.

CHAPTER 2

Call me by my name

Hate speech and identity

Claudia Bianchi
University Vita-Salute San Raffaele

Abstract

The aim of this chapter is to examine social labels not only as tools of *description* of social identity but also as means of *construction* of our and other people's identities. I will endorse an Austinian, *performative* perspective on social labels, and focus on a particularly hateful kind of labels, namely slurs. Rather than analysing what slurs mean or *say*, I will devote my attention to what speakers *do* with slurs—and to the different kinds of speech acts that they allow speakers to perform. Firstly, I will characterise how standard, derogatory uses of slurs contribute to shaping toxic and harmful identities for both their targets and their speakers, as well as their non-targeted addressees. Secondly, I will show how appropriated, non-derogatory uses of slurs can help to constitute

How to cite this book chapter:
Bianchi, Claudia. 2024. 'Call me by my name: Hate speech and identity'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 27–54. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-2.

positive identities for targets, setting the boundaries of groups and communities. While slurs reinforce oppressive social norms and hierarchies, and may even legitimate discriminatory actions against targets, appropriation is a way to disrupt such unfair norms and hierarchies, to subvert the subordinate position imposed on targets, and to reclaim strong, positive, proud identities. From this perspective, language is a powerful tool of exclusion, oppression, and discrimination—but, hopefully, also of inclusion, emancipation, and self-determination.

Keywords: slurs, identity, speech acts, appropriation, derogation

2.1 Introduction

In the Pulitzer Prize-winning novel *Middlesex* by Jeffrey Eugenides, 16-year-old Callie Stephanides goes to the New York Public Library and looks up in Webster's Dictionary the word used by her doctors to describe her condition:

> *Hermaphrodite* -1. One having the sex organs and many of the secondary sex characteristics of both male and female… See synonyms at MONSTER.
>
> There it was, *monster*, in black and white, in a battered dictionary in a great city library. A venerable old book, the shape and size of a headstone, with yellowing pages that bore marks of the multitudes who had consulted them before me … Here was a book that contained the collected knowledge of the past while giving evidence of present social conditions … she stared down at that word. *Monster*. Still there. It had not moved. And she wasn't reading this word on the wall of her old bathroom stall … the synonym was official, authoritative; it was the verdict that the culture gave on a person like her. *Monster*. That was what she was. (Eugenides 2002: 430–431)

Words are tools that allow us to communicate with others, to describe objects and individuals inhabiting the world. At the same time, we use words to give a certain order to reality: we assign names to objects and people, and classify them into categories and groups. Most of the time, we have the impression that we are merely reflecting a reality that is given to us; even with regard to social reality, we think of language as a mirror of individuals and groups that exist independently of us. By endorsing an Austinian, *performative* perspective on language (Austin [1962] 1975), I will show that language does not just mirror reality—but *shapes* and transforms social reality, and especially social identities, groups, and hierarchies. That is why people have the right to choose what they wish to be called, either as a group or on an individual basis: they have the right to be called by their name.

The aim of this chapter is to examine social labels not only as devices of *description* but also as means of *construction,* of our and other people's social identities. I will focus on a particular kind of hateful social labels, namely derogatory epithets, or *slurs*. Rather than analysing what slurs mean or *say*, I will turn my attention to what speakers *do* with slurs—to the different kinds of speech acts that they allow speakers to perform. On the one hand, I will characterise how standard, derogatory uses of slurs contribute to shaping toxic and harmful identities for both their targets and their speakers, as well as non-targeted addressees. On the other, I will show how appropriated, non-derogatory uses of slurs help to constitute positive identities for targets and target groups, setting the boundaries of groups and communities.

The chapter is structured as follows. In Section 2.2, I will show how category labels and slurs function as devices of social management, and of social control. In Section 2.3, drawing on Austin's speech acts theory, I will introduce the performative perspective on hate speech and slurs. In Section 2.4, I will provide some examples of how slurs help to construct a strengthened dominant group for the speakers and their addressees, and a weakened group for the targets, hence contributing to 'outgrouping' targets and 'ingrouping' addressees. In Section 2.5, I will present *appropriated*

uses of slurs, namely non-derogatory uses, typically by members of the target group, that are intended to foster camaraderie and to display power and a sense of belonging. In Section 2.6, I will illustrate how, while standard uses of slurs reinforce oppressive social norms and hierarchies and may even legitimate discriminatory actions against targets, appropriation is a way to disrupt such unfair norms and hierarchies, to subvert the subordinate position imposed on targets, and to reclaim positive identities—both for individuals and groups. From a performative perspective, language is a powerful tool of exclusion, oppression, and discrimination—but, hopefully, also one of inclusion, emancipation, and self-determination.

2.2 Social labels and slurs

Words are key devices of social control. We classify people and groups with the help of social labels, which we then use to justify and legitimate our beliefs, emotions, and actions towards individuals and social categories. Labels, indeed, influence our expectations of individuals and our behaviour towards them, while also projecting stereotypes and prejudices. They are a sort of *lens* through which we see and interact with others, and through which we learn to see ourselves. In this sense, social labels are ways in which we control and discipline individuals. For example, once a label such as 'man' or 'woman' has been applied to someone, we expect particular appearances, feelings, attitudes, and behaviours from them; non-conformity with such expectations will be acknowledged, condemned, and sometimes even punished. Part of the function of social labels is to make classifications seem natural, obvious, and rational, and to conceal their contingency and historicity, hence suppressing the need to either justify or criticise such categorisations (on social labels, see inter alia Rothbart and Taylor 1992; Haslam, Rothschild, and Ernst 2000, 2002; Bastian and Haslam 2006; Haslam and Levy 2006; Prentice and Miller 2007; for an overview, see Leslie 2017).

This normative dimension of language is especially evident for forms of expression that fall under the label of 'hate speech'. The definition of this term is highly contentious. Hate speech concerns 'insulting, degrading, defaming, negatively stereotyping or inciting hatred, discrimination or violence against people in virtue of their race, ethnicity, nationality, religion, sexual orientation, disability, gender identity, for example' (Brown 2017, 419–420), and may include a wide variety of expressive forms, ranging from words to gestures, from sounds to images, and from symbols to communicative behaviour. Here I will deal only with an uncontroversial instance of hate speech, that constituted by slurs. Slurs are particular social labels (such as 'dyke' or 'wop') targeting individuals and groups of individuals on the basis of race, nationality, religion, disability, gender, sexual orientation, and so on.

Since the start of the 21st century, scholars working on slurs have identified a number of features that characterise their linguistic behaviour (see for example Hornsby 2001; Hom 2008; Potts 2005, 2007; Richard 2008; Croom 2011, 2013; Camp 2013; Anderson and Lepore 2013a, 2013b; Jeshion 2013a, 2013b; Bianchi 2014a, 2014b, 2021; Bolinger 2017; Nunberg 2018; Cepollaro 2020; for a recent overview, see Hess 2022). Here I will address only those features of slurs that make them powerful devices for the construction of social identities—in both negative and positive ways.

First, slurs convey hatred of and contempt for their targets, displaying unique derogatory force. Indeed, most scholars consider slurs to be more offensive than non-slurring pejoratives (terms like 'stupid' or 'idiot', targeting individuals rather than groups of people). As Robin Jeshion puts it, 'Slurs are widely regarded as extraordinarily pernicious, far more so than many other pejoratives like "jerk" or "idiot"—harming their target's self-conception and self-worth, often in ways that are common to the social group as a whole' (Jeshion 2013b: 314). Indeed, while pejoratives only express the negative attitude of the speaker towards a particular individual, slurs target an entire social category: 'That explains how the impact of a slur can be more explosive and threatening than any expression that merely gives voice to the speaker's point

of view, however charged it is or how emphatically it is uttered' (Nunberg 2018: 286).[1] Slurs denigrate the members of a target group *because* they are members of that group, thereby exemplifying the social aspect of hateful language: unlike insults, which denigrate individuals because of something that they *do*, slurs denigrate individuals because of something that they *are*—their (real or perceived) social traits.

Another feature characterising slurs is that their derogatory force *evolves* over time, reflecting the values and dynamics of the society: expressions that were once neutral (such as 'Negro' or 'Coloured') have become derogatory, while expressions that were once insulting (like 'gay' or 'Tory' and 'Whig') are no longer perceived as offensive. Chris Hom, for example, points out, that

> As target groups gradually integrate into the dominant society, and active discrimination subsides, the derogatory content of the corresponding epithets will typically fade. Examples of gradual decline might include epithets for Irish immigrants such as 'mic' or 'paddy' (for American English), terms that were much more antagonistic one hundred and fifty years ago in the United States. (Hom 2008: 427–428)

1 The intuition that slurs are more offensive than non-slurring pejoratives has recently been experimentally confirmed by Cepollaro, Sulpizio, and Bianchi (2019). Their pilot study showed that, on average, slurs are indeed perceived as more offensive than non-slurring insults, but only when presented in isolation ('wop' versus 'idiot'). In fact, when slurs occur in atomic predications of the form 'Claudia is a wop', they are perceived as less offensive than when they occur in isolation. According to them, a *decrease* in offensiveness in atomic predications could be explained in terms of the *information* provided by slurs. In addition to denigrating and dehumanising, slurs, unlike insults, also function to *describe* the subject, to provide factual information regarding features such as nationality, ethnicity, sexual orientation, and the like. On the contrary, a non-slurring insult like 'asshole' does not provide any specific descriptive information about the subject; it simply expresses a negative attitude.

Additionally, slurs inflict long-lasting *harm* not only on their targets but also on bystanders. Empirical studies show that racial insults and slurs cause physical or psychological damage to targets: such damage ranges from nightmares and post-traumatic stress to hypertension, psychosis, and suicide (Delgado [1982] 1993; D'Augelli 1992; Swim et al. 2001, 2003; Cowan and Mettrick 2002). Slurs also increase the gap between targets and dominant groups, even as far as non-racist group members are concerned. The non-racist members of the dominant group feel relieved not to have to undergo similar abuse, while members of the target group treat even non-racist members of the dominant group with hostility and suspicion (Matsuda [1989] 1993). Moreover, empirical studies by Greenberg and Pyszczynski show that slurs have a detrimental impact not only on targets but also on bystanders: ethnic slurs prompt *negative evaluations* of the target group by those who overhear the slur (Greenberg and Pyszczynski 1985; Kirkland, Greenberg, and Pyszczynski 1987).

More recent findings go even further. Experimental studies by Carnaghi and collaborators investigate the effects of homophobic slurs on the self-perception of heterosexual males, showing that when exposed to homophobic slurs they are motivated to underline their masculinity and claim a distinctly heterosexual identity by distancing themselves from homosexuals: 'derogatory language not only activates prevalently negative images about gays but also triggers identity-protective strategies in heterosexual males, thereby creating an even stronger gap between heterosexuals and homosexuals' (Carnaghi, Maass, and Fasoli 2011: 1663; see also Fasoli, Maass, and Carnaghi 2015).

2.3 How to do things with slurs[2]

We said in Section 2.1 that words are devices not only of *description* but also of *construction* of social reality. This is in line with John Austin's performative perspective on language, which focuses not

2 I borrow the title of this paragraph from Croom (2013).

on what words *say*, but on what speakers *do* with them. According to Austin's speech acts theory, we must distinguish three different acts within the same total speech act—for example, the uttering of a sentence like

(1) Shoot them!

The *locutionary* act is the act of saying something, the act of uttering certain expressions that are well formed from a syntactic point of view and are meaningful. The *illocutionary* act corresponds to the particular force that an utterance like (1) has in a particular context (order, request, entreaty, challenge, and so on): by uttering a sentence we can bring about new facts, undertake obligations and legitimate attitudes and behaviours, institute new conventions, and modify social reality. The *perlocutionary* act corresponds to the effects brought about by performing an illocutionary act, and to its consequences (intentional or non-intentional) on the feelings, thoughts, or actions of the participants.

Although it does not explicitly address relations of power imbalances and inequalities, Austin's analysis provides the theoretical framework to clarify issues of oppression and subordination. Drawing on Austin's work, Rae Langton draws on the speech acts account in order to understand hate speech (Langton, Haslanger, and Anderson 2012). Hateful labels such as slurs are expressions used not only to describe but also to *do* things, to perform certain speech acts: indeed, slurs do not merely *mirror* phenomena of racism, sexism, and homophobia, or *cause* occurrences of racism, sexism, and homophobia, but do themselves *constitute* forms of racism, sexism, and homophobia. In a speech acts framework, we may conceive of acts performed using slurs in three different ways:

i) as *locutionary* acts that *represent* discrimination and oppression;
ii) as *perlocutionary* acts that *cause* discrimination, and produce changes in attitudes and behaviours, including oppression and violence;

iii) as *illocutionary* acts that *constitute* racial or gender discrimination, legitimate beliefs, attitudes, and behaviours of discrimination, and advocate oppression and violence: 'Austin's distinction between illocutionary and perlocutionary acts offers a way to distinguish speech that *constitutes* racial oppression, and speech that *causes* racial oppression' (Langton, Haslanger, and Anderson 2012: 758; see also Bianchi 2014a, 2018, 2021).

Here I will narrow my focus to the illocutionary perspective on slurs, to their distinctive performative and normative power: they are clear examples of how we can evaluate, assault, harm, and even subordinate individuals with words. Following Catharine MacKinnon (1987: 202), and drawing on her previous work on pornography as a form of hate speech (Langton [1993] 2009: 35), Langton identifies three distinctive kinds of illocutionary acts: a) subordination; b) assault; c) propaganda (Langton, Haslanger, and Anderson 2012: 758).

a) *Subordination*. The first class of illocutions that a speaker can perform by using hate speech and slurs includes authoritative acts of subordination. While insults hurt people by communicating one person's dislike, displeasure, or disapproval of another, slurs inflict harm—they do something, they have normative power: in addition to changing beliefs about their targets, they subordinate their targets. Slurs are connected with networks of subordination and help to enact wide-ranging systems of oppression or more local policies, as in

(2) Fagots stay out!

—the infamous (misspelt) sign installed in the 1940s at Barney's Beanery and displayed there for decades. Acts of subordination such as (2) are directed at both target and non-target addressees: with slurs we *classify* people as inferior, *legitimate* racial, religious, or gender discrimination, and *deprive* minorities of powers and

rights.[3] Quill Kukla (writing as Rebecca Kukla) holds a view similar to Langton's:

> slurs exercise power by positioning the interpellator above the one interpellated on some sort of hierarchy, at least locally. I can insult someone as an equal ('Wow, you're being an asshole!') but I can't slur someone as an equal; the use of the slurring name not only *reflects* but *constitutes* a kind of subordinating speech, which positions the one slurred in a less empowered position than the one using the slur. (Kukla 2018: 20–21; see also Maitra 2012; McGowan 2012; Nunberg 2018)

b) *Assault*. A second class of illocutions that a speaker can perform by using a slur includes assault-like speech acts such as *persecuting* and *degrading*. Assault-like speech acts are typically (but not exclusively) performed with *second-person* uses of slurs (see Jeshion 2013a, 2013b), as in

(3) Wop!

By using slurs, speakers may directly attack, persecute, or degrade their targets. Slurs are weapons of verbal abuse: the focus is on the targeted group and individuals. By uttering (3), the speaker is not merely *asserting* something, but is performing an illocutionary act of persecuting, degrading, or threatening—an act directed towards both a particular individual and all Italians.

c) *Propaganda*. The third class of illocutions that a speaker can perform by using a slur includes propaganda-like speech acts such

---

3 Following Austin's taxonomy, Langton classifies authoritative subordinating speech acts as verdictives or exercitives. In the class of verdictives Austin includes acts (formal or informal, and concerning facts or values) of giving a verdict, estimate, or appraisal (such as acquitting, reckoning, assessing, diagnosing). In the class of exercitives Austin includes acts of exerting powers, rights, or influence (such as appointing, voting, ordering, warning). In Langton's view, slurs are used to classify people as inferior (verdictives) and to legitimate racial oppression, religious or gender discrimination, and to deprive minorities of powers and rights (exercitives).

as *inciting* and *promoting* racial or gender discrimination, hate, and violence. They are typically (but not exclusively) performed with *third-person* uses of slurs, as in

(4) Claudia is a wop.

Shifting the focus from targets to addressees, the speaker's utterance of (4) may be regarded as an act of propaganda, an act that incites and promotes discrimination: the act of propaganda is primarily addressed to 'prospective haters' (Langton, Haslanger, and Anderson 2012: 758). Some uses of slurs, in other words, come with an invitation to assume a certain perspective—they shape the interlocutors' responses and guide their thinking—but also allow speakers to claim an *affiliation* with a particular group, their beliefs and their attitudes, their discriminatory and sometimes even violent behaviours.[4]

When we use a slur third-personally with the members of our own group against a target group, we aim to create or reinforce both the target's subordinated identity and our own dominant identity. At the same time, we attempt to shape the identity of our *addressees*: we present them not only as being outside the target group and inside the dominant group, but also as willing to share our derogatory stance against the target group. As Kukla points out:

> Slurring others *together* is a special kind of speech act that enforces and constitutes ingroup boundaries and memberships. It powerfully positions not just the one uttering the slur, but also the audience who hears and recognizes the slur within the ideology that gives the slur its primary force and meaning. (Kukla 2018: 22–23)

In the following section, I will present some examples of the complex performative dimension of gender and racial slurs—with which we contribute to establishing group membership and to setting boundaries on acceptable beliefs, attitudes, and behaviours.

4 On affiliation with dominant groups, see Bolinger (2017) and Nunberg (2018).

2.4 In and out

We said earlier that from a performative perspective, slurs are tools of social management: they police beliefs, attitudes, and behaviours, and legitimate discriminatory speech and physical acts. This social control function is particularly evident in slurs targeting women, where it usually takes a highly sexualised form. Indeed, when women are the object of hate speech, the words used to assault them are usually variations of 'slut': the reference to women's sexual conduct is evidence of how strongly our societies monitor women and their sexual behaviour—and also evidence of how women tend to be reduced to their body and their sexuality. Moreover, this kind of slur targets not only women who are not behaving 'properly' in the sexual domain, but more generally women who do not conform to gender norms and expectations— primarily women who participate in the public sphere.[5] Sexualised and violent messages, death and rape threats have the ultimate goal not only of condemning women's opinions but also of undermining their presence in the public sphere.

Moreover, such *assaultive* words have a sort of *propaganda* boomerang effect on 'good' women who are merely bystanders. As Lynne Tirrell writes:

> Sometimes … a [slur] is used by a member of the dominant group to a hearer who is a member of the subordinate group as a way of labeling the third person with a label that boomerangs from the target back to the hearer. For example, Fred and Ethel see Lucy do something silly, and while Ethel laughs, Fred scornfully says, 'Lucy is such a bimbo.' 'Bimbo' is a gendered term, and its use here sets boundaries on acceptable and unacceptable female behavior … [Fred's] use of the derogatory term sets gender boundaries for

5 Some examples of hate campaign targets, with death and rape threats, are Kamala Harris, vice president of the USA; Laura Boldrini, former president of the Italian parliament; Michela Murgia, Italian author and activist; and Caroline Criado Perez, British author, journalist, and activist who, in fighting for the Women's Room project, aimed to increase the presence of female experts in the media.

> Ethel even though he was hurling the term at Lucy. (Tirrell 2012: 192)

In a similar vein, Kukla points out that a word such as ‘slut’ not only helps constitute the identity of its target as an ‘abject’ woman who is sexually available (‘just a thing that has sex’)—and in this way *outgroups* her—but at the same time contributes to constructing the identity of ‘good’ women ‘who do not desire or take pleasure in sex’—and in this way *ingroups* them. Interestingly, according to Kukla, ‘slut’ carries with it a third identity:

> men who are always ready for sex and will take sex when they can get it—because the concept of a ‘slut’ requires that there be plenty of men available to have sex with them, even though they are abject. Such men are not themselves particularly abject or objectified; they are just acting as men *naturally* do. (Kukla 2018: 27)

Unsurprisingly, slurs may be used to police the appearance and behaviour of members of target groups even by individuals who belong to the same oppressed group. Some slurs—such as ‘Banana’, ‘Oreo’, ‘Apple’, ‘Coconut’ and ‘Bounty Bar’—are meant to criticise what is perceived as ‘racial betrayal’. ‘Banana’ is a word targeting Chinese Americans perceived as having yellow skin and a white heart; ‘Oreo’ targets African Americans perceived as having black skin and a white heart; ‘Apple’ targets Native Americans who are allegedly white on the inside; ‘Coconut’ targets those Desis,[6] Latinos, and Afro-Caribbeans who are perceived as being brown outside but white on the inside; finally, ‘Bounty Bar’, as in the coconut-filled chocolate bar, is a slur targeting Black people in positions of authority in England. Once more, this kind of slur aims to outgroup actual members of the target group who do not behave properly.

6 According to the Cambridge Dictionary, a Desi is a ‘person who comes from or whose family comes from India, Pakistan, or Bangladesh but who lives in another country’ (s.v. ‘desi (*n*)’, last updated 7 May 2021, https://dictionary.cambridge.org/dictionary/english/desi).

The most forceful example of the performative power of slurs is provided by Lynne Tirrell in an article titled 'Genocidal language games'. Tirrell studied the changing speech practices in Rwanda in the years prior to the 1994 genocide of the Tutsi, focusing on the performative force of slurs such as 'snake' or 'cockroach'—deeply derogatory terms licensing extermination and murder (on the role of hate speech during the Rwandan genocide, see inter alia Chretien et al. 1995; Des Forges 1999; Sibomana 1999; Thompson 2007).[7] Racist propaganda was broadcast throughout the country, primarily by the Radio Télévision Libre des Mille Collines (RTLMC). The role of language before and during the genocide was recognised by the UN war crimes tribunal in 2003: the tribunal charged the RTLMC leader, Ferdinand Nahimana, with genocide, incitement to genocide, and crimes against humanity.

According to Tirrell, slurs display the *functional variation feature*—that is, they allow the performance of various speech acts, serving many different functions: enacting power, inciting crime, and rationalising cruelty. Furthermore, slurs reinforce unjust networks of power and help to constitute subordinate identities. In other words, slurs express the *insider/outsider function*: 'Using such terms helps to construct a strengthened "us" for the speakers and a weakened "them" for the targets, thus reinforcing or even realigning social relations' (Tirrell 2012: 174–175). Moreover, the negative message communicated by slurs concerns an allegedly essential (sometimes even biological) aspect of the target, and thereby creates and enforces a hierarchy (*essentialism condition*). Finally, slurs are *action-engendering* within a context: they delineate what kinds of treatments are permissible with respect to those who are classified in this way.[8]

7 The Rwandan genocide occurred between April and July 1994. During this period of around 100 days, members of the Tutsi minority ethnic group, as well as some moderate Hutu, were killed by armed militias. The most widely accepted scholarly estimates range from around 500,000 to 662,000 Tutsi deaths.

8 André Sibomana, a survivor of the massacre, powerfully describes how slurs helped erase the Tutsis' identities as human beings: 'Soon it was

Tirrell remarks that slurs are most effective when they are *connected to networks of oppression* and discrimination, with the weight of history and social censure behind them, but underlines that this connection to practices of subordination need not be conscious or acknowledged by the speakers enacting the practice:

> When a ten-year-old boy in the USA calls one of his classmates 'fag,' he is unlikely … to think about, much less have mastery of, the broader social context of homophobia and hate crimes against homosexuals. Just the same, that child uses a term that brings a heavy social history and oppressive apparatus to bear on his classmate … Although this speaker is a child, many adults speak with similar epistemic limitations, day in and day out. Few of our words lead to genocide, but we must consider our own diction and ask what apparatuses of power we invoke to control or harm others. (Tirrell 2012: 206)[9]

---

not even necessary to encourage the population to kill. Violence feeds on violence, like a fire. People went mad and lost all points of reference. They killed and killed and killed. Or rather, they stopped killing to "work". They weren't crushing skulls with their rifle butts anymore; they were stamping on vermin. The meaning of words changed and language adjusted to this new concept of life which identified different levels in the human species. Tutsi and their Hutu accomplices were really no longer viewed as human beings, but as things, dirt which had to be eliminated, poisonous snakes which had to be destroyed, whatever their age' (Sibomana 1999: 57–58).

9 To give an example of such networks of oppression against homosexuals, according to ILGA, the International Lesbian, Gay, Bisexual, Trans and Intersex Association, a worldwide federation of more than 1,700 organisations from over 160 countries and territories campaigning for lesbian, gay, bisexual, trans, and intersex human rights, 'As of December 2020, 69 States continue to criminalise same-sex consensual activity'. There are currently six UN Member States (Brunei, Iran, Mauritania, Nigeria, Saudi Arabia, and Yemen) that impose the death penalty on consensual same-sex sexual acts. In five additional UN member states (Afghanistan, Pakistan, Qatar, Somalia, and the United Arab Emirates) 'certain sources indicate that the death penalty could potentially be imposed for consensual same-sex conduct, but there is less legal certainty on the matter' (ILGA 2020).

2.5 Non-derogatory uses: appropriation

Slurs are emblematic of social practices of subordination and discrimination. Yet they can sometimes be used in non-derogatory ways. Most scholars agree that in certain contexts the derogatory force of slurs is, to a certain extent, neutralised or at least diminished (for a survey on non-derogatory uses of slurs, see the special issue of the *Grazer Philosophische Studien*, edited by Cepollaro and Zeman 2020).[10] *Reporting* slurs, for example, is generally perceived as less offensive than *using* them (see inter alia Potts 2005; Schlenker 2007; Langton et al. 2012; Anderson and Lepore 2013a, 2013b; Wieland 2013; Anderson 2016; Capone 2016; Bach 2018).[11] In addition, it is a matter of debate whether slurs occurring in fictional contexts (such as novels, films, and songs) maintain their derogatory power. Another interesting example are fictional slurs, namely slurs made up by writers to target fictional groups or individuals, such as robots or vampires.[12] Finally, according to some

10 Nevertheless, some authors defend a prohibitionist view according to which the mere phonetic realisation of slurs triggers a reaction of offence in *any* context. Anderson and Lepore, for example, take a *silentist* stance and suggest removing slurs from use until their offensive potential fades away, and avoiding any use or mention in any context: 'we insist upon *silentism* as policy. A use, mention, or interaction with a slur, *ceteris paribus* … constitutes an infraction … We cringe when confronted by slurs *because* they usually admit of no tolerable uses' (Anderson and Lepore 2013a: 39).

11 Some empirical studies on the perceived offensiveness of slurs and non-slurring insults ('jerk', 'asshole', etc.) in direct and indirect speech found that the speaker who utters a slur in a report is perceived as less offensive than a speaker using an unembedded slurring utterance such as (3), but to some degree offensive nevertheless. Additionally, quotation marks (as in 'Mary said: "Claudia is a wop" ') can *seal* part of the derogatory import of slurs (Cepollaro, Sulpizio, and Bianchi 2019).

12 The web site *Tropedia* (2021) lists a large variety of fictional slurs. To provide just a few examples, in the comic book *Top Ten*, robots are sometimes referred to as 'clickers', a term that carries the same connotations as the N-word (robots prefer to be called 'Ferro-Americans' or 'Post-organics'); in the TV show *Battlestar Galactica* robots are called 'toasters', and in the movie *I Robot*, 'canners' (presumably short for 'can

scholars, slurs can occur in non-derogatory contexts such as *pedagogical contexts*, where the speaker is objecting to discriminatory discourse, as in:

(5) Institutions that treat Chinese people as chinks are racist (Hom 2008: 429).

All these cases are more or less contentious. There are contexts, however, which are unanimously considered as non-derogatory, namely contexts of appropriation (or reclamation). Appropriation of slurs is the phenomenon whereby speakers (typically but not exclusively in-groups) use a slur for non-derogatory purposes, usually to express intimacy and solidarity, and sometimes as an empowering tool of social and political struggle. The best-known examples are the appropriation of 'Black' by the African American community in the 1960s, of 'queer' by the homosexual community in the 1990s, and the more recent appropriation of 'nigga' by the African American community.[13] Such uses are usually (but not always) taken to convey solidarity rather than hatred or contempt, and are often employed to help achieve political goals and fight oppression.[14]

Two broad types of appropriated contexts are usually identified:

opener'). In the *Harry Potter* series, 'Mudblood' is a slur frequently used for Muggle-born wizards, a word implied to be on par with the N-word in terms of nastiness. In the movie *Blade* vampires are usually called 'suckheads', while in the TV show *True Blood* they are called 'fangs'.

13 Appropriation is a well-documented practice in sociolinguistics: there are examples of appropriation of slurs targeting race ('nigga'), gender ('bitch', 'slut'), sexual orientation or gender identity ('gay', 'queer'), ability status ('deaf'), and so on.

14 There is little consensus on the best account of appropriation. Several alternative theories have been proposed: the ambiguity account (Potts 2007; Hom 2008), the echoic account (Bianchi 2014b), the expressivist account (Richard 2008; Jeshion 2013a, 2020), the indexical account (Ritchie 2017), and Anderson's account in terms of communities of practice (Anderson 2018).

a) friendship contexts—where the non-derogatory use has no conscious political, social, or cultural intent: the slur is used as a term of endearment, or to express camaraderie and solidarity as a form of banter and mock impoliteness;[15]
b) political appropriation contexts—where civil rights groups or artists (writers, poets, comedians, song lyricists) reclaim the use of the slur as a tool of deliberate political and social fight. The slur 'queer' has undergone such a process of conscious political appropriation: 'QUEER can be a rough word but it is also a sly and ironic weapon we can steal from the homophobe's hands and use against him' (Anonymous 1990).[16]

Appropriation may be an effective instrument for fighting discrimination, allowing in-groups to demarcate the group, showing a sense of intimacy and solidarity and reminding targets that they are objects of discrimination. In Hom's words, appropriation

> is a means for the targeted group to recapture political power from the racist group by transforming one [of] its tools, it is a means for 'toughening up' other members of the targeted group by desensitizing them to uses of the epithet, it is a means of in-group demarcation to bring members of the targeted group closer together and to remind members of the targeted group that they are, indeed, a targeted group. (Hom 2008: 428)

Through appropriation, targets assume a critical stance against derogatory uses of a slur and attempt to disrupt entrenched sociocultural norms—they do not merely replace or erase offensive

15 On banter and mock impoliteness, see Leech (1983) and Culpeper (1996).

16 See Bianchi (2014b) and Anderson (2018). Jeshion (2020) identifies two types of appropriation: 'pride reclamation' and 'insular reclamation', respectively.

uses, but *subvert* them.[17] An appropriated slur, then, 'is the same word, with the same history, but with a new future' (Tirrell 1999: 60).[18]

Appropriation subverts the perception of both targets and slurs. Experimental studies by Galinsky and colleagues (2003, 2013) show that self-applying a slur results in in-groups feeling more powerful, and being perceived as more powerful by both targets and non-targets. Further empirical works show that appropriation changes the perception of a slur: self-ascribing a slur reduces its perceived negativity (Galinsky et al. 2013).

This reduction of perceived negativity may eventually lead to the *neutralisation* of the slur: certain words appear to have lost their slur status as a result of a process of appropriation. The best-known example is the slur 'queer', mentioned above. At the beginning of the 1990s, the gay community started a process to take control over the term. Gradually, appropriated uses of the slur became widespread, weakening the connection between the word and the oppressive norms governing it. Eventually, they extended to out-groups: academics were the first to start using the term 'queer' in ways licensed by the gay community, with expressions referring to research fields such as Queer Studies and Queer Theory. The term became customary in general culture, and has become a neutral label for gender non-conforming people (see Brontsema 2004).

A similar process of (non-linguistic) appropriation and neutralisation has involved a symbol, the downward-pointing pink

---

17 See Hornsby (2001: 134): 'they trade on the fact of the word's having had its former hateful or contemptuous element. Where words are appropriated for a new use, old non-descriptive meanings are not brushed away: they are subverted'.

18 In a similar vein, Adam Croom emphasises the function of 'normative reversal' of appropriated uses: 'the non-derogatory in-group use of slurs is especially prevalent in communities highly influenced by "counter-culture" norms (i.e., norms adopted in opposition to, and for the purpose of subverting, other entrenched sociocultural norms that a group contests), such as those associated with hip-hop culture' (Croom 2013: 191).

triangle that was placed on the shirts of gay men in Nazi concentration camps—to identify and dehumanise them. In the 1970s and 1990s, activists reclaimed the symbol as one of liberation: the upward-pointing pink triangle has since become a symbol of gay power and pride (see Jensen 2002).[19]

2.6 Conclusion

My aim was to examine slurs as means of construction of our and other people's social identities. Rather than analysing what slurs mean or say, I have endorsed an Austinian, performative perspective, and focused on what speakers *do* with slurs—on the different kinds of speech acts that they allow speakers to perform, in a negative and in a positive way.

On the one hand, I have characterised how standard, derogatory uses of slurs contribute to shaping toxic and harmful identities for both their targets and their speakers, as well as non-targeted addressees. Indeed, derogatory uses of slurs both draw on and reinforce networks of oppressive identities in complex ways, and unfairly set the boundaries of groups and acceptable in-group behaviour.

First, slurs help hate-speakers to *outgroup* targets, and constitute their identities as subordinated subjects, by ranking them

19 There are concerns about the process of appropriation: see for example Herbert (2015), Anderson (2018), and Herbert and Kukla (2016: 594): 'a term undergoing reclamation, when used by the wrong person in the wrong way, can have the opposite effect: when used by an outsider it reverts to being a slur or a pejorative … this is complicated by the fact that part of what is at issue and unsettled in such reclamation projects is often the boundaries of "the" community. There are no strict and stable rules for who counts as the right person or what counts as the right kind of use … This makes the project of repurposing traditionally subordinating, outgrouping speech especially dangerous (Herbert 2015). "Bitch" used skilfully by someone in the right position can be hilarious and empowering; used just an indefinable bit off-key, it can reinforce sexism and be alienating and hurtful'. On 'nigga', see Kennedy (2003) and Rahman (2012).

as having inferior worth, legitimating discriminatory behaviour towards them, and depriving them of powers and rights.

Additionally, they help hate-speakers to constitute their own identities as members of a dominant, powerful, and intimidating group: ‘The [N-word] can turn a bigot from a hapless, inconsequential “I” into an intimidating, menacing “we” ’ (Nunberg 2018: 286). Slurs are devices for displaying both distance from the target group and membership in a dominant group.

Finally, slurs help hate-speakers to *ingroup* non-targeted hearers, and constitute their identities in harmful ways. By using slurs, bigots present their addressees not only as having the ‘right’ identity (the speaker’s own social identity) but also as likely to hold the same derogatory attitude towards the target group. This is why slurs often evoke a feeling of complicity in their hearers. Liz Camp observes that ‘it seems that any standard form of engagement with the slurring utterance threatens to make us complicit in the bigot’s way of thinking, despite our finding it abhorrent’ (Camp 2013: 330). Adam Croom concurs: ‘the racial slur “nigger” is explosively derogatory, enough so that just hearing it mentioned can leave one feeling as if they have been made complicit in a morally atrocious act’ (Croom 2011: 343).

On the other hand, I have shown how appropriated, non-derogatory uses of slurs help to constitute positive identities for targets and target groups, setting the boundaries of groups and communities. Appropriated uses of slurs may derail standard harmful dynamics of identity construction—and actually initiate an opposite, subverted, positive dynamic. When members of a target group use an appropriated slur, they repurpose the word and perform a variety of potentially positive speech acts.

First, appropriated slurs help targets to *display* insider status, and to constitute their own identities as members of a powerful and proud group.

Additionally, appropriated slurs help targets to *recognise* someone else’s insider status, or even to *invite* someone into a group—that is, they help targets to *ingroup* relevant hearers. By using appropriated slurs, targets present their addressees as having the

'right' identity (the speaker's own social identity), or at least the 'right' insider status.[20] In Herbert and Kukla's words:

> This process does not just *reflect* the realities of community membership but also helps to *constitute* it … Part of being an insider *is* being recognized as one. Crucially, the relevant sort of recognition is not mere passive, conscious acknowledgment, but the kind of recognition that is built into practice. (Herbert and Kukla 2016: 584)

Of course, appropriated uses of slurs sometimes display an oppositional nature: they may be used to *set* the boundaries of the group, and to *outgroup* non-targets, by constituting their identities with hostility and suspicion.[21]

From a performative perspective, language is a powerful tool of exclusion, oppression, and discrimination—but, hopefully, also one of inclusion, emancipation, and self-determination. A theoretical comprehensive awareness of such complex dynamics of identity construction will hopefully help to highlight not only the harms of hate speech, but also the outcomes of processes of self-empowering initiated by oppressed individuals or groups. Slurs are usually connected to unfair systems of social power; they help to reinforce oppressive social norms and to license unjust and even violent actions against their targets. Appropriation is indeed a way of destabilising oppressive social norms and systems of this sort: through appropriation, in-groups disrupt and subvert the subordinate position that has been imposed upon them, and claim for themselves a strong, positive identity. A hateful instrument of

20 See Herbert and Kukla (2016: 583): '*displaying* insider status, *inviting* someone into a group, *settling* the boundaries of a group and the norms it shares, *recognizing* someone else's insider status, *closing ranks* against someone and thereby outgrouping them, and so forth'.

21 This goes partially against Herbert and Kukla (2016: 588): 'Peripheral speech can provide tools for building a positive (in the sense of nonoppositional, not necessarily in the sense of evaluatively good) identity, and this is an ethically and theoretically important function that language can serve'.

injustice and subordination is turned against the oppressors, and transformed into an expression of power and pride.

References

Anderson, Luvell. 2018. 'Calling, addressing, and appropriation'. In *Bad Words: Philosophical Perspectives on Slurs*, edited by Ernest Sosa, 6–28. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198758655.003.0002.

Anderson, Luvell. 2016. 'When reporting others backfires'. In *Indirect Reports and Pragmatics. Interdisciplinary Studies*, edited by Alessandro Capone, Ferenc Kiefer, and Franco Lo Piparo, 253-264. Cham: Springer International Publishing.

Anderson, Luvell, and Ernie Lepore. 2013a. 'Slurring words'. *Noûs* 47: 25–48. https://doi.org/10.1111/j.1468-0068.2010.00820.x.

Anderson, Luvell, and Ernie Lepore. 2013b. 'What did you call me? Slurs as prohibited words'. *Analytic Philosophy* 54 (3): 350-363. https://doi.org/10.1111/phib.12023.

Anonymous. 1990. 'Queers read this'. Accessed 27 May 2022. http://www.qrd.org/qrd/misc/text/queers.read.this.

Austin, John L. [1962] 1975. *How to do Things with Words*, 2nd edition, edited by James O. Urmson and Marina Sbisà. Oxford: Oxford University Press.

Bach, Kent. 2018. 'Loaded words: on the semantics and pragmatics of slurs'. In *Bad Words: Philosophical Perspectives on Slurs*, edited by David Sosa, 60–76. Oxford: Oxford University Press.

Bastian, Brock, and Nick Haslam. 2006. 'Psychological essentialism and stereotype endorsement'. *Journal of Experimental Social Psychology* 42 (2): 228–235. https://doi.org/10.1016/j.jesp.2005.03.003.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Rome and Bari: Laterza.

Bianchi, Claudia. 2018. 'Perspectives and slurs'. In *Beyond Semantics and Pragmatics*, edited by Gerhard Preyer, 187–198. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198791492.003.0011.

Bianchi, Claudia. 2014a. 'The speech acts account of derogatory epithets: some critical notes'. In *Liber Amicorum Pascal Engel*, edited by Julien Dutant, Davide Fassio, and Anne Meylan, 465–480. Geneva: Université de Genève.

Bianchi, Claudia. 2014b. 'Slurs and appropriation: an echoic account'. *Journal of Pragmatics* 66 (May): 35–44. https://doi.org/10.1016/j.pragma.2014.02.009.

Bolinger, Renée Jorgensen. 2017. 'The pragmatics of slurs'. *Noûs* 51 (3): 439–462. https://doi.org/10.1111/nous.12090.

Brontsema, Robin. 2004. 'A queer revolution: reconceptualizing the debate over linguistic reclamation'. *Colorado Research in Linguistics* 17: 1–17. https://doi.org/10.25810/dky3-zq57.

Brown, Alexander. 2017. 'What is hate speech? Part 1: The myth of hate'. *Law and Philosophy* 36 (4): 419–468. https://doi.org/10.1007/s10982-017-9297-1.

Camp, Elisabeth. 2013. 'Slurring perspectives'. *Analytic Philosophy* 54 (3): 330–349. https://doi.org/10.1111/phib.12022.

Capone, Alessandro. 2016. *The Pragmatics of Indirect Reports*. Cham: Springer International Publishing.

Carnaghi, Andrea, Anne Maass, and Fabio Fasoli. 2011. 'Enhancing masculinity by slandering homosexuals: the role of homophobic epithets in heterosexual gender identity'. *Personality and Social Psychology Bulletin* 37 (12): 1655–1665. https://doi.org/10.1177/0146167211424167.

Cepollaro, Bianca. 2020. *Slurs and Thick Terms: When Language Encodes Values*. London: Rowman & Littlefield.

Cepollaro, Bianca, Simone Sulpizio, and Claudia Bianchi. 2019. 'How bad is to report a slur? An empirical investigation'. *Journal of Pragmatics*, 146: 32–42. https://doi.org/10.1016/j.pragma.2019.03.012.

Cepollaro, Bianca, and Dan Zeman. 2020. 'Editors' introduction: the challenge from non-derogatory uses of slurs'. In 'Non-derogatory uses of slurs', edited by Bianca Cepollaro and Dan Zeman, special issue, *Grazer Philosophische Studien* 97 (1): 1–10. https://doi.org/10.1163/18756735-09701002.

Chrétien, Jean-Pierre, Jean-François Dupaquier, Marcel Kabanda, and Joseph Ngarambe. 1995. *Rwanda: Les Médias du génocide*. Paris: Karthala.

Cowan, Gloria, and Jon Mettrick. 2002. 'The effects of target variables and setting on perceptions of hate speech'. *Journal of Applied Social Psychology* 32 (2): 277–299. https://doi.org/10.1111/j.1559-1816.2002.tb00213.x.

Croom, Adam. 2013. 'How to do things with slurs: studies in the way of derogatory words'. *Language and Communication* 33: 177–204. https://doi.org/10.1016/j.langcom.2013.03.008.

Croom, Adam. 2011. 'Slurs'. *Language Sciences* 33: 343–358. https://doi.org/10.1016/j.langsci.2010.11.005.

Culpeper, Jonathan. 1996. 'Towards an anatomy of impoliteness'. *Journal of Pragmatics* 25: 349–367. https://doi.org/10.1016/0378-2166(95)00014-3.

D'Augelli, Anthony Raymond. 1992. 'Lesbian and gay male undergraduates' experiences of harassment and fear on campus'. *Journal of Interpersonal Violence* 7 (3): 383–395. https://doi.org/10.1177/088626092007003007.

Delgado, Richard. 1993. 'Words that wound: a tort action for racial insults, epithets, and name-calling'. In Matsuda *et al.*, *Words that Wound*,

89–110. Originally published in 1982 in *Harvard Civil Rights-Civil Liberties Law Review* 17: 133–181.

Des Forges, Alison. 1999. *Leave None to Tell the Story: Genocide in Rwanda*. New York: Human Rights Watch.

Eugenides, Jeffrey. 2002. *Middlesex*. New York: Farrar.

Fasoli, Fabio, Anne Maass, and Andrea Carnaghi. 2015. 'Labelling and discrimination: do homophobic epithets undermine fair distribution of resources?' *British Journal of Social Psychology* 54 (2): 383–393. https://doi.org/10.1111/bjso.12090.

Galinsky, Adam D., Kurt Hugenberg, Carl Groom, and Galen V. Bodenhausen. 2003. 'The reappropriation of stigmatizing labels: implications for social identity'. In *Identity Issues in Groups*, edited by Jeffrey Polzer, 221–256. Bingley: Emerald Group Publishing Limited. https://doi.org/10.1016/S1534-0856(02)05009-0.

Galinsky, Adam D., Cinthya S. Wang, Jennifer A. Whitson, Eric M. Anicich, Kurt Hugenberg, and Galen V. Bodenhausen. 2013. 'The reappropriation of stigmatizing labels: the reciprocal relationship between power and self-labeling'. *Psychological Science* 24: 2020–2029. https://doi.org/10.1177/0956797613482943.

Greenberg, Jeff, and Tom Pyszczynski. 1985. 'The effect of an overheard ethnic slur on evaluations of the target: how to spread a social disease'. *Journal of Experimental Social Psychology* 21: 61–72. https://doi.org/10.1016/0022-1031(85)90006-x.

Haslam Nick, Louis Rothschild, and Donald Ernst. 2002. 'Are essentialist beliefs associated with prejudice?' *British Journal of Social Psychology* 41 (1): 87–100. https://doi.org/10.1348/014466602165072.

Haslam, Nick, Louis Rothschild, and Donald Ernst. 2000. 'Essentialist beliefs about social categories'. *British Journal of Social Psychology* 33 (1): 113–127. https://doi.org/10.1348/014466600164363.

Haslam, Nick, and Sheri R. Levy. 2006. 'Essentialist beliefs about homosexuality: structure and implications for prejudice'. *Personality and Social Psychology Bulletin* 32 (4): 471–485. https://doi.org/10.1177/0146167205276516.

Herbert, Cassie. 2015. 'Precarious projects: the performative structure of reclamation'. *Language Sciences* 52, 131–138. https://doi.org/10.1016/j.langsci.2015.05.002.

Herbert, Cassie, and Rebecca Kukla. 2016. 'Ingrouping, outgrouping, and the pragmatics of peripheral speech'. *Journal of the American Philosophical Association* 2 (4): 576–596. https://doi.org/10.1017/apa.2016.31.

Hess, Leopold. 2022. 'Slurs: semantic and pragmatic theories of meaning'. In *The Cambridge Handbook of Philosophy of Language*, edited by Piotr Stalmaszczyk, 450–466. Cambridge: Cambridge University Press.

Hom, Christopher. 2008. 'The semantics of racial epithets'. *Journal of Philosophy* 105 (8): 416–440. https://doi.org/10.5840/jphil2008105834.

Hornsby, Jennifer. 2001. 'Meaning and uselessness: how to think about derogatory words'. In *Midwest Studies in Philosophy*, edited by Peter A. French and Howard K. Wettstein, XXV, 128–141. https://doi.org/10.1111/1475-4975.00042.

ILGA. 2020. 'ILGA World updates state-sponsored homophobia report: "there's progress in times of uncertainty"'. Press release, 15 December 2020. https://ilga.org/news/ilga-world-releases-state-sponsored-homophobia-december-2020-update/.

Jensen, Erik N. 2002. 'The pink triangle and political consciousness: gays, lesbians, and the memory of Nazi persecution'. *Journal of the History of Sexuality* 11: 319–349. https://doi.org/10.1353/sex.2002.0008.

Jeshion, Robin. 2020. 'Pride and prejudiced: on the appropriation of slurs'. *Grazer Philosophische Studien* 97: 106–137. https://doi.org/10.1163/18756735-09701007.

Jeshion, Robin. 2013a. 'Expressivism and the offensiveness of slurs'. *Philosophical Perspectives* 27 (1): 231–259. https://doi.org/10.1111/phpe.12027.

Jeshion, Robin. 2013b. 'Slurs and stereotypes'. *Analytic Philosophy* 54 (3): 314–329. https://doi.org/10.1111/phib.12021.

Kennedy, Randall. 2003. *Nigger: The Strange Career of a Troublesome Word*. New York: Vintage.

Kirkland, Shari, Jeff Greenberg, and Tom Pyszczynski. 1987. 'Further evidence of the deleterious effects of overheard derogatory ethnic labels: derogation beyond the target'. *Personality and Social Psychology Bulletin* 13: 216–227. https://doi.org/10.1177/0146167287132007.

Kukla, Quill Rebecca. 2018. 'Slurs, interpellation, and ideology'. *Southern Journal of Philosophy* 56: 7–32, Spindel Supplement. https://doi.org/10.1111/sjp.12298.

Langton, Rae. 2009. 'Speech acts and unspeakable acts'. In *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*, 25–63. Oxford: Oxford University Press. Originally published in 1993 in *Philosophy and Public Affairs* 22 (4, Autumn): 293–330.

Langton, Rae, Sally Haslanger, and Luvell Anderson. 2012. 'Language and race'. In *The Routledge Companion to the Philosophy of Language*, edited by Gillian Russell and Delia Graff Fara, 753–766. New York: Routledge.

Leech, G. 1983. *Principles of Pragmatics*. London: Longman.

Leslie, Sarah-Jane. 2017. 'The original sin of cognition: fear, prejudice and generalization'. *Journal of Philosophy* 114 (8): 393–421. https://doi.org/10.5840/jphil2017114828.

MacKinnon, Catharine. 1987. *Feminism Unmodified: Discourses on Life and Law*. Cambridge, MA: Harvard University Press.

Maitra, Ishani. 2012. 'Subordinating speech'. In Maitra and McGowan, *Speech and Harm*, 94–120.

McGowan, Mary Kate. 2012. 'On "whites-only" signs and racist hate speech: verbal acts of racial discrimination'. In Maitra and McGowan, *Speech and Harm*, 121–147.

Maitra, Ishani, and Mary Kate McGowan, eds. 2012. *Speech and Harm: Controversies over Free Speech*. Oxford: Oxford University Press.

Matsuda, Mari. 1993. 'Public response to racist speech: considering the victim's story'. In Matsuda *et al.*, *Words that Wound*, 17–51. Originally published in 1989 in *Michigan Law Review* 87 (8, August): 2320–2381. https://doi.org/10.2307/1289306.

Matsuda, Mari, Charles Lawrence III, Richard Delgado, and Kimberlè Crenshaw. 1993. *Words that Wound: Critical Race Theory, Assaultive Speech and the First Amendment*. Boulder, CO: Westview Press.

Nunberg, Geoffrey. 2018. 'The social life of slurs'. In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 237–295. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0010.

Prentice, Deborah A., and Dale T. Miller. 2007. 'Psychological essentialism of human categories'. *Current Directions in Psychological Science* 16 (4): 202–206. https://doi.org/10.1111/j.1467-8721.2007.00504.x.

Potts, Christopher. 2007. 'The centrality of expressive indices'. *Theoretical Linguistics* 33 (2): 255–268. https://doi.org/10.1515/tl.2007.019.

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Rahman, Jacquelyn. 2012. 'The N word: its history and use in the African American community'. *Journal of English Linguistics* 40 (2): 137–171. https://doi.org/10.1177/0075424211414807.

Richard, Mark. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.

Ritchie, Katherine. 2017. 'Social identity, indexicality, and the appropriation of slurs'. *Croatian Journal of Philosophy* 17 (2): 155–180.

Rothbart, Myron, and Marjorie Taylor. 1992. 'Category labels and social reality: do we view social categories as natural kinds?' In *Language, Interaction and Social Cognition*, edited by Gun Semin and Klaus Fiedler, 11–36. London: Sage.

Schlenker, Philippe. 2007. 'Expressive presuppositions'. *Theoretical Linguistics* 33 (2): 237–245. https://doi.org/10.1515/TL.2007.017.

Sibomana, André. 1999. *Hope for Rwanda*, *Conversations with Laure Guilbert and Hervé Deguine*. London: Pluto Press.

Swim, Janet, Lauri Hyers, Laurie Cohen, and Melissa Ferguson. 2001. 'Everyday sexism: evidence for its incidence, nature, and psychological impact from three daily diary studies'. *Journal of Social Issues* 57: 31–53. https://doi.org/10.1111/0022-4537.00200.

Swim, Janet K., Lauri L. Hyers, Laurie L. Cohen, Davita C. Fitzgerald, and Wayne H. Bylsma. 2003. 'African American college students'

experiences with everyday racism: characteristics of and responses to these incidents'. *Journal of Black Psychology* 29 (1): 38–67. https://doi.org/10.1177/0095798402239228.

Thompson, Allan, ed. 2007. *The Media and the Rwanda Genocide*. London and Ann Arbor, MI: Pluto Press.

Tirrell, Lynne. 2012. 'Genocidal language games'. In Maitra and McGowan, *Speech and Harm*, 174–221.

Tirrell, Lynne. 1999. 'Derogatory terms: racism, sexism and the inferential role theory of meaning'. In *Language and Liberation: Feminism, Philosophy and Language*, edited by Kelly Oliver and Christina Hendricks, 41–79. Albany, NY: SUNY Press.

Tropedia. 2021. 'Fantastic slurs'. Last updated 28 June 2021. https://tropedia.fandom.com/wiki/Fantastic_Slurs.

Wieland, Nellie. 2013. 'Indirect reports and pragmatics'. In *Perspectives on Pragmatics and Philosophy*, edited by Alessandro Capone, Franco Lo Piparo, and Marco Carapezza, 389-411. Dordrecht: Springer.

II

The interaction of lexical strategies and context

CHAPTER 3

A typology of insults

A corpus-based study of Italian political debates on Twitter

Margarita Borreguero Zuloaga
Complutense University of Madrid

Abstract

Political discourse has undergone a radical change in recent decades due both to a new conception of politics as entertainment for citizens and to the use of social networks as the primary site of political debate and interaction, among other factors. One of the main linguistic characteristics of this new political discourse is the presence of linguistic elements that fulfil the pragmatic function of insulting opponents. Our study aims to establish a typology of insulting strategies in political discourse based on an analysis of a corpus of tweets by Italian politicians. Our point of departure is an encompassing notion of insult that considers its illocutionary traits and perlocutionary effects. This notion overcomes the

How to cite this book chapter:
Borreguero Zuloaga, Margarita. 2024. 'A typology of insults: A corpus-based study of Italian political debates on Twitter'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 57–109. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-3.

concept of insult as epithet (such as slurs and other negatively connotated adjectives) and offers a broader perspective on textual constructions where the negatively connotated lexical elements are nouns or verbs, or where the rhetorical devices are key in fulfilling the insulting function. Therefore, three types of insults will be examined here: slurs or derogatory epithets, other insulting epithets, and rhetorical insults.

Keywords: hate speech, insult, political discourse, social networks, Italian

3.1 Political discourse and social networks

Italian political discourse has undergone a radical change since the 1990s (see Mazzoleni 1998; Dell'Anna 2009, 2010; Scaramella 2016). It no longer involves careful argumentative constructions that employ precise lexical and syntactic structures with the aim of persuading the opponent in the tradition of Ancient Greek and Roman rhetorical discourse. Today political discourse has evolved to become closer to everyday language, and particularly to spoken language in daily interactions (Gallardo 2018, 2022). This has been called the 'mirroring paradigm' and means that domain-specific vocabulary and formal register have been abandoned and characteristics of spoken language such as vague terms, impromptu speech, and anacoluthon are often found in political statements both inside and outside parliament (Antonelli 2017: 21–23, 48–50, 54–63). Debates are not often organised as an interchange of arguments and counterarguments because political discourse is no longer primarily argumentative but narrative (Antonelli 2017: 4–5; Gallardo 2022: 61–72). Carefully planned arguments have been substituted by *argumenta ad hominem*—that is, spontaneous attacks on opponents, which are rarely based on facts. Some politicians, such as Silvio Berlusconi, Umberto Bossi, and Beppe Grillo, seem to have played a key role in this process in the Italian scene (Antonelli 2017: 21–35). In fact, spontaneity and improvisation can be observed in political interviews, statements to the press, and—overall—in interactions in mass media.

Regarding computer-mediated communication (social networks, microblogging), studies on the level of legibility of political tweets applying the Gulpease index and the type–token ratio show the high degree of legibility of these texts, which are oversimplified in language (Antonelli 2017: 48–49; Combei 2020). There has hence been a levelling between the speech of politicians and that of the average citizen, which has deprived political discourse of its former aura. This can be considered a strategy to reach different types of voters, especially those with a lower level of education, and to distance the new parties from traditional political parties, which are seen as part of the political elite.

It has been said that mass media have reshaped political discourse, in that they have transformed politics into just another form of entertainment (Antonelli 2017: Ch. 4). The focus has moved away from the problems of civil society that politicians are faced with and towards scandals, rumours, and trivial anecdotes. Media attention is mainly devoted to what politicians say on TV or on social media instead of what they propose in the traditional loci of power such as the parliament or even the press.

One of the clearest changes is the relocation of the pathos dimension in political discourse. According to Aristotle's precepts, pathos has the function of causing the audience to experience emotions in order to predispose them to hear the argumentative part of the discourse—that is, it was a mean of persuasion; in contemporary political discourse, however, emotions have taken the place of arguments and the content of discourse is therefore reduced to the expression of emotions while ideas and facts occupy a marginal position (Spina 2016; Antonelli 2017: 5–7). Although this type of discourse was considered prototypical of populist parties some years ago (Combei 2020),[1] today the

1 A linguistic definition of populism is based on a particular rhetoric and discursive style: the polarisation of the opposition between two groups (*we* vs *they*: we Italians vs they foreigners or immigrants, we the people vs they the political elite; see Paris 2020: 78–80); the role of the implicit meaning to avoid a conscious reception of the message and the possibility of discussing it (Lombardi Vallauri 2019); the use of rhetorical

predominance of emotions in political language is pervasive even in speeches by politicians who belong to more traditional parties (Antonelli 2017: 50–51). However, in the case of Italian politicians it is fair to say that the leaders of the right-wing parties (mainly Matteo Salvini from Lega Nord and Giorgia Meloni from Fratelli d'Italia) are particularly prolific in producing emotional discourse and hate speech.

One way of raising negative emotions in political discourse is by discrediting opponents, attacking or mocking the facts and claims that they present. This discourse strategy has received the metaphorical name of 'flaming' and it is so pervasive that when a group of Italian journalists decided to write a *Manifesto della comunicazione non ostile* (Manifesto of non-hostile communication) they had to declare in point 9 that 'insults are not arguments'. Flaming is fostered by the anonymity and the disembodiment in interactions that take place on social networks (Palermo 2020: 2).

According to Testa (2018), flaming is always successful because 'the mechanism of discrediting never fails. Refuting a discrediting narrative makes it stronger. Presenting a non-discrediting narrative against it legitimates it. Ignoring it [i.e., not reacting to it] underlines the (guilty) helplessness of those who are discredited' (Testa 2018, our translation). This communicative success will explain its dissemination in political discourse and, more specifically, in political discourse in social media.

---

figures such metaphors, metonymy, hyperboles, etc., and the presence of polyphony marked by quotation marks; the strategy of refuting the debate based on argumentative discourses by denying the opponent the right to speak (Petrilli 2019b); the use of colloquialism, slang and a plain communicative style (Combei 2020: 106–107). The speaker is always emotionally implicated in the discourse and the main attitudes are negativism and pessimism, appealing to emotions (linked to patriotism and national unity in the case of right-wing populism) and *intimisation* by referring to personal experiences.

3.1.1 The role of the social networks

Social networks (henceforth SNs) have important advantages for political communication: the possibility of producing and disseminating messages at a massive level; the immediacy that was not possible with traditional mass media; the intertextuality (i.e., the ability to comment on another's words just by reposting a message or a video without having to reproduce their discourse, which leads the audience to believe that there is less manipulation in quoting mechanisms than in the press); the illusion of an interaction with citizens; the de-territoriality (i.e., the possibility of reaching a larger audience who are not necessarily affected in a direct way by the political actions of the speaker, but who may contribute to further dissemination of their messages and who could allow other people, who may be geographically distant, to get to know the speaker) (Spina 2016; Theocaris et al. 2020: 2–3).

The ease of publishing a message on a SN has led to intense posting activity by politicians and their communication advisors, with a rhythm that exceeds just a single daily post. This unceasing bombardment of information, criticisms, mockeries, harsh comments, and so on favours a permanent campaign atmosphere, where differences between election and non-election periods become blurred and it becomes harder to distinguish the discourses produced by government representatives from those issued by the opposition, although members of extreme political partiers (both far right and far left) use a more aggressive vocabulary (Torregrossa et al. 2023: 461).

But it has been the formal limitations of the texts, especially on Twitter,[2] that have determined a new type of political discourse, in which texts are shorter, more emotional, and discrediting. There is no space for well-developed ideas, even less for argumentative texts, so politicians have opted to reduce their messages to two types: a) praising themselves for what they have done or are about

2 Although Elon Musk renamed this SN as 'X' in August 2023, 'Twitter' continues to be the most widespread name among its users.

to do; and b) discrediting others for what they have done or said (Van Dijk 2006).[3] In both cases the aim is to persuade their audience by provoking positive or negative emotions, not by offering a reasoned argument about the qualities or opportunities of the course of action they are praising or criticising.

3.1.2 Twitter and the 'new' political discourse

Twitter, which is more precisely defined as a microblog than as a SN, has attracted the attention of scholars dealing with the 'new' political discourse, not only because it is the most used SN for political propaganda all around the world but also because the texts are easily accessible and open to anyone (even those without an account[4]) and it is possible to search texts by author, topic (especially if marked with a hash, #), or keyword.

There are several factors that explain why politicians prefer Twitter over other SNs such as Facebook or Instagram. First of all, Twitter can be considered a 'non-mediated field' (Testa 2019, 2020)—that is, no one controls how much a politician posts, a clear violation of the Italian *par condicio law* (28/2000) which states that the visibility of a political party on mass media should be determined by the number of votes in the last election. Indeed, new political parties such as Lega and Fratelli d'Italia have disseminated their messages mainly through Twitter and other SNs.[5]

3 Van Dijk (2006) considers than emphasising positive information towards Us and negative information towards Them, on the one hand, and de-emphasising negative information towards Us and positive information towards Them, on the other, are key strategies in polarised discourse.

4 This was at least so until Elon Musk's introduction in 2023 of new rules on data access.

5 'Lega' is commonly used as an abbreviation for both the historical Italian party Lega Nord and for its recent informal successor Lega per Salvini Premier, established in December 2017 by Matteo Salvini. No distinction is made between the two parties in today's speech, one being the continuation of the other, so the terms 'Lega Nord' and simply 'Lega' are used to refer to the same party.

In addition, there is no mediation of disputes, debates, and discussions, unlike in debates broadcast on TV and radio.

This means that Twitter is the perfect place to gain visibility and to draw the attention of a huge number of users. Moreover, citizens seem to have established an awkward equation between the visibility of a politician on SNs, the politician's importance and the quality of their political agendas.

The short time span between the creation of the message in the mind of the politician or the communication advisor and the followers reading that message is another determining factor: there is no time (and no space) for articulated speeches. The only objective is to make an impact on the audience. Given that incendiary news is disseminated more quickly, politicians will often choose to make an impact through messages conveying disturbing information or by inspiring negative emotions regarding a particular fact.

One final characteristic of the way Twitter has reshaped political discourse is that the messages are linked to a particular politician, for politics has become personal—it is no longer just a question of a party or an ideology. In addition, citizens have the illusion of having 'direct contact' with the politicians in that they are able to respond to their posts. Most politicians, however, do not read, much less respond to, citizens' posts or messages. The use of SNs is unidirectional on their side (Antonelli 2017: 11): they are not interested in knowing what the people think about their acts or decisions—they simply use SNs as a means of political propaganda.

Among the linguistic characteristics of this type of text (see Brocca, Garassino, and Masia 2016), our study will focus on the presence of insults, which are very frequently used in political discourse to delegitimise an opponent. The frequency of insulting lexical elements and discursive strategies is indicative of a change in the social consideration of insults. While insults were until relatively recently considered a sign of a low level of education, a lack of argumentative resources, and male chauvinism, the politicians of the 2020s belonging to different political orientations,

with different degrees of education, and without gender distinctions use insults (with the exception of blasphemy; see Dell'Anna 2009; Antonelli 2017; Faloppa 2020). Andrino and Pérez Colomé (2021) report that in the campaign for the regional elections in Madrid in April 2021, 79,840 tweets with insults were published, 5346 of which were directed at candidates or parties. Insults are often used as lexical tools to convey hate speech but not all types and forms of insults count by themselves as hate speech, which is a more complex phenomenon (see Chapter 1, Section 1.3, in this volume).

SNs foster the proliferation and dissemination of insults, which are of course not only found in tweets posted by politicians. In fact, insults are a linguistic feature that allow a Twitter user to be identified as a hater or a troll (Pistolesi 2020), normally interacting under a false profile and hiding their real identity.[6]

3.1.3 Aims and structure

The aim of this chapter is to analyse the form, frequency, and types of insults found in tweets published by Italian politicians between 2020 and 2022. Given that lexical insults are easily identifiable by automatic filters in most SNs, we have observed that insulting strategies in this type of text do not always match the most prototypical slurs and insulting epithets reported in previous Italian studies (De Mauro 2016; see Section 3.4.2 below). Politicians instead show a preference for negatively connotated terms that acquire insulting functions in specific contexts and for more elaborate discursive strategies based on rhetorical figures.

Based on the data in our corpus, a second aim is to establish a typology of all the insulting mechanisms found in these short texts to demonstrate the variety and richness of linguistic devices

6 According to Pistolesi (2020: 97–98), the main difference between a hater or flamer and a troll is the use of insults. While haters use insults frequently, trolls do so less often, instead looking to disrupt other people's conversation with provocative, senseless, or offensive actions.

that are intended to fulfil a denigratory function. Finally, we note the need to widen the concept of 'insult' to accommodate these new textual forms, provided that the pragmatic function is not modified.

For this reason, it is important to determine what is meant by 'insult' in this specific context. In Section 3.2 we offer a definition of insult in the framework of the theory of speech acts, thus taking into account its illocutionary characteristics and its perlocutionary effects. This definition highlights the pragmatic nature of insults and does not circumscribe insults to a pre-established set of linguistic characteristics. In Section 3.3 the corpus extracted from Twitter is presented and some methodological decisions are addressed, while in Section 3.4 the typology of insults that emerged from the analysis of our corpus is introduced based on two main criteria: the role of the addressee (Section 3.4.1) and the linguistic mechanisms at play. Regarding this last criterion, we distinguish between insults based on lexical elements, mainly but not solely epithets (Section 3.4.2), and insults of a more discursive and rhetorical nature which are constructed following the scheme of rhetorical figures (Section 3.5), mainly metaphors (Section 3.5.1) but also hyperboles, parallelisms, and irony (Section 3.5.2). The chapter ends with the conclusions of our study regarding the creativity of political discourse on SNs (Section 3.6).

3.2 Insulting as a speech act

Insults are complex social phenomena that have a variety of forms and fulfil different functions according to different cultural contexts, languages, speakers, and communicative aims (Domaneschi 2020: 10). They have been at the centre of 21st-century multidisciplinary research, in fields such as psychology, anthropology, sociology, law, philosophy of language and linguistics, among others (see, among many others, Cepollaro 2020; Domaneschi 2020; Faloppa 2020; Bianchi 2021; Nitti 2021, with specific reference to Italian; see also Chapter 2 in this volume).

From a psychological point of view, insults have been studied with regard to how they shape and transform our social identity, because they can weaken the sense of belonging to society among some groups of individuals. The psychological effects of insults are intertwined with the sociological perspective: insults reinforce social asymmetries and discrimination; they contribute to marginalising ethnic, national, religious, or gender-based minorities. Therefore, legal studies consider some insults a crime under certain circumstances because they threaten social cohesion and democratic values.[7] There is, of course, a big debate about insults and the limits of freedom of expression to which different countries have reacted in different ways, with the divergence between the legislation in EU countries, the UK, and the USA being particularly striking (Domaneschi 2020: 8–10; Faloppa 2020).

For our purposes, insults will be approached from a pragmatic perspective, and specifically within the framework of the theory of speech acts (Austin 1962; Searle 1969). Uttering an insult is a speech act *per se*, with a particular illocutionary force and perlocutionary effects. Following Austin's classification, it can be considered a verdictive speech act similar to judging, evaluating, or condemning. Regarding the illocutionary force, we may say that an insult conveys the speaker's intention to anger, to humiliate, to shame, to disregard, or to hurt someone or something with their words, but also to show their power, to attract attention, to force someone to do something, and even to show affect (Domaneschi 2020: xiv).[8]

Moreover, regarding the functions of language, insulting has both an emotive and a referential function. The emotive function

7 These circumstances mainly include the cases when insults are uttered in public, damaging the honour of the insulted person by accusing him/her of committing certain acts with prior knowledge of their falsity or with reckless disregard for the truth.

8 In this last case, which will not be taken into account in our study, insults are means to reinforce social ties between interlocutors and are then considered strategies of positive politeness or, to use Zimmerman's (2003: 57) terminology, anti-politeness.

is linked to the fact that insults convey the speaker's negative emotions, such as disdain, scorn, contempt, or disgust. At the same time, insulting someone involves offering a negative representation or evaluation of the insulted person; there is a claim, an assertion about someone, there is an implicit or explicit predicative function by which one characteristic or quality is attributed to a subject. This underlying structure distinguishes insults from curses, which express a wish for the future. To the extent that there is a type of representation (i.e. a link between the words and the extra-linguistic reality), it is possible to speak of a referential function, although the emotive functions overwhelm it in the act of insulting.

To understand why insults are so frequent in political discourse, it is important to analyse their perlocutionary effects. These can be divided into three types:

a) Effects on the addressee (the political opponent in our case): insults cause a wide variety of emotions ranging from intimidation to fear and rage.[9] One of the most common effects is offence—that is, awareness of having been morally, psychologically, or economically harmed (Domaneschi 2020: 42). They are face-threating acts (Culpeper 1996; Palermo 2020: 2) that cause politicians to lose their credibility and their authority, and sometimes their honour; this loss may have direct consequences for their career. Insults are a very effective means of delegitimising opponents and discrediting them in front of potential voters.

9 Scholars have discussed how important the effects on the addressee and the speaker's intention are in defining a speech act as an insult. In other words, can we define something as an insult if no one feels insulted or if no one intends to be insulting, even when our words are perceived as offensive? This question is not relevant to this study in that every post in our corpus reveals a clear intention to insult a target, which will be accepted as a sufficient criterion to consider the posts as insults, even if we have no access to the target's reaction (i.e. to their perception of those posts as insults), except when the readers retweet and comment on them.

b) Effects on supporters: these are the most interesting effects for the purpose of our research. Attacking a rival is perceived as a sign of strength on the part of the speaker and, as a consequence, the speaker is viewed as a competent, courageous, and coherent person. In other words, insults reinforce the speaker's public image (Palermo 2020); this means that the prestige of the insulting person is increased, which may have an electoral return (see Cavazza and Guidetti 2014). Insults also help to strengthen political positions, by acting as political and ideological propaganda.

   In addition, insults encourage discrimination, hatred, and violence against some individuals, groups, and communities. By doing so, they bolster the sense of belonging to a dominant group among like-minded audiences, thereby reinforcing social prejudices and stereotypes. This is often expressed through the well-known opposition between 'we' and 'they', which creates identity borders to separate two groups (Van Dijk 2006; Paris 2019): the group of people who share the same political ideology and world view as the author of the post and the group of people who are denigrated or represented by the humiliated political opponent.

   Therefore, despite the common negative evaluation regarding the act of insulting, the fact is that insults have a *covert prestige*, to use Labov's words, and convey values such as authenticity, closeness to the people, genuineness, courage, and so on (Labov, cited in Domaneschi 2020: 131).

c) Effects on 'neutral' observers: even when the audience do not have a clear political position in favour of the speaker, it has been proved that insults weaken political opponents and help to normalise discriminatory behaviours and attitudes. In fact, they can transform harmless individuals into a threatening group (Bianchi 2021: 11).

However, without denying the importance of the theory of speech acts in defining what can be considered an insult, the pragmatic analysis shows that insults are not always easily identifiable and

are highly context-dependent (Alfonzetti 2009: 67). What can be considered an insult in one context may not be considered as such in a different one. Terms such as 'communist' and 'fascist' describe historical movements and ideologies, but in political debates on- and offline they are employed to vilify the opponent with the audience's complicity.

Insults are often accompanied by other aggressive acts such as accusations, threats, and curses, which are not always easy to distinguish from insult itself. Another way of approaching the task of defining insults is within the framework of prototype theory. The characteristics of a prototypical insult are the following (Alfonzetti 2009: 71–77):[10]

a) An insult is a verdictive act: a negative judgement or a negative evaluation about a person, regarding their physical characteristics, personality, facts and actions, moral qualities, etc.
b) An insult is an expressive act: the speaker expresses an emotion, a feeling regarding the addressee such as hate, rage, contempt, or disdain.
c) The speaker has the intention of causing offence, or of angering, vilifying, or harming the addressee.
d) The insult has perlocutionary effects, i.e. it psychologically affects the recipient.
e) The addressee must be present in the communicative situation.

10 Regarding characteristics (e), (g), and (j) in the list, insults on SNs are not prototypical insults because the addressee is never present in the same communicative situation in which the insult is produced and is not always directly addressed (see Section 3.4.1). Being of written nature, paralinguistic, kinetic, and proxemic elements play no role in this type of insult. For other speech acts that are similar to insults but do not share all of the prototypical features, see Alfonzetti (2009: 73–74), who considers that defamation is a different speech act, while in this study—as we shall see later in this chapter—defamation is considered a specific type of insult.

f) The addressee must interpret the insult as offensive. And this interpretation is necessarily based on a common axiological system.
g) The insult is directed to the recipient in a vocative form.
h) The use of negatively connotated adjectives, nouns, adverbs, and so on.
i) The syntactic structure is reduced to a noun phrase (normally as apostrophe), categorisation structures N+*di*+N (N + of + N), assertive sentences, rhetoric questions, and emphatic constructions (e.g. *che X che sei*, 'what a [X] you are').
j) Insults are accompanied by several paralinguistic (high volume), kinetic, and proxemic phenomena.

According to Domaneschi (2020: 64), three contextual factors seem to determine the identification of an insult: the speaker's status, the place of production, and the power that the speaker has in that place. When these conditions are met, saying something (particular words or expressions) becomes doing something (offending, angering, humiliating). Insulting is not just a matter of uttering a negative evaluation about someone; it also presents the speaker as having the power, the capacity, and the right to do so. Politicians have a high social status and hold a privileged position in accessing economic resources and information. They also occupy a privileged position on Twitter, shown by their number of followers and the reactions and comments raised by each of their tweets. This position is the source of their *auctoritas*, the moral locus from which they maintain the right to insult the opponent. But a clarification is required here: the relationship between the insulting and the insulted subjects does not necessarily pre-date the insult itself; on the contrary, it can be a consequence of the concrete speech act of insulting.

In this study, following Canobbio (2010), insults will be defined as linguistic elements (words, phrases, clauses, sentences, textual structures) that speakers use when performing speech acts that are intended to cause offence to the addressee. What is important

in this definition is that the criteria for distinguishing what is an insult and what is not are not strictly linguistic (let alone lexical) but pragmatic: an insult is defined based on the speaker's intentions and the harm it can cause to the addressee and the audience (Alfonzetti and Spampinaro Beretta 2010). From the politeness theory perspective, insults constitute an act of non-mitigated disagreement that threaten the positive face of the addressee (Brown and Levinson 1987; Palermo 2020) and block any further negotiation (Moïse 2006). They represent a point of high tension in verbal interaction because they are manifestations of non-cooperative interaction and impoliteness (Culpeper 1996), a type of verbal violence which might evoke (and provoke) physical violence.

3.3 The corpus

In order to analyse insults as speech acts in political discourse on SNs, and particularly on Twitter, we have manually collected and analysed a small sample of 250 tweets from 27 politicians belonging to seven different political parties, from left wing to right wing: Movimento Cinque Stelle (M5S, Five Star Movement), Partito Democratico, Azione, Italia Viva, Forza Italia, Fratelli d'Italia, and Lega.[11] The names of the politicians are displayed in Table 3.1.

The tweets were collected between August 2020 and May 2022, and they deal with a wide variety of topics: the COVID-19 pandemic, immigration, social revolts, parliamentary activities, new laws. No specific hashtag or keyword was selected.

11 Our corpus is very small compared to the usual dimensions of Twitter corpora in other studies, because we have chosen to undertake a qualitative analysis. No automatic filters were used in the selection of tweets in order to avoid a selection based on lexical criteria or hashtags. We carried out a manual search on Twitter at different times and looking at different political accounts for two years, and selected only those tweets with a clear insulting intention. It is our intention to enlarge the corpus; here we are presenting only some preliminary results.

Table 3.1: Italian politicians and political parties represented in the corpus.

M5S	Partito Demo-cratico	Azione	Italia Viva	Forza Italia	Fratelli d'Italia	Lega	Inde-pendent
L. Azzolina	M. Di Maio	C. Calenda	M. Renzi	S. Berlusconi	D. Santanchè	M. Salvini	C. Cottarelli
C. Sibilia		M. Richetti	M. E. Boschi	M. Gasparri	G. Meloni	C. Borghi	
V. Raggi		F. Carpano	L. Noja	L. Ronzulli	I. La Russa		
N. Morra			T. Bellanova	E. Vito	W. Rizzetto		
D. Toninelli			L. Nobili				

In collecting the tweets, the aim was to achieve a balance in the ideology of their authors, the topics, and the time of year. However, it was not easy to obtain the same number of tweets for each political party, as some parties, such as Fratelli d'Italia, were much more active on Twitter than others, as shown in Table 3.2. The average number of tweets per party is 15–17 but M5S and Forza Italia are underrepresented in our corpus, while Fratelli d'Italia is overrepresented (46.8 per cent of the tweets in our corpus were posted by their members).

Table 3.2: Number of tweets per political party in our corpus.

Political party	Number of tweets	Percentage of tweets in the corpus
M5S	5	2.0
PD	17	6.8
Azione	17	6.8
Italia Viva	13	5.2
Forza Italia	8	3.2
Fratelli d'Italia	117	46.8
Lega	17	6.8

The same imbalance is found in the number of tweets per politician. It was not possible to obtain an equal number of tweets from each politician, again because some politicians publish not only a higher number of tweets than others but also a higher number of tweets containing insults. The most active politician on Twitter is without doubt Giorgia Meloni, the leader of Fratelli d'Italia who became prime minister in October 2022, and she also authored a high number of tweets with insulting mechanisms in our corpus (92). Other politicians whose tweets often have a clear insulting function are Mario Di Maio (17), Daniela Santanchè (12), Matteo Salvini (10), Matteo Renzi, and Matteo Richetti (9 each).

We are conscious of this imbalance in our corpus, but for the purposes of our study—which does not aim to correlate ideologies and political movements with insulting strategies—we consider a corpus of 250 tweets, all exhibiting insulting strategies, to be sufficient as a first step in our research to explore the discursive dimension of the insulting strategies in contemporary political discourse. The focus here is on the linguistic mechanisms that are subordinate to the insulting function and not on the language of different political parties from a comparative approach.

3.4 Types of insults

Insults will be classified according to two different criteria: a) the addressee, and b) the linguistic mechanisms involved.

3.4.1 The role of the addressee

An insult is a communicative event with two main participants: the addresser and the addressee. Depending on the addressee, insults can be classified as injury, defamation, or blasphemy.[12] Injury is an insult directly addressed to a specific person, not necessarily in

12 Not all scholars agree with this classification (see, e.g., Alfonzetti 2009). For an attempt at clarifying concepts such as offence, defamation, outrage, contempt, and slander, see Domaneschi (2020: Ch. 2).

the presence of others—in other words, it can be a private event; defamation is an insult about someone or something in front of an audience which is not the target of the insult (Palermo 2020), and not necessarily in the presence of the insulted person; blasphemy is an insult addressed to God or to a person, object, or place considered sacred or linked to divinity in some way (Domaneschi 2020: 125–129). This last type of insult is extremely rare in Italian political discourse because it offends the sensibilities of a substantial proportion of the population and is thus carefully avoided even by politicians who openly claim to be atheist. Moreover, blasphemy has sociological connotations, in that it is usually considered a sign of a low level of education, of a limited capacity to present personal opinions in well-constructed discourse, and of impoliteness.

Insults on Twitter are clearly of the second type, because politicians insult opponents or rivals as a discourse strategy to gain credit or to reinforce their own political position in front of an audience of possible voters. According to the Collins Dictionary, 'defamation is the damaging of someone's good reputation by saying something bad and untrue about them'.[13] One way of determining whether an insult can be considered injury or defamation is by analysing personal pronouns. As shown in Figure 3.1, the most used personal pronouns in politician's tweets are those of the third person, while the second person, either singular or plural, which is the form found in injuries, is rarely used.

13 *Collins Dictionary*, s.v. 'Defamation (*n.*)', accessed 25 July 2024, https://www.collinsdictionary.com/dictionary/english/defamation.

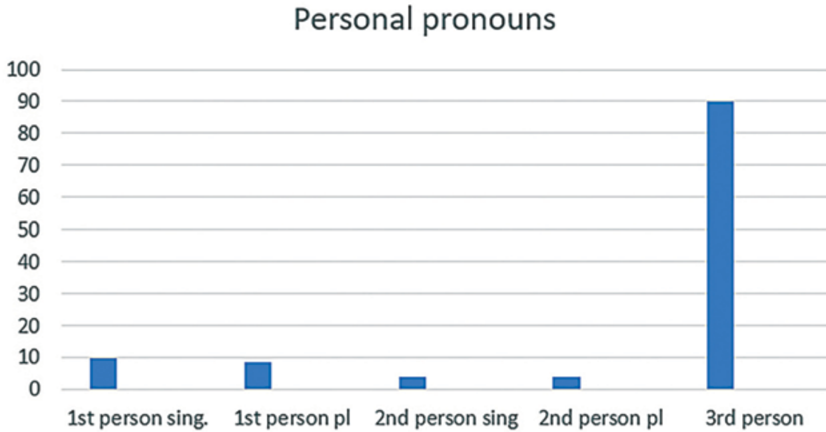


Figure 3.1: Personal pronouns in Italian politicians' tweets.

An example of injury using the second-person singular is:

(1) '**You** can attack me as much as you like. […] And **remember**: Rome will never vote Lega Nord, will never vote for the person who cried "thieving Rome" or for those fascists and racists that **you** put on your lists' (@virginiaraggi, 9 September 2021)

An example of a politician directly addressing the audience is (2), where Giorgia Meloni creates a very dynamic text by placing the first-person pronoun and possessive (*i miei libri*, 'my books'; *io*

'I') in opposition to the third-person singular (a university professor); she uses the well-known opposition 'we' vs 'they' (*noi di destra*, 'we right-wing parties') while still giving the impression that she is interacting with the audience (who are addressed using the second-person plural pronoun *vi*, 'you'):

(2) 'Do **you** think it is acceptable that a university professor should joke about the fact that **my books** have been turned upside down to symbolise that **I** should be hanged? This is one of the "brains" teaching respect, tolerance and freedom of expression to the youth. Thank goodness **we** right-wing parties are the haters…' (@GiorgiaMeloni, 28 May 2021)[14]

Giorgia Meloni
@GiorgiaMeloni

Ma vi sembra normale che un docente universitario scherzi sui miei libri ribaltati per simulare il fatto che io venga appesa? Ecco una delle tante "menti" che insegnano ai giovani rispetto, tolleranza e libertà di pensiero. Menomale che i seminatori di odio siamo noi di destra...

IL PROFESSORE UNIVERSITARIO CHE SCHERZA SU UNA POLITICA APPESA A TESTA IN GIÙ.

14 Many tweets are multimodal in that besides text they also contain pictures, audio, and video. Their meaning is significantly conditioned by this multimodality, but the study of the global meaning of these posts would require a semiotic analysis. In our study only the linguistic component of the tweets will be described and we are well aware of the limitations of the analysis derived from this decision. In any case, we have not altered the tweets and they are reproduced in this chapter in their original form, to allow the reader to fully understand the text. The only minor edits to the tweets involved recropping and underlining.

However, most tweets take the form not of an interaction but of a description, as in (3).

(3) 'This "gentleman" is simply insane' (@matteosalvinimi, 15 July 2021)

Based on the use of personal pronouns, then, 95 per cent of the tweets in our corpus can be considered defamation, as shown in Figure 3.2.

Nevertheless, an insult on a blog or a SN is never a private insult and, in this sense, every insult on Twitter, regardless of the type of pronouns being used, can be considered defamation—that is, the damaging of someone's good reputation by saying something bad and untrue about them.[15] Besides, the target is never present when the text is materially produced (written and posted on the SN), thus one of the main criteria for an insult to be considered an injury is not fulfilled. Consequently, the data in Figure 3.2 are to

15 *Collins Dictionary*, s.v. 'Defamation (*n.*)', accessed 25 July 2024, https://www.collinsdictionary.com/dictionary/english/defamation.

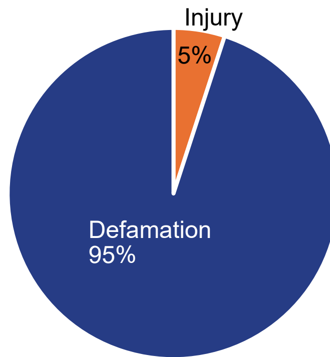


Figure 3.2: Injury vs defamation in politicians' tweets.

be understood as follows: 5 per cent of the tweets imitate a direct dialogue with the target, using second-person pronouns and addressing the insult directly to the target, but the fact of posting the insult on a SN turns it into defamation.

3.4.2 Linguistic mechanisms as insulting strategies: the epithets

According to the types of epithets that are used to convey an insult, we can identify two types of insults: slurs or derogatory insults, and other insulting epithets.

Slurs or derogatory insults

Slurs are insults that are addressed to individuals based on their belonging to a particular group (normally a minority group in a given community) defined on the basis of race, gender, religion, nationality, and so on. This type of insult has received considerable attention in Italian, particularly from philosophers of language (Cepollaro 2020; Bianchi 2021) and linguists (Faloppa 2020). They are very common in SNs.

In Table 3.3 we offer a list of the most common Italian slurs based on previous inventories (De Mauro 2016; Domaneschi 2020: 18–19). Some of them are found cross-linguistically and are

easy to understand; others are specific to the Italian culture (e.g. the derogatory use of white-collar professions to show distrust or suspicion) and unusual in other cultures. This explains why the English translations offered in the table do not always function as derogatory insults in English-speaking contexts, but this is not the place to analyse the specific context of use of each of them. Slurs found in our corpus are marked in bold.

Table 3.3: Examples of Italian slurs.

Types of derogatory insult	Examples
Race	*negro, asiatico, giallo* 'nigger, Asian, yellow'[16]
Sexual orientation	*frocio, lesbica, paraculo, travestito* 'fag, lesbian, bastard, transvestite'
Nationality	***cinese***, *albanese, bulgaro, beduino, ebreo, giudeo, zulù, mongolo, turco, sodomita* 'Chinese, Albanian, Bulgarian, Bedouin, Jew, Jewish, Zulu, Mongol, Turk, sodomite'
Religion	***islamista*** 'Islamist'
Region or city (North vs South Italy)	*terrone, polentone, meridionale, genovese* 'peasant [a pejorative term for Southern Italians], polenta-eaters [a pejorative term for Northern Italians], southerner, Genoese'
Social stereotypes	*gesuita, mammalucco, ayatollah, mafioso* 'Jesuit, Moor, ayatollah, mafia man'
Humble professions	*pescivendolo, cafone, buffone, carrettiere, parrucchiere, pecoraio, portinaia, scaricatore di porto* 'fishmonger, oaf, buffoon, cart driver, hairdresser, shepherd, doorman, docker'
'Well-respected' professions	*accademico, professore, avvocato, leguleio, paglietta, cattedratico, politico* 'intellectual, professor, lawyer, university professor, politician'
Political orientation	***comunista, fascista, nazista, populista, antisemita, immigrazionista, grillino*** 'communist, fascist, Nazi, populist, antisemite, immigrationist, Grillo supporter'[17]

16 Translations to English are only approximate and meant to help the reader, as it is very difficult to find an insult that will cause the same impact and that will point to the same characteristic as the original Italian insult.

17 Beppe Grillo, an Italian comedian, founded M5S in 2009; the party has been very active in Italian politics ever since.

The characteristic of this type of insult is that even if it appears in a negated assertion such as:

(4) Carlo non è frocio.

'Carlo is not a faggot.'

The negative connotation and the offence directed to the group (in this case homosexual people) does not disappear because it is entailed by the epithet (Domaneschi 2020: 111). However, slurs are not common in political discourse because politicians are aware that, by insulting a minority group, they may lose potential voters. The slurs in our corpus are thus almost entirely limited to political orientation ('communist', 'fascist', 'Nazi', 'populist', etc.) because 'one of the quickest ways for an extremist to discredit anyone who disagrees with them is to call them a sexist, a fascist a racist, a nazi or any other "ist" word, primarily because they are deeply damaging and "sticky" labels' (Bule 2017). In addition, the interpretation of an epithet that uses a political orientation as a derogatory insult implies sharing a common axiological system (Alfonzetti 2009: 72). Slurs relating to political orientation have therefore become an effective way of discrediting opposing perspectives, causing deep fractures in civil society and democratic institutions.

It is also possible to find slurs directed towards groups that do not have the right to vote in the country, such as the so-called 'illegal immigrants', or foreign citizens such as Chinese people. Right-wing parties also insult individuals who belong to religions other than Catholicism by using slurs such as 'Islamic'. Examples (5) and (6) contain some of the slurs mentioned above:

(5) 'Italy, Europe, Western world: shame! To leave women and children in the hands of **Islamic** throat-slitters is not human [...]' (@matteosalvinimi, 15 August 2021)

Italia, Europa, Occidente:
vergogna!
Lasciare donne e bambini in mano
ai tagliagole islamici, dopo anni
di battaglie e sofferenza, non
è umano. Qualcuno al governo
dovrebbe rileggersi "La rabbia
e l'orgoglio" e "La forza della
ragione" della grandissima Oriana
Fallaci.

(6) '[...] Anyone who winks to the anti-vaxxers supporters in the name of generic "freedom" is putting Italy at risk. Let's listen to the science, not to **populists**.' (@marcodimaio, 29 July 2021)

Anche #Fedriga sconfessa
#Salvini sul #GreenPass. La
battaglia contro il Virus dovrebbe
essere di tutti, senza distinzioni
politiche. Chi strizza l'occhio ai
NoVax in nome di una generica
"libertà" mette in pericolo l'Italia.
Ascoltiamo la scienza, non i
populisti

Other insulting epithets

Insulting epithets addressed to opponents on the basis of individual characteristics and not because of their belonging to a group are found more frequently. The insulting function of these epithets is based on the relation between the source of meaning which points to a specific experiential area (a scatological element, for instance) and the target, the insulted person, who is often considered as deviating from an idealised model. This explains why people with physical or mental disabilities are often targeted and thus stigmatised (Domaneschi 2020: 25–28).

Many of these epithets (or nouns used as epithets) have no intrinsic negative value (e.g. nouns referring animals and vegetables). They become insults in specific contexts in which the participants in the communicative interaction share a common cultural background and axiological system. Some of the most frequent Italian insulting epithets, arranged by semantic fields, are shown in Table 3.4 (De Mauro 2016; Domaneschi 2020: 18–19; Faloppa 2020; Palermo 2020).

Table 3.4: Examples of Italian insulting epithets

Types of insulting epithets	Examples
Psychological characteristics (in reference to mental disability)	*imbecille, idiota, cretino, minorato, tonto, ritardato, inetto, analfabeta, folle* 'imbecile, idiot, dumb, retard, stupid, incompetent, illiterate, crazy'
Physical characteristics (in reference to physical deformity and disability)	*gobbo, zoppo, abnorme, handicappato* 'hunchback, cripple, abnormal, handicapped'
Character traits (in reference to negative behaviours)	*imbroglione, pigro, scansafatiche, tirchio, bigotto, falso, ipocrita, intrigante, cattivo* 'swindler, lazy, lazybones, miser, sanctimonious, false, hypocrite, meddlesome, evil'

Types of insulting epithets	Examples
Sexual organs and sexual attitudes	*cazzo, cacchio, minchia* (when manipulated within a nominal expression: *testa di cazzo* 'dickhead'), *coglione, puttana, rotto in culo, cornuto* 'dick, euphemism for *cazzo*, prick, moron, bitch, fag, cuckold'
Criminal activities	*criminale, ladro, terrorista, assassino* 'criminal, thief, terrorist, murder'
Scatological elements	*stronzo, pezzo di merda, cesso* 'asshole, piece of shit, toilet'
Animals	*maiale, asino, troia, cagna, vacca, zoccola, pappagallo* 'pig, donkey, female pig [whore], bitch, cow [whore], sewer rat [whore], parrot'
Vegetables	*finocchio, broccolo, pera cotta, (testa di) rapa* 'fennel [faggot], broccoli [fool], cooked pear [fool], turnip [block head]'

This type of insult, particularly those regarding psychological characteristics, character traits, and criminal activities, is much more frequent. It is possible to establish a scale in which criticism of social abilities (*folle*, 'crazy'; *insensate*, 'insane') or degree of competence (*incompetente*, 'incompetent'; *incapace*, 'unable') occupies a lower position than attacks on the moral quality of the person (*vergognoso*, 'shameful'; *indegno*, 'ignoble'; *bugiardo*, 'liar'). In our data, moral insults, which are considered more harmful to public image and prestige, are by far the most frequent. This means that politicians try to cause moral harm to their opponents as a recurrent strategy to discredit them.

In many cases, these epithets are used to modify nouns and verbs that describe politicians' words and actions, such as *folli misure restrittive* ('crazy restrictive measures'), *vergognose affermazioni* ('shameful claims'), *coprifuoco insensato* ('foolish curfew'). They can also be nominalised, as in (7), or can be part of an attributive structure, as in (8):

(7) '[…] Now the government goes directly to Libya to bow and scrape and to kiss the slippers of the Libyan tribe leaders. So have the **incompetents** that govern us humiliated Italy.' (@GiorgiaMeloni, 19 December 2021)

(8) '[…] In a moment of deep crisis, it is **shameful** that the government continues to ruin citizens and companies. […]' (@GiorgiaMeloni, 8 September 2021)

A quantitative analysis of the two types of insults described so far shows that, contrary to what is found in the literature, epithets are not the most common linguistic strategy for insulting used by politicians on SNs. In fact, slurs represent only 6 per cent of

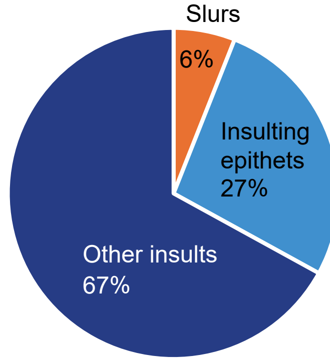


Figure 3.3: Preliminary quantitative analysis of insulting mechanisms in our corpus.

the insulting mechanisms (for the reasons mentioned above) and insulting epithets only 27 per cent, as can be seen in Figure 3.3.

In fact, many of the insulting lexical elements found in our corpus are not adjectival in nature but nominal or verbal: nouns and verbs convey negative connotations as frequently as adjectives, as can be seen in Figure 3.4. They refer to:

a) Agents: *aggressore* ('aggressor'), *odiatore* ('hater'), *truffatore* ('cheater')
b) Attitudes: *sdegno* ('disdain'), *ipocresia* ('hypocrisy'), *intolleranza* ('intollerance')
c) Actions: *bugie* ('lies'), *latrocinio* ('robbery'), *follia* ('madness')
   i Fraudulent actions: *rovinare* ('to ruin'), *danneggiare* ('to damage'), *attaccare* ('to attack')
   ii Negated positive actions: *non avere idee* ('not to have ideas'), *non meritare* ('not to deserve'), *non sapere* ('not to know')
   iii Endured actions (presented from the point of view of the victim): *essere attaccato* ('to be attacked'), *essere calpestato* ('to be stepped on'), *essere parte lesa* ('to be the offended party').

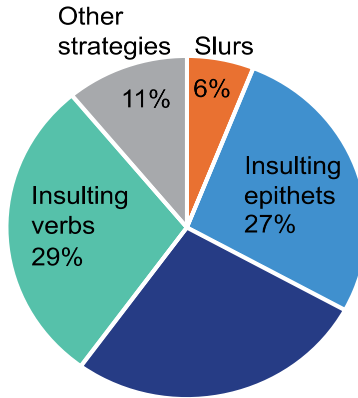


Figure 3.4: A quantitative analysis of insulting strategies in our corpus.

Some examples to illustrate this use are in (9) and (10) (cf. *Così hanno ridotto l'Italia* 'So have they humiliated Italy' in (7) and *massacrare* 'to ruin' in (8) above):

(9) 'We are depriving the students of years of life, I hope it will be soon possible to go back to school safely. But to know that all this depends on Minister Azzolina, known for her **incompetence**, is not reassuring […]' (@matteosalvinimi, 23 November 2021)

(10) ‘The government has approved a mandatory #greenpass, a pass that **jeopardises** citizens’ freedom, further **destroys** the economy and [...] It is the umpteenth **shame** [...]’ (@GiorgiaMeloni, 22 July 2021)

In some cases, the negatively connotated noun has the textual function of a labelling tag or anaphoric encapsulator, both of which are very frequent in journalistic and political discourse (see, among others, D’Addio Colosimo 1988; Francis 1994; Conte 1999a,b; Borreguero 2006, 2018; González Ruiz 2008, 2010; Lala 2010; Llamas 2010a,b; Izquierdo Alegría and González Ruiz 2013; López Samaniego 2015; Korzen 2016). This type of anaphor summarises a previous idea by adding a valuative tag, like in the case of *estratto del delirio di Conte* (‘an excerpt of Conte’s delirium’) in (11):

(11) '"The security decrees have thrown thousands of immigrants into the streets deployed in the suburbs and the countryside: Salvini has failed as a minister; it is a fact". An excerpt of #Conte's **delirium** for Il Corriere. But does he really think that he can fool the Italian people forever?' (@marcodimaio, 9 July 2021)

Although a lexical study would certainly be interesting,[18] our research focuses on other insulting strategies which are not exclusively lexical and have a more discursive nature. We will call them

18 For instance, linguistic creativity in the configuration of new insults is a very interesting field of research that will not be dealt with in this study. In most languages, some prefixes and suffixes have acquired pejorative values and are found particularly frequently in the formation of insults. In Italian, this is the case for *-uccio* (*professoruccio*) and *sub-* (*subnormale*). See Domaneschi (2020: Ch. 1) for these and other linguistic characteristics of lexical insults.

rhetorical insults, and they represent 11 per cent of the insulting devices in our corpus.

3.5 Rhetorical insults

Rhetorical insults are discursive insults: the insulting function is not fulfilled solely by lexical elements, but results from a more complex and developed discursive structure. In fact, lexical elements such as those analysed in Section 3.4.2 are integrated into carefully planned textual structures.

We have labelled these structures 'rhetorical insults' because they are based on rhetorical figures, such as parallelism, metaphors, and irony, among others. They are highly polyphonic in that they quote, summarise, and attribute words to others, words that let the reader deduce what the ideas or behaviour of the insulted person are.

Rhetorical insults also require a certain cultural common knowledge in order to identify idioms and referents and also a specific understanding of the highlights of the political scene (the most recent facts and declarations, the most active or prominent politicians).

Finally, an important advantage is that, unlike slurs, they cannot be easily detected by automatic filters and can then be disseminated in a more efficient way.

We will focus on four rhetorical structures that are particularly frequent in our corpus: metaphors, hyperbole, parallelism, and irony. We will then illustrate how cultural referents play a role as part of the insulting strategy.

3.5.1 Metaphors

Metaphors are omnipresent in political discourse and have been approached from rhetorical, cognitive, and textual perspectives (see Otieno, Owino, and Attyang 2016). They are considered a very powerful rhetorical strategy due to their persuasive potential. In fact, metaphors structure our understanding of political, social,

and economic issues. The conceptual metaphor 'politics is war', for instance, structures the way we think about politics as a battle to be won. Another example is the use of the metaphor of war to explain governmental actions during the COVID-19 pandemic in 2020 (see Castro Seixas 2021).

According to Lakoff and Johnson (1980), a conceptual metaphor is a pervasive culture-wide disposition to conceive one fixed sort of thing in terms of another fixed sort of thing. In every metaphor, a source area and a target area can be identified. The source area is the cultural or experiential area from which the literal meaning of the expression introducing the metaphor stems, while the target area is a more abstract area offering more effective interpretations of the metaphor. Conceptual metaphors can be universal or culture specific. Our cultural backgrounds influence our perception of the world and our use of metaphors. In many cases, metaphors represent subconscious choices on the part of the speaker, based partly on the conceptual structures shared by members of their community (Otieno, Owino, and Attyang 2016: 23).

Metaphors help to shape the structure of political categorisation and argumentation. A good example is the conceptual metaphor 'politics is a game' as opposed to 'politics is war', which shapes our perception of politics. Metaphors reflect social and cultural constructions to conceptualise the political world but have a less culture-specific nature (i.e. they are more generalised) than metaphors employed to describe personal and familiar relationships, for example.

In the case of the metaphors employed by politicians as insulting strategies, the main source areas are animals (e.g. animalised behaviours such as the way pigs eat and live are attributed to political opponents), dirty places (e.g. places and activities related to the political sphere are described as cesspits or swamps), criminal activities (e.g. politicians are accused of holding the country to ransom and, as a consequence, the country is presented as a victim). An example of this last type of metaphor can be seen in (12), while (13) is a good illustration of how animalised behaviours are

used to portray political opponents. The animal metaphor is an efficient way of dehumanising rivals, and of reducing them to the cognitive and moral level of a beast (Domínguez and Zawislawska 2006; Domaneschi 2020: 92).[19] In this case the image of a jackal pouncing on its prey is used to depict the desire of some parties to administrate European funds:

(12) 'The US vice-president and idol of the left, Kamala Harris, says that illegal immigration will be persecuted: the US will defend its own borders, and will "push back" anyone who illegally crosses them. Like every other nation in the world. Except Italy, **hostage** of immigration-friendly left-wing parties' (@GiorgiaMeloni, 8 June 2021)

Giorgia Meloni
@GiorgiaMeloni

Il vice presidente USA e idolo della sinistra Kamala Harris dice che l'immigrazione illegale sarà contrastata: gli USA difenderanno i propri confini, "respingendo" chi entra illegalmente. Come fa qualsiasi Nazione al mondo. Tranne l'Italia, ostaggio della sinistra immigrazionista

19 In fact, some Italian politicians receive nicknames based on these animal metaphors: Berlusconi, *il caimano* ('the caiman'), Craxi, *il cinghialone* ('the big wild boar'), Salvini, *il capitone* ('the large eel') (Domaneschi 2020: 130).

(13) ‘While Renzi plays the game of destruction, #WeGoOn-WithConti in order to not allow the **jackals to pounce** on the Italians’ safe known as the #RecoveryPlan’ (@DaniloToninelli, 13 January 2021)

3.5.2 Hyperboles, parallelisms, and irony

Metaphors are not the only rhetorical devices found in our corpus. Hyperbole is another traditional strategy that has been used in political discourse since ancient times, with Cicero and Quintilian two of its most emblematic representatives. Hyperbole is an exaggeration in the description of a state of affairs: it exceeds the credible limits of facts in a given context (Claridge 2011: 5), but to be effective it has to have its basis in an intersubjective perception of the state of affairs. The literal and the corresponding hyperbolic expression are part of the same scale.

While hyperbole is a mechanism of linguistic creativity and an important contributor to language change (Claridge 2011), it also is a powerful means of manipulation because it is aimed directly at the addressee’s emotions. When confronted with hyperbole, the audience’s focus is not on the message but on the emotions the hyperbole inspires.

We have several examples in our corpus that primarily feature harsh criticism by some right-wing parties (mainly Fratelli d’Italia) towards the left-wing government:

(14) 'A left-wing party that lives on Mars [...]' (@FratellidItalia, 1 September 2021)

(15) 'The attempted silencing of the opposition continues [...] Welcome to North Korea' (@GiorgiaMeloni, 31 May 2021)

By using these hyperboles, Fratelli d'Italia and its leader, Giorgia Meloni, transform a rational criticism of the government into a claim whose main purpose is to provoke an emotional response in the audience: instead of criticising politicians' lack of awareness of people's real problems, they prefer to say that the government is living on another planet (the planet is often Mars and they speak of a Martian left-wing party); instead of accusing the mass media of discriminating against the opposition on some TV programmes, they compare the political situation with a dictatorship, making the audience forget the substantial distance between a democratic system such as the Italian one and an authoritarian system such as that of North Korea.

Hyperbole also has a side-effect: it undermines the credibility of truthful claims. 'The more false claims that we see, the less likely we are to believe the truthful claims that try to counter them and that is how we get to the point where we no longer believe anything, even if it's backed by good science' (Bule 2017).

Another rhetorical device which is very frequent in our corpus is irony, which requires a particular interpretative effort on the part of the audience. Readers first need to decodify the text and process its literal meaning, and then they need to understand that this literal meaning is negated and the message is different, based on the interplay between the literal meaning and readers' knowledge of the world (in our case, the political relationships). This means that the writer presupposes that the audience has a good knowledge of the political situation and is able to infer the writer's intentions; in other words, a complicity may arise between the speaker and the interpreter. The interpreter then becomes aware of the insincerity of the claims. For an ironic message to be effective it must be clear for the interpreter that the speaker does not believe what he says (Pistolesi 2020: 90).

Let us look at two examples by Matteo Salvini, the leader of the right-wing party Lega. In (16) Salvini claims 'I am longing to meet the nice German rammer'. In fact, he has no desire at all to meet the person in question because the meeting will take place at a trial. To understand the irony, the audience must be acquainted with the fact that Salvini and Carola Rackete, the German sea captain, had a conflict when the former Italian Minister of Foreign Affairs did not grant permission for her boat, *Sea Watch 3*, carrying 53 people rescued from a shipwreck in the Mediterranean Sea, to make landfall on the Italian coast. After 15 days Rackete decided to approach the Italian island of Lampedusa and was arrested. This caused an international conflict with Germany and a series of trials followed. This post refers to one of these trials. Salvini has always been very critical of Rackete's activities and decisions, so the adjective 'nice' is clearly ironic. The presence of the emoji reinforces the sense of irony.

The tweet in (17), also by Salvini, contains praise of the former government led by Giuseppe Conte. The interpretation of Salvini's text requires a good knowledge of recent Italian history. In May 2018 Conte was appointed prime minister due to an agreement between Salvini (Lega) and Luigi Di Maio (M5S), but this government failed because Salvini broke up the coalition and Conte resigned in August 2019. In September 2019 a new government was formed thanks to a coalition between M5S and Partito Democratico, and Conte was again appointed prime minister. Salvini was not part of this new government, which explains the resentment evident in his post. This post is a comment on a news article reporting that 57.2 per cent of Italians wanted Conte to resign,

featuring a malicious picture of both Di Maio, the new Minister of Foreign Affairs, and Conte. So, the exclamation 'what a surprise' means 'it is not a surprise', and the claim that 'they are so capable and they're providing such a clear show of efficiency, unity and dignity' must be understood as meaning the opposite.

(16) 'While millions of Italians live among difficulties, uncertainties and fear, for some the most important thing is to prepare other trials against me. I am longing to meet the nice German rammer.' (@matteosalvinimi, 17 January 2021)

(17) 'Look, what a surprise. And yet they are so capable and they're providing such a clear picture of efficiency, unity and dignity.' (@matteosalvinimi, 22 January 2021)

Ma dai, che sorpresa. Eppure sono così bravi e stanno dando una così bella prova di efficienza, unità e dignità...!

Finally, we will discuss one further rhetorical figure: parallelism. Insults are reinforced when they are inserted into a parallel structure—that is, when two syntactic structures follow a similar pattern. Parallelisms and dichotomies are very useful in creating contrasts between different situations, for example comparing what happens in different places, as in (18). They do not constitute an insult by themselves but reinforce a textual construction and enhance the insulting potential of a post.

In (18) Giorgia Meloni compares what happens in Spain and in Italy regarding immigration policy, a warhorse issue for her party, and this comparison is followed by the lexical element *buonisti* ('do-gooders') intended to insult left-wing parties accused of not taking the right measures to stop immigration. The pattern is: adverbial complement subject + verb + object. While the adverbial

complements are introduced by the same preposition and the subject is the same in the two clauses (in fact, it is elliptical in the second), the two verbs are in a relation of contextual antonymy ('protect' vs 'open wide'). In (19) the lexical insult (*schifosi*, 'disgusting') precedes the parallel structure in which Salvini expresses his wish for English football fans to be locked up.

(18) '**In Spain**, left-wing parties **protect their country's borders**. **In Italy,** they **open the ports wide** to illegal immigration. To protect one's country's borders is a duty, but it's hard to understand for the do-gooders' (@GiorgiaMeloni, 15 June 2021)

Giorgia Meloni
@GiorgiaMeloni

In Spagna, la sinistra difende i propri confini. In Italia, lascia i porti spalancati all'immigrazione clandestina. Difendere le proprie frontiere è un dovere, ma i buonisti a tutti i costi faticano a comprenderlo

(19) ‘These English men (disgusting, not fans) have taken defeat well. Instead of kneeling on the pitch, I hope they will kneel in a jail cell.’ (@matteosalvinimi, 12 June 2021)

Salvini’s post is a good example of how the textual construction of an insult takes advantage of different mechanisms: lexical, using the insulting epithet ‘disgusting’, and rhetorical, with two different figures, one formal (parallelism) and one semantic (irony). He plays even with the rhyme *schifosi, non tifosi*.

To sum up, the written nature of these texts, no matter how often a politician writes them and how quickly he/she is supposed to react to the latest news, allows for a minimum of discourse planning. This explains the complex and accurate structure of some of the posts and the use of rhetorical figures as insulting strategies. There seems to be a selection of rhetorical figures with a clear preference for metaphors, hyperboles, irony, and parallelism, as shown in Figure 3.5.

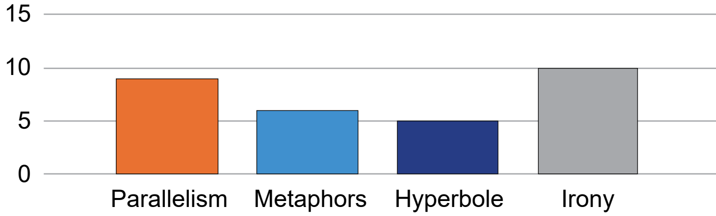


Figure 3.5: Percentages of rhetorical figures as insulting strategies.

3.5.3 The role of cultural referents

We will complete our analysis by briefly discussing the presence of cultural referents as part of insulting strategies. The number of tweets containing these references in our corpus is very low, but it is nonetheless an interesting strategy, aimed at a different type of audience. In these tweets it is possible to identify a hypotext—that is, a text that is referred to by the post and that the audience should be able to identify (Palermo 2013). This is the case for the title of Pirandello's novel *Uno, nessuno, centomila* (1926) in (20), to refer to the changing political support that Conte seeks for different parliamentary votes; and for the quotation from Feuerbach's essay *Die Naturwissenschaft und die Revolution* (The science of nature and the revolution), 'Der Mensch ist, was er isst' (the man is what he eats) in (21). Nicola Morra accompanies his tweet with a picture of Salvini eating a hamburger, taken from Salvini's own timeline (Salvini is well known for sharing moments from his private life, which is a communicative strategy avoided by other politicians) with the clear intention of denigrating him.

The question remains: to what extent are readers able to identify the hypotext and to complete the intentional message hidden by these references?

(20) 'Conte is to me one, no one, a hundred thousand. [...]' (@DSantanche, 4 January 2020)

(21) 'Given that man is what he eats, if the man eats badly, he lives badly.' (@NicolaMorra63, 22 July 2020)

The use of cultural referents in political communication deserves a more in-depth analysis in further studies to assess whether it is a cross-linguistic characteristic of this type of political discourse or is restricted to some cultures. Besides, the audience addressed in this type of tweet does not seem to be the average citizen, to the extent that these texts presuppose a certain knowledge of literary, philosophical, musical, and cinematographical referents, among others. A further question then regards whether we are facing a communicative strategy in political discourse that distinguishes different types of audiences in SNs and privileges a type of hate speech based mainly on discursive constructions and not so on the use of negatively connotated lexical items.

3.6 Conclusions

Our research on insults in political discourse has tried to establish a minimal taxonomy of insults in a particular context, SNs—and, more precisely, Twitter—in order to contribute to the analysis of hate speech both in contemporary political discourse and in computer-mediated communication. We conceive of insults as speech acts—following previous research by Canobbio (2010), Domaneschi (2020), Palermo (2020), Bianchi (2021), Nitti (2021)—and our definition is based on prototypical insults uttered with intention of causing offence to the addressee or the person addressed in discourse. However, the consideration of insult as speech act may be controversial from some perspectives,[20] first because insults may respond to different communicative intentions (including the intention to reinforce social ties among interlocutors) and have thus different illocutionary forces. Another argument put forward by scholars critical of this conception is that the perlocutionary effects of insults are variegated and highly dependent on the cultural and social context. We agree with this view on the complex pragmatic nature of insults, but our study is limited to one specific

20 I am grateful to one of the anonymous reviewers for drawing my attention to this point.

type of insult with a clear illocutionary force and communicative intention. As we have hoped to show, insults in digital political discourse are aimed both at the political opponent and at the audience, and have a double objective: to belittle and humiliate the rival and, by doing so, to persuade the audience of potential voters of the speaker's (or speaker's party's) superiority as a candidate for a political position. The speaker, as in most speech acts, does not have control over the perlocutionary effects of their acts—that is, the effective reactions and responses of the opponent and the audience, that may in effect be very variegated, and therefore not central to the definition proposed in this chapter.

Therefore, insulting mechanisms in political discourse on SNs differ from insults in other contexts, for example daily interactions or street fighting. From the data that we have analysed, it emerges that although lexical insults represent 89 per cent of the insulting mechanisms in our corpus, slurs and derogatory insults, which have been the focus of so much research, are limited mainly to those relating to political orientation (6 per cent). This is for two main reasons: a) slurs directed at minority groups are carefully avoided because they can have a direct effect on the number and type of potential voters—however, some political parties use slurs to denigrate groups who are not allowed to vote (illegal immigrants or foreign citizens); and b) slurs are easily detected by automatic filters in SNs, at least in the most widespread languages (not only English, but also Spanish, Italian, French, and German among the European languages), which can lead to the deletion of the post.

On the other hand, epithets with a negatively connotated value are as frequent as nouns and verbs with the same axiological character. These elements, which are the most frequent in our corpus (83 per cent), are used to discredit the actions, words, and attitudes of opponents and rivals, and belong mostly to the semantic field of criminal activities or unethical acts.

As we have seen, politicians try to create an impact on the audience by creating complex textual constructions (in just 280 characters). Hence, several strategies are often combined to

produce insults of a more textual than lexical nature: negatively connotated terms belonging to different grammatical categories are embedded within rhetorical figures. Moreover, politicians rely on citizens' previous encyclopaedic knowledge and their capacity to draw inferences from ironic texts and cultural referents.

These texts therefore involve thorough discourse planning that contradicts the idea that insults are emotional, uncontrollable reactions in a moment of anger or rage. The analysed tweets show a careful lexical selection, revealing a conscious construction of a well-defined identity and a discursive strategy—a way to take part in political life, to attract potential voters, to construct one's public persona and to damage rivals. Insults constitute a powerful tool in current political debate that has emerged from the convergence of populism as a rhetorical style that contaminates every political party with SNs as the main channel for the dissemination of political propaganda (Gil de Zúñiga, Michalska, and Römmele 2020: 587–588).

There are, however, some factors that limit the validity of our study: first of all, our corpus is very small and therefore the results may be biased by the selection of the tweets in terms of their quantity but also of their authorship and the unequal distribution between politicians of different ideological orientations. Further studies based on larger corpora of digital texts produced by actors of the political sphere will assess the validity of these results. Besides, it would be interesting to explore whether there is a link between the different types of insults and the ideology of the speakers—that is, whether discursive insults are a strategy that is characteristic of left- or right-wing politicians. On the other hand, the taxonomy of insults may be enlarged or modified when contrasted with larger corpora. Not only the percentages may vary but studies on other languages may show that politicians from different cultural and linguistic backgrounds may differ in the frequency and type of insults employed in computer-mediated-communication. This study intends to be a first step in the establishment of a taxonomy of insults taking as main criteria their linguistic nature (lexical or discursive).

References

Alfonzetti, Giovanna. 2009. 'Gli insulti: alcuni criteri di categorizzazione'. In *Studi linguistici in memoria di Giovanni Tropea*, edited by Salvatore Carmelo Trovato, 67–78. Alessandria: Edizioni dell'Orso.

Alfonzetti, Giovanna, and Margherita Spampinato Beretta. 2010. 'L'arte dell'insulto o il rispondere per le rime'. In *Actes du XXVe Congrès International de Linguistique et de Philologie Romanes*, vol. 5, edited by Maria Iliescu, Heidi M. Siller-Runggaldier, and Paul Danler, 3–11. Berlin: De Gruyter.

Andrino, Borja, and Jordi Pérez Colomé. 2021. 'Los insultos en Twitter: una práctica común que crece con la campaña de Madrid', *El País*, 3 May 2021.

Antonelli, Giuseppe. 2017. *Volgare eloquenza: come le parole hanno paralizzato la politica*. Bari and Rome: Laterza.

Austin, John L. 1962. *How to Do Things with Words.* Oxford: Oxford University Press.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.

Borreguero Zuloaga, Margarita. 2018. 'Los encapsuladores anafóricos: una propuesta de clasificación'. *Caplletra* 64: 179–203.

Borreguero Zuloaga, Margarita. 2006. 'Naturaleza y función de los encapsuladores en los textos informativamente densos (la noticia periodística)'. *Cuadernos de Filología Italiana*, 13: 73–95.

Brocca, Nicola, Davide Garassino, and Viviana Masia. 2016. 'Politici nella rete o nella rete dei politici? L'implicito nella comunicazione politica italiana su Twitter'. *iPhiN*-Beiheft 11: 66–79.

Brown, Penelope, and Stephen Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Bule, Guise. 2020. 'Hyperbole in politics: forget what you know and prepare to get emotional'. *All the News Now*, last updated 25 October 2020. Accessed 2 July 2022. https://atnnow.com/hyperbole-in-politics-forget-what-you-know-and-prepare-to-get-emotional/.

Cavazza, Nicoletta, and Margherita Guidetti. 2014. 'Swearing in political discourse. why vulgarity works'. *Journal of Language and Social Psychology* 33 (5): 537–547. https://doi.org/10.1177/0261927X14533198.

Canobbio, Sabina. 2010. 'Insulti'. In *Enciclopedia dell'Italiano Treccani*, edited by Rafaelle Simone. Rome: Treccani. https://www.treccani.it/enciclopedia/insulti_(Enciclopedia-dell'Italiano)/.

Castro Seixas, Eunice. 2021. 'War metaphors in political communication on COVID-19'. *Frontiers of Sociology* 5: 583680. https://doi.org/10.3389/fsoc.2020.583680.

Cepollaro, Bianca. 2020. *Slurs and Thick Terms: When Languages Encodes Values.* London: Rowman & Littlefield.

Claridge, Claudia. 2011. *Hyperbole in English: A Corpus-based Study on Exaggeration*. Cambridge: Cambridge University Press.

Combei, Claudia Roberta. 2020. '*Parlo come mangio!* Il lessico populista su Twitter'. *CLUB Working Papers in Linguistics* 4: 103–122

Conte, Maria Elisabeth. [1988] 1999a. 'Deissi testuale ed anafora'. In *Condizioni di coerenza. Ricerche di linguistica testuale*, 11–27. Alessandria: Edizioni dell'Orso.

Conte, Maria Elisabeth. [1996] 1999b. 'Anaphoric Encapsulation'. In *Condizioni di coerenza. Ricerche di linguistica testuale*, 107–114. Alessandria: Edizioni dell'Orso.

Culpeper, Jonathan. 1996. 'Towards an anatomy of impoliteness'. *Journal of Pragmatics* 25: 349– 367. https://doi.org/10.1016/0378-2166(95)00014-3.

D'Addio Colosimo, Wanda. 1988. 'Nominali anaforici incapsulatori: un aspetto della coesione lessicale'. In *Dalla parte del ricevente: percezione, comprensione, interpretazione: atti del XIX congresso internazionale di studi della SLI ‑ Società di Linguistica Italiana (Roma, 1985)*, edited by Tullio De Mauro, Stefano Gensini, and Maria Emanuela Piemontese, 143‑151. Rome: Bulzoni.

De Mauro, Tullio. 2016. 'Le parole per ferire'. *Internazionale*, 27 September 2016. https://www.internazionale.it/opinione/tullio-de-mauro/2016/09/27/razzismo-parole-ferire.

Dell'Anna, Maria Vittoria. 2010. *Lingua italiana e politica*. Rome: Carocci.

Dell'Anna, Maria Vittoria. 2009. 'La politica va all'insulto'. *Treccani Magazine*, 24 April 2009. https://www.treccani.it/magazine/lingua_italiana/speciali/eloquio/dellanna.html.

Domaneschi, Filippo. 2020. *Insultare gli altri.* Turin: Einaudi.

Domínguez, Pedro J., and Magdalena Zawislawska. 2006. 'Animal names used as insults and derogation in Polish and Spanish'. *Philologia Hispalensis* 20: 123–156. http://doi.org/10.12795/PH.2006.v20.i02.01.

Faloppa, Federico. 2020. *#ODIO: manuale di resistenza alla violenza delle parole*. Turin: UTET.

Francis, Gillian. 1994. 'Labelling discourse: an aspect for nominal-group lexical cohesion'. In *Advances in Written Text Analysis*, edited by Malcolm Coulthard, 83–101. London: Routledge.

Gallardo Paúls, Beatriz. 2022. *Signos rotos: fracturas de lenguaje en la esfera pública*. Valencia: Tirant Humanidades.

Gallardo Paúls, Beatriz. 2018. *Tiempos de hipérbole: inestabilidad e interferencias en el discurso político.* Valencia: Tirant.

Gil de Zúñiga, Homero, Karolina Koc Michalska, and Andrea Römmele. 2020. 'Populism in the era of Twitter: how social media contextualized new insights into and old phenomenon', *New Media and Society* 22 (4): 585–594. https://doi.org/10.1177/1461444819893978.

González Ruiz, Ramón. 2010. 'Gramática y discurso: nominalización y construcción discursiva en las noticias periodísticas'. In *Estrategias argumentativas en el discurso periodístico*, edited by Concepción Martínez Pasamar, 119–146. Frankfurt: Peter Lang.

González Ruiz, Ramón. 2008. 'Las nominalizaciones como estrategia de manipulación informativa en la noticia periodística: el caso de la anáfora conceptual'. In *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística* (SEL), edited by Inés Olza Moreno, Manuel Casado Velarde, and Ramón González Ruiz, 247–259. Pamplona: Universidad de Navarra.

Izquierdo Alegría, Dámaso, and Ramón González Ruiz. 2013. 'Encapsulación y etiquetas discursivas en el discurso parlamentario: función argumentativa a partir de un corpus paralelo'. *Oralia* 16: 185–219.

Korzen, Iørn. 2016. 'Come riassumere un messaggio politico? Strategie e strutture di incapsulazione anaforica nei discorsi del parlamento europeo'. In *La lingua variabile nei testi: atti del XV Congresso della Società Internazionale di Linguistica e Filologia Italiana*, edited by Giovanna Ruffino and Marina Castiglione, 445–461. Florence: Franco Cesati.

Lala, Letizia. 2010. 'Incapsulatori'. In *Enciclopedia dell'Italiano Treccani*, edited by Rafaelle Simone. Rome: Treccani. https://www.treccani.it/enciclopedia/incapsulatori_%28Enciclopedia-dell%27Italiano%29/.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Llamas Saíz, Carmen. 2010a. 'Argumentación en la noticia periodística: el caso de la anáfora conceptual metafórica'. In *Estrategias argumentativas en el discurso periodístico*, edited by Concepción Martínez Pasamar, 147–170. Frankfurt: Peter Lang.

Llamas Saíz, Carmen. 2010b. 'Interpretación del discurso ajeno: la anáfora conceptual metafórica en la noticia periodística'. *Revista de Investigación Lingüística* 13: 107–126.

Lombardi Vallauri, Edoardo. 2019. *La lingua disonesta: contenuti impliciti e strategie di persuasione*. Bologna: Il Mulino.

López Samaniego, Anna. 2015. 'Etiquetas discursivas, hiperónimos y encapsuladores: una propuesta de clasificación de las relaciones de cohesión referencial'. *RILCE* 31 (2): 435–462. https://doi.org/10.15581/008.31.1551.

Mazzoleni, Gianpietro. 1998. *La comunicazione politica*. Bologna: il Mulino.

Moïse, Claudine. 2006. 'Analyse de la violence verbale: quelque principe méthodologique'. In *Actes des XXVIe journées d'*études *sur la parole* (Dinar, 12–16 June 2006), edited by Association Francophone de la Comunication Parlée, 103–114.

Nitti, Paolo. 2021. *L'insulto: la lingua dello scherzo, la lingua dell'odio*. Florence: Franco Cesati.

Otieno, Raphael Francis, Francis Rew Owino, and Judith Miguda Attyang. 2016. 'Metaphors in political discourse: A review of selected studies'.

Journal of English and Literature 7 (2): 21–26. https://doi.org/10.5897/IJEL2015.0856.

Palermo, Massimo. 2020. 'L'insulto ai tempi dei social media: costanti e innovazioni'. In *Accordi e disaccordi in rete: aspetti linguistici, comunicativi e psicosociali*, edited by Angela Ferrari, Letizia Lala, and Filippo Pecorari, special issue, *Lingue e culture dei media* 4 (2): 2–15. https://doi.org/10.13130/2532-1803/14330.

Palermo, Massimo. 2013. *Linguistica testuale dell'italiano*. Bologna: il Mulino.

Paris, Orlando. 2019. 'Nuovi razzismi: la struttura dei discorsi dell'odio su Facebook'. In Petrilli, *Hate speech*, 77–99.

Petrilli, Raffaela, ed. 2019a. *Hate speech: l'odio nel discorso pubblico*. Rome: Round Robin.

Petrilli, Raffaella. 2019b. 'Il meccanismo dell'odio nel discorso politico'. In Petrilli, *Hate speech*, 41–58.

Pistolesi, Elena. 2020. 'La multiforme vita del troll: provocatore, flamer o dissidente?' *Quaderns d'italià* 25: 83–102. https://doi.org/10.5565/rev/qdi.483.

Scaramella, Carla, ed. 2016. *Discorsi d'odio e social media: criticità, strategie e pratiche d'intervento.* Rome: Arci/Cittalia.

Searle, John. 1969. *Speech Acts. An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Spina, Stefania. 2016. *Fiumi di parole: discorso e grammatica delle conversazioni scritte in Twitter.* Loreto: StreetLib.

Testa, Annamaria. 2020. 'Cambiare la politica cambiando le emozioni', *Internazionale*, 20 January 2020. https://www.internazionale.it/opinione/annamaria-testa/2020/01/20/cambiare-politica-emozioni.

Testa, Annamaria. 2019. 'I discorsi d'odio non vanno sottovalutati'. *Internazionale*, 9 July 2019. https://www.internazionale.it/opinione/annamaria-testa/2019/07/09/discorsi-odio.

Testa, Annamaria. 2018. 'Come funziona il meccanismo del discredito'. *Internazionale*, 25 July 2018. https://www.internazionale.it/opinione/annamaria-testa/2018/07/25/meccanismo-discredito.

Theocaris, Yannis, Pablo Barberá, Zoltán Fazekas, and Sebastian A. Popa. 2020. 'The dynamics of political incivility on Twitter', *SAGE Open* 10 (2): 1–15. https://doi.org/10.1177/2158244020919447.

Torregrossa, Javier, Sergio D'Antonio-Maceiras, Guillermo Villar-Rodríguez, Amir Hussain, Erik Cambria, and David Camacho. 2023. 'A mixed approach for aggressive political discourse analysisi on Twitter'. *Cognitive Computation* 15: 440–465. https://doi.org/10.1007/s12559-022-10048-w.

Van Dijk, Teun A. 2006. 'Discourse and manipulation'. *Discourse and Society* 17 (3): 359–383. https://doi.org/10.1177/0957926506060250.

Zimmerman, Klaus. 2003. ‘Constitución de la identidad y anticortesía verbal entre jóvenes masculinos hablantes de español’. In *Estudios del discurso de cortesía en español. Actas del primer coloquio del programa EDICE. El análisis de la cortesía comunicativa: categorización pragmalingüística de la cultura española*, edited by Diana Bravo, 47–59. Stockholm: Stockholm Universitet.

CHAPTER 4

The words of hate speech

A lexical study of homotransphobia in an Italian Twitter corpus

Elena Sofia Safina
University of Naples Federico II

Abstract

The everyday discursive production of dehumanising representations and stereotypical beliefs regarding the LGBTQIA+ community undermines the self-respect of both individuals and the target group by damaging their social agency and entitlement dimensions. This chapter proposes a quantitative and qualitative analysis of the TWEER corpus, which consists of 5660 Italian tweets on queer topics. The aims of the work are quantifying the presence of hate speech online and describing the main linguistic features that characterise such language in Italian. Quantitative analysis consists in a manual annotation of the corpus based on a fine-grained scheme comprising six labels (Sanguinetti et al. 2018): hate speech, intensity of hate, aggressiveness, offensiveness, irony,

How to cite this book chapter:
Safina, Elena Sofia. 2024. 'The words of hate speech: A lexical study of homotransphobia in an Italian Twitter corpus'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 111–157. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-4.

and stereotype. We found that hate speech covers 13 per cent of the entire corpus, but only 6 per cent of those tweets contained an explicit inciting of hatred, while most hateful tweets were superficially polite, containing dangerous prejudices against LGBTQIA+ individuals. We then propose a lexical-semantic study on a subcorpus which focuses on isolating the most representative meaning clusters in explicitly hateful texts. By analysing each lexical word, we found three main clusters, namely references to politics, health, and ethics, while mentions of sexual identity issues were far rarer, confirming that even explicit hate relies on a heteronormative matrix rather than an impulsive intolerance of certain kinds of sexual orientations or gender identities.

Keywords: corpus analysis, sentiment analysis, hate speech, homotransphobia

4.1 Introduction

From a linguistic point of view, hate speech has been a focus of research particularly in computational linguistics and natural language processing (NLP; Basile et al. 2020). The main focus of the computational studies was first on hate speech detection (Warner and Hirschberg 2012), based on sentiment analysis annotations (Patti, Bosco, and Damiano 2017). More recent studies have positively evaluated detection systems implemented with modern deep learning tools (Chakravarthi et al. 2022), such as fine-tuned large language models like HateBERT (Caselli et al. 2021) and RoBERTa (Nozza 2022), which have demonstrated high performance in detecting homotransphobia in YouTube comments. Many other subtasks have been developed over the years, such as target detection (Silva et al. 2016), author profiling (Mishra and Del Tredici 2018), and automatic user detection (Musto et al. 2019). These aims first required a search for the linguistic indicators of hatred, primarily identified as foul language and explicit incitement to physical violence evidenced by words such as 'hitting', 'eliminating', or 'fighting'.

Although the indicators of offensiveness and aggressiveness are inherently objective measures for categorising hate speech, it is equally true that the co-occurrence of aggressiveness and offensiveness is quite rare in corpora, and offensive language is also a factor of ambiguity for machine performance (Davidson et al. 2017; Malmasi and Zampieri 2017; Pamungkas, Cignarella, and Basile 2018). For instance, offensive language can also occur in harmless texts with an ironic or expressive function, which inevitably leads to confusion in the detection task. Moreover, this type of co-occurrence is common to any content that could be described as hate speech and cannot therefore constitute a distinctive feature of a particular discriminatory discourse, such as homotransphobic speech.

One of the limitations that NLP studies have encountered over the years is the ineffectiveness of generic hate speech detection systems (Chakravarthi et al. 2022; Nozza 2022), since each hate type has its own linguistic peculiarities, especially at the lexical and semantic levels, which may elude recognition by a general classifier. Some studies have therefore looked at the detection of specific types of hate speech in Italian, such as the automatic recognition of misogyny (Attanasio and Pastor 2020; Fersini, Nozza, and Rosso 2020; Muti and Barrón-Cedeño 2020). However, according to more recent studies (Chakravarthi et al. 2022; Nozza 2022; Locatelli, Damo, and Nozza 2023), homotransphobic speech still receives little attention from NLP researchers compared with other types of hate.

Another notable example is the manual annotation of a xenophobic and racist hate speech corpus (Poletto et al. 2017; Sanguinetti et al. 2018) based on a rich and fine-grained annotation model. The authors of these papers proposed a scheme that aimed both to detect hate speech and to describe additional strategies of linguistic expression of hatred, such as aggressiveness, offensiveness, irony, or stereotype.

Hate speech has been studied from a more qualitative perspective in social sciences (Leonard et al. 2022), philosophy of language (Bianchi 2017), and Critical Discourse Analysis (CDA),

with a focus on the structural and semantic features of hate messages (Assimakopoulos, Baider, and Millar 2017). Work within the CDA framework has specifically helped to identify some of the referential strategies and frames evoked in institutional homotransphobic discourse (Reddy 2002), in the press (Mongie 2016), and in computer-mediated communication (Socciarelli 2019). These works have also contributed to highlighting the importance of sociocultural context analysis in hate speech studies.

Qualitative analysis has proved to be essential, even combined with some NLP tasks: an important work by Locatelli, Damo, and Nozza (2023) applied a multimodal process of annotation to a homotransphobic Twitter corpus by integrating hate and topic detection tasks. Starting from a qualitative review of the main themes involved in homotransphobia, the researchers managed not only to quantify the presence of hate speech at the cross-linguistic level but also to create a taxonomy of the most frequent topics for each language.

Lavender and queer linguistics (Liddicoat 2009; Norocel 2011; Peterson 2013) can also be considered benchmarks in the study of homotransphobia, since these approaches analyse the discursive construction of the heteronormative model and consider it the conceptual matrix of homotransphobic discourse (Bucholtz and Hall 2004; Coates 2013), even if the explicit incitement of hate is not the main focus of those works.

Relying on the strands of computational and queer linguistics research, this work addresses two different methodologies and goals. Indeed, Section 4.2 specifically aims at detecting the presence of hate speech in a Twitter corpus; thus, we provide a sentiment analysis based on a six-label annotation scheme to quantify the percentage distribution of different hateful features, such as aggressiveness, offensiveness, irony, and stereotype throughout the corpus.

Since quantitative methodology does not allow for an in-depth study of tweets, Section 4.3 instead looks more closely and qualitatively at the words adopted by haters, interfacing with a very narrow subcorpus. The main goal of Section 4.3 is to isolate the

meanings that differentiate homotransphobic speech from other types of hate, creating a spectrum of types and degrees of hate speech. Therefore, based on the extraction of all lexical words of the subcorpus, this section reviews the main semantic clusters involved in Italian homotransphobia, going beyond a description of slurs and searching for apparently neutral meanings that could foster dangerous prejudices against the LGBTQIA+ community.

4.2 Quantitative computational analysis

4.2.1 Corpus construction and description

This section describes the sentiment analysis of online texts automatically collected from Twitter (now X). As the main goal was to detect hate speech against LGBTQIA+ people, we decided to name the corpus 'TWEER', a portmanteau of 'tweet' and 'queer'.

The corpus was built between June and July 2019, and contains 5660 tweets in Italian about queer topics. The corpus is the outcome of three data-filtering operations on a bigger dataset called TWITA (Basile, Lai, and Sanguinetti 2018), kindly made available for this research by the Computer Science Department at the University of Turin. This larger corpus consisted of 500 million tweets, and had already been used for hate speech detection purposes, particularly for xenophobia and racism-related hate speech (Poletto et al. 2017; Sanguinetti et al. 2018).

For the data filtering, we first selected specific production periods by including only those tweets produced by users in June–July 2018 and March–April 2019. The first period covered the Pride months, and while the second related to the World Congress of Families XIII (WCF), held in Verona on 29–31 March 2019.[1] The choices were therefore based on the hypothesis that a greater concentration of political events could have aroused frequent

1 According to its official website, WCF (2019) is a large, international public event that aims to unite leaders, organisations, and families to affirm, celebrate, and defend the natural family as the only stable foundation of society.

discussion by users on Twitter about homotransphobia, whether in support or against such demonstrations.

We opted for a keyword-based approach, selecting a set of keywords associated with the queer target, such as *gay*, *omosessual**, *lgbt*, *lesbica* (lesbian), *bisex, bisessual** (bisexual), *trans, transessual** (transsexual), *transgender*, *queer*, *gender*, *genderfluid*.[2] Along with these keywords, we selected another set of typical Italian homotransphobic slurs, such as all the terms that correspond to 'faggot' in Italian: *froci, finocchi, culatton*, ricchion*, checc*, succhiacazz*, ciucciacazz*, rottinculo, rotto in culo, pigliánculo, piglia in culo*, and in regional variants such as *caghin** (Sardinia), *bulìccio* (Liguria), *busone* (Emilia-Romagna) *bucaiol** (Tuscany), *garrusu* (Sicily) (see Chapter 5 in this volume for the variable intrinsic offensiveness of these terms). Finally, we added a set of neutral keywords represented by single words or phrases, with and without hashtags, which described the main queer topics on social media, such as #loveislove, #famigliarcobaleno (rainbow family), #unionicivili (civil unions), #wcf, #wcfverona, #congressomondialedellefamiglie (World Congress of Families), #contronatura (unnatural), and #pride.

Because of the huge number of tweets obtained, we used a third random filter to reduce the corpus to 6000 tweets, and after off-topic tweets were removed 5660 tweets were left. The final version of the corpus was manually annotated by the author according to the scheme and guidelines described in the next section.

4.2.2 Annotation scheme: tagset design and issues

The annotation task was completed manually by using a tagset consisting of six labels that had been already used by Sanguinetti et al. (2018) (see also Chapter 5 in this volume for an adaptation of

2 Here and throughout the chapter, an asterisk (*) indicates that we also included inflected and derived forms of the word. For example, *omosessual** includes the singular form *omosessuale* (homosexual) but also the plural *omosessuali* and the derived form *omosessualità* (homosexuality).

this tagset). The annotation model relies on a set of variables that the European Court of Human Rights considers in the analysis of hate speech cases; the model attempts to encompass all those variables in a single coherent framework. The tagset includes, besides a hate speech label, labels for aggressiveness, offensiveness, irony, stereotype, and intensity of hate (Sanguinetti et al. 2018: 2800).

The hate speech tag presents a binary choice of values (yes/no). Confirming the presence of hate speech in the tweet depends on the co-occurrence of two factors: the *target*, thus a reference to the LGBTQIA+ community, and the *action*, meaning the illocutionary force of the utterance, thus the intention of spreading, inciting, promoting, or justifying hatred or violence towards the queer target, or a message that aims to dehumanise, delegitimise, hurt, or intimidate the target (Sanguinetti et al. 2018: 2800).

(1) It's only fair that the government is against lesbians.[3]

If the 'yes' label applies, one of five degrees of intensity must be selected: the degree of intensity can be between 1 and 4, or 0 to indicate the absence of hate speech. Indeed, intensity is the only hate speech-dependent tag, while the other four categories are more descriptive and are independent of each other. The two lower degrees of intensity (1 and 2) describe implicit discrimination, while the higher degrees (3 and 4) describe explicit hatred. Definitions and examples are given in the following.

Degree 1: There is no explicit incitement to violence, but the text negatively depicts the queer target. It could be a derogative judgement against a single person or the whole social minority, designed to promote prejudices or to discredit the target:

(2) Trans people didn't even have the decency, they always parade their obscenity!!

Degree 2: There is still no explicit incitement, but the tweet aims to dehumanise or delegitimise the targeted group by questioning

3 All the examples provided in this section have been created by the author to clarify the meaning of each label of the tagset.

their fundamental rights, which are described as a threat to the in-group's rights:

(3) The government only cares about faggots and their civil unions, who is caring about Italian workers?

Degree 3: There is explicit incitement towards discriminatory or violent acts, but users do not refer to themselves as the direct promoter of the violent actions:

(4) The church should refuse to let these homosexual perverts in on Sundays!!!

Degree 4: There is explicit incitement towards discriminatory or violent acts, and they are promoted by the author in person:

(5) As soon as I find that slutty lesbian, I swear I'll smash her head!!!

Turning to the independent categories, the aggressiveness label has three possible values (weak, strong, or absent) and refers to the user's willingness to be aggressive or violent through the justification of discriminatory acts against the target (weak label, as in (6)) or by promoting violent actions against the target (strong label, as in (7)):

(6) It's normal that a gay couple has been hurt on the street! They were kissing each other!!!!

(7) I want all those faggots out of my neighbourhood!

The offensiveness label could be considered complementary to the previous label as it takes into account the target's, rather than the hater's, point of view. Relying on the same three values, offensiveness focuses on the lexicon employed in the tweet. For example, if the tweet contains a negative representation of the target by means of the expression of negative qualities, it receives the weak label (8), while if the message features highly disparaging lexical items, it receives the strong label (9):

(8) Bisexual people don't exist at all. They're just moody.

(9) Here we go again! Another bastard faggot paedophile in the Church!

While the previous tags are useful in detecting homotransphobic speech, the following tags were added to the tagset by Sanguinetti et al. (2018) to investigate the implicit strategies that may express hateful content. The irony tag, which has only two values ('yes' and 'no'), indicates the presence of any kind of ironic, satirical, or sarcastic expression in the tweet. This linguistic feature is quite important because it could mitigate the spread of hate speech. It is also particularly challenging for automatic hate speech detection systems, because 'sometimes, the presence of figurative language even baffles human annotators. Moreover, external world knowledge is often required in order to infer whether an utterance is ironic' (Pamungkas, Cignarella, and Basile 2018: 204):

(10) Yes dude, gay families exist, and I am a flying unicorn.

Finally, the stereotype label (which can have the value 'yes' or 'no') is associated with all tweets that contain false prejudicial beliefs about the LGBTQIA+ community, which are disseminated in order to justify existing discrimination or to lay the foundations for new discriminatory phenomena:

(11) Children would grow up very bad with two mums or dads.

Based on this annotation scheme, the following section will present a quantitative analysis of the results of the manual annotation task.

4.2.3 Results and discussion

We conducted a quantitative analysis of the distribution and frequency of the labels in the annotated TWEER corpus by using the R statistical tool. What emerges from the distribution of hate

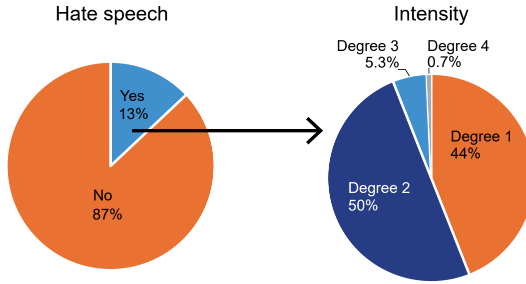


Figure 4.1: Distribution of hate speech and intensity labels in TWEER.

speech labels (Figure 4.1) is that hate speech covers 13 per cent of the entire corpus, or 742 tweets. To better understand this data, it is important to compare the frequency of hate speech with the distribution of the intensity label. Indeed, less than 6 per cent of the tweets labelled as hate speech (44 tweets) explicitly incite violent or discriminatory actions (degrees 3 and 4), while a larger number of tweets convey implicit and mitigated hate speech (degrees 1 and 2).

Turning to aggressiveness and offensiveness, we found two opposing trends: although each category has been tagged in less than 10 per cent of total tweets (Figure 4.2), there are more tweets labelled with weak aggressiveness than tweets labelled strongly aggressive, while tweets with strong offensiveness are more frequent than tweets labelled weakly offensive.

However, given that the offensiveness category is independent of the hate speech category, this unexpected trend of strong

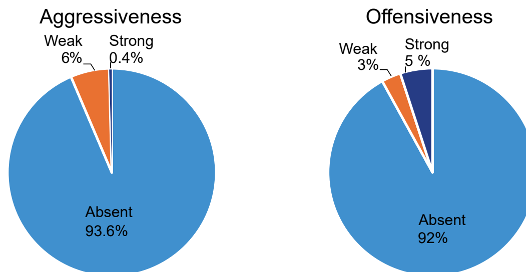


Figure 4.2: Distribution of aggressiveness and offensiveness labels in TWEER.

lexical offensiveness is understandable as a typical linguistic feature of social networks rather than a homotransphobic trait. In fact, 20 per cent of tweets labelled as offensive refer to totally harmless messages (hate speech=No). Rösner and Krämer (2016) describe the absence of a traceable and verified identity on social networks as the cause of the *online disinhibition effect*, the feeling of anonymity and deindividuation that may lead to the extended use of uncivil language; however, offensive words are sometimes used for benign purposes, such as conveying irreverent and ironic meanings, or highlighting the emphasis of the utterance by using insults as filler words.

With regard to the distribution of the irony tag (Figure 4.3), the corpus does not show frequent use of mitigation strategies (only 5 per cent of total tweets). Conversely, stereotype is the most frequent label in TWEER, accounting for 12 per cent of total tweets.

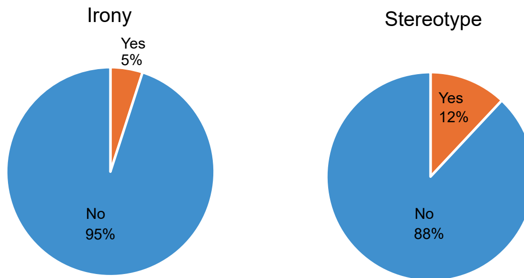


Figure 4.3: Distribution of irony and stereotype labels in TWEER.

In summary, the four labels are distributed in descending order as follows: stereotype (12 per cent of tweets), offensiveness (8 per cent of tweets), aggression (6 per cent of tweets), and irony (4 per cent of tweets). Most hate speech is conveyed in a moderate, mild, and polite form, relying on the effectiveness and immediacy of stereotypes. This is confirmed by the data, given the low component of aggression and offensiveness and the low levels of explicit hate messages (degrees 3 and 4) in the corpus. Similar results are reported by Locatelli, Damo, and Nozza (2023), who found fewer

negative sentiments but more prejudicial homotransphobic views in their dataset made up of 25,000 tweets in Italian.

The quantitative analysis illustrates how the hate speech detection task is difficult and treacherous, partly due to the use of implicit strategies by users, but also due to the significant degree of variability in the linguistic structures that spread hate online, which are difficult to trace by rigid measurement metrics such as those used in this study. The limits of the annotation scheme and of this first quantitative analysis will be further explored in Section 4.4.

In conclusion, the significant distribution of stereotype labels, especially in explicit harmful tweets (hate speech=Yes; intensity degrees 3 and 4) leads us to our next research aim: a more in-depth analysis of stereotypes and the lexical and semantic description of online homotransphobia.

4.3 Lexical and semantic analysis

4.3.1 Corpus construction and description

This section describes a lexical study of a restricted corpus of tweets with the aim of investigating which meanings contribute to the construction of homotransphobic discourse. Starting from the TWEER corpus (see Section 4.2), we collected each lexical word contained in tweets with degrees 3 and 4 under the intensity label in order to avoid any kind of ambiguity in the interpretation of those texts.[4] For this more qualitative analysis, we explored a subcorpus made up of 43 tweets, 38 with intensity degree 3, and 5 with intensity degree 4, ultimately comprising 665 types and 1246 tokens.[5] In terms of lexical words, the subcorpus contains items from the lexical classes of nouns, adjectives, and verbs.

4 By lexical words we mean all words with descriptive-referential, as opposed to purely grammatical, content.

5 By ‘type’ we mean each word of the corpus with descriptive-referential meaning, while ‘tokens’ refers to all the occurrences of each type-word in the corpus.

The research goal of this study is to identify which meanings are 'activated' and 'salient' (Arduini and Fabbri 2013) in homotransphobic discourse, to describe which semantic spheres are involved in stereotypical representations of the queer community, and to detect possible characteristic clusters among lexical words that realise these semantic spheres. As we mentioned in Section 4.1, there are several studies in queer and lavender linguistics that investigate reference strategies and LGBTQIA+ framing (Motschenbacher and Stegu 2013; Motschenbacher 2019). These problem-oriented works focus on how the words used to refer to the target such as 'gay', 'queer', 'transgender', and 'lesbian' are qualified by frequent co-occurrences or collocations (Socciarelli 2019). By analysing frequency, keyness, and collocations, researchers were able to establish not only which words qualified the target but also which semantic frames (Fillmore 1985) are usually employed in referring to that target, intended as sets of words associated with stable cognitive structures.

In his work on the Corpus of Contemporary American English (COCA), Motschenbacher (2018) identified six 'basic sexual usage categories' intended to represent the main clusters of meanings associated with the queer target, namely Identity, Gender, Partner, Relationship, Desire, and Practice. However, what we see from his analysis is that homosexual or transgender targets are often addressed within a political frame rather than within an affective or erotic frame. In fact, the author found more co-occurrences of politics and social identity words than expressions of desire, sexuality, and erotic practices. Only two out of eight target words, 'bisexual' and 'homosexual', concerned desire and sexual practices.

We adopt a similar methodology, but considering every lexical word contained in the tweets, even if it does not qualify queer people in co-text, in order to describe the whole semantic environment in which homotransphobia is observed in terms of argumentations and specific topics.

This study does not therefore use tokens as measuring units, but semantic families that consist of one word and all its inflected and derived forms. For example, all the occurrences of *pedofilo*

(paedophile), *pedofili* (paedophiles), and *pedofilia* (paedophilia) have been grouped into the unit *pedofilo** (paedophile). This measuring unit has the advantage of treating related meanings together, thereby directly individuating core meanings in homotransphobic speech.

Firstly, we use the AntConc software to compile a frequency wordlist. The most frequent lexical words are listed in Table 4.1.[6] The tables presented in this chapter will have a column on the right reporting the semantic family of reference and a column on the left indicating the number of occurrences of all the inflected and derived forms relating to that semantic family. The asterisk (*) indicates the presence of inflected and derived forms related to the semantic family, while a double asterisk (**) signals that we found different spellings of the same unit of meaning due to informal writing or to the hashtag function on Twitter. For example, in the nine occurrences of the unit of meaning 'Matteo Salvini', the presentation of the name varied with respect to the lack of blank spaces between the first and last name ('MatteoSalvini'), lowercase letters ('matteosalvini'), and reference by last name ('Salvini'), but we considered all those forms as the same unit of meaning. The subcorpus contains 222 semantic families.

It is important to highlight that the words *gay, lgbt, famiglie, froci, checche, gender, omosessual**, and *ricchione* in the list had already functioned as keywords for the construction of the TWEER corpus (see Section 4.2.1), which probably explains why their frequency is higher, as evidenced by 'gay' being the most frequent word. We looked at the whole set of words from a problem-oriented, corpus-based approach. Corpus-based studies typically analyse corpus data to validate, refute, or refine a hypothesis, while the corpus-driven approach claims that the corpus itself should be considered the source of the hypothesis about language

6 Because of the very narrow dimensions of the subcorpus, we provide a list of every word that occurred more than once in the dataset. Thus, in this qualitative analysis, the frequency criterion only serves as an auxiliary descriptive data point.

(McEnery and Hardie 2012: 5–6). More precisely, the problem-oriented element in the corpus-based approach indicates that the analysis of language features is informed by critical knowledge about a specific social problem—in our case, homotransphobia (Motschenbacher 2019).

With this in mind, we derive four main clusters of meanings typically associated with homotransphobic hate (tables 4.2, 4.3, 4.4, and 4.5), and two more clusters containing general offensive and violent language (tables 4.6 and 4.7). These clusters can be interpreted as abstract models that are useful in schematising homotransphobic discourse.

Table 4.1: Most frequent words in the TWEER subcorpus.

Frequency	Semantic family
25	gay*
9	matteo salvini** (Italian Politician)
8	lgbt
7	bambini* (children); inesistenti* (non-existent); legge (law)
6	froci* (faggots); italiani* (Italians); roma (Rome); virginia raggi** (Italian politician)
5	famiglie* (families); natura (nature); zingari (gypsies); sessuali* (sexual)
4	gender*; pedofilo* (paedophiles); uomo (man)
3	africani* (African people); delinquere (to commit a crime); diritti (rights); donna (woman); fermate [imperative mood] (stop); ius (Latin = right); lobby; negri (niggers); stop; vivere (to live)
2	basta [hortative] (enough!); checche (faggots); civili (civil); coglione (asshole); depravata (depraved); eros; etici* (ethical); farmaco (drug, medication); fontana (Italian politician Lorenzo Fontana); governolega (Lega Government); inutile (useless); immorali* (immoral); liberando (dalla cacca) (to break free from the shit); madre* (mother); merda (shit); movimento ([political] movement); nazioni (nations); palle [informal] (balls, testicles); popoli* (people/folk); razza (race); repubblica (republic); ricchione (faggot); rottoinculo* ([literal] broken ass, wreck); spazziamo (via) (let's sweep away)

First cluster: Politics

Confirming the findings of Motschenbacher's (2018) study, mentioned above, Politics is the most significant cluster, as references to the political discourse were the most frequent in the subcorpus (36 per cent of total semantic families). As we see from Table 4.2, the politics cluster comprises all those words that describe Italian and international political topics and major figures: first, we find references to leading Italian politicians, especially from the right, such as Francesco Storace, Lorenzo Fontana, and Matteo Salvini, the most cited referent in the subcorpus, since he was in government during the data-collection phase, representing his political party Lega Nord, addressed as 'governolega' in the corpus.

Predictably enough, the main figures from the left are the recipients of hate content, such as in (14), while those from the right are considered both as role models and as reliable political partners, as in (12) and (13):[7]

(12) SEMPRE PIU' PUTIN! COSI' SI FA. Mondiali, i cosacchi controlleranno i gay: 'Effusioni in pubblico segnalate alla polizia'.

'MORE AND MORE PUTIN! THIS IS WHAT YOU DO. World Cup, the Cossacks will control the gays: "Public displays of affection reported to the police"'.

(13) Quando salvini caccerà gli extracomunitari e i gay, in italia non ci sarà più delinquenza e ci sarà lavoro per tutti gli italiani.

'As soon as Salvini expels non-EU citizens and gays, there will be no more crime in Italy and there will be jobs for all Italians.'

7 All the following examples are from the TWEER corpus.

(14) @virginiaraggi @Roma Un privato può affittare la casa a chi vuole. Chi discrimina in realtà è lei. E la smetta di sculettare a favore di gay e immigrati.

'@virginiaraggi @Roma A private individual can rent the house to whoever he wants. The one who's actually discriminating is you. And stop strutting around in favour of gays and immigrants.'

We also find references to female left-wing personalities, including Monica Cirinnà, a senator of the Italian Republic and representative of the political campaign for civil rights for the LGBTQIA+ community, and Virginia Raggi, the mayor of Rome in 2019. The tweet referring to Raggi in (14) displays not only homotransphobic discrimination but also sexist vilification of her professional role. Along with Italian politicians, we also found references to powerful groups, such as Bilderberg (*bidelberg*** in TWEER), and to famous journalists, such as Bianca Berlinguer and Ezio Mauro. There were three further subclusters relating to politics:

a) Law and criminality: This category is particularly interesting because, alongside neutral institutional terms such as *governo* (government), *ministro* (minister), and *repubblica* (republic), homotransphobia occurs in a criminal frame, such as in (15), with terms with negative meanings such as *delinquere* (to commit a crime), *criminale* (criminal), *mafioso* (mafia man), *sentenza* (conviction), *polizia* (police).

(15) @RadioSpada Tu sei un coglione non un ministro, lasciamo vivere i bambini come madre natura comanda e sterminiamo la #LGBT sporca e maledetta assicurazioni a delinquere di stampo mafioso!!![8]

'@RadioSpada You're an asshole, not a minister, let the children live as Mother Nature intends and exterminate the dirty and cursed #LGBT, criminal mafia conspiracy.'

8 In the example in (15), *assicurazioni* (insurances, assurances) is a malapropism for *associazioni* (associations).

b) Freedom and fundamental rights: As is usual in the argumentative strategies of many in-groups (Van Dijk 2004), especially in highly polarised societies, homotransphobic speech portrays the perpetrators of hate as the victims by depicting the in-group as the target of discrimination through the use of words such as *discrimina** (discrimination), *odiano* (they hate), and *razzista* (racist). From the perspective of the in-group, the use of expressions such as *ideologia* (ideology), *cambiare* (to change), *controlleranno* (they will control), *libertà* (freedom), *ribellarci* (to rise up/rebel against something), and *dittatura* (dictatorship) is a specific argumentative strategy intended to frame straight and cisgender people as being subordinate to a superior out-group power, namely the 'LGBT lobby' (16).

(16) Eliminare e cancellare le leggi gayste e il gaysmo di stato subito.

'You must remove pro-gay laws and the National Gay cult, now.'

c) Geography and immigration: The final subcluster is the most relevant, because it demonstrates typical hate speech behaviour, namely the assimilation strategy (Van Dijk 2004). Assimilation has been described as a rhetorical strategy that aims to dehumanise and objectify specific groups of people, such as social minorities. Along with the construction of a social dichotomy through the use of deictic 'us', sometimes replaced by generalisations such as the 'people' or the 'Nation', assimilation tends to depict the out-group, 'them', using figures or demonstratives, thus erasing the minority's human traits or cultural peculiarities (Orrù 2017: 35). In our case, alongside the individual dehumanisation of the queer minority, we find *assimilation of minorities*, since the hateful message addresses a large and heterogeneous group of people, which is perceived and represented as a compact group with no internal differences. We found this assimilation of minorities in 13 out of 43 tweets. In one case (17), the author represented the enemy of the in-group through the juxtaposition of ethnic,

religious, political, sexual, and even professional groups, talking about a 'long list' of hated subjects with 'no distinctions' needed.

(17) Mi viene un rutto ogniqualvolta leggo Salvini si/no/ma. Salvini sta cambiando il vento che aveva già affondato l'Italia. Non è ora di 'distinguo'. Spazziamo via zecche rosse, froci/e, giornalai superpagati, toghe indegne. Lista lunga. Poi vediamo. Chi dubita è dall'altra parte.

'I'm going to burp every single time I read Salvini yes/no/maybe. Salvini is changing the wind that had already sunk Italy. No more time to 'distinguish'. Let's sweep away red ticks ([figurative] communists), fags, overpaid newsagents ([ironic-derogative] journalists), unworthy robes ([figurative] magistracy). Long list. Then we'll see. Anyone with doubts is on the other side.'

Of the assimilated minorities, the group with immigrant status is referred to most often. As we can see from Table 4.2, there are several references to the migration frame, such as *immigrati* (immigrants), *extracomunitari* (non-EU citizens), *profughi* (asylum seekers), and *rifugiati* (refugees). Moreover, we find specific minorities addressed by their geographic or ethnic origins, *africani* (Africans) and *zingari* (gypsies), by religion, *musulmani* (Muslims), *ebrei* (Jewish), and also by the racist slur *negri* (niggers).

We can easily contextualise these frequent references to immigrants in Italian political debate since the early 2010s (Orrù 2017), where the representation of migrants arriving by sea as a *wave of invasion* has gained significant ground in the media agenda. In addition to this, by being deprived of a specific identity, immigrants and sexual minorities are merely characterised by *otherness*. Thus, just as immigration is believed to lead to an inevitable drift towards a loss of security and national identity, homosexual orientation is held to result in social disorder and the dissolution of traditional values.

Table 4.2: List of words in the Politics cluster.

Frequency	Semantic family
Politicians and journalists	
9	matteo salvini** (Italian politician)
5	virginia raggi** (Italian politician)
2	fontana (Italian politician Lorenzo Fontana); governolega (Lega Nord government)
1	cirinnà (Italian senator Monica Cirinnà); eziomauro (Italian journalist Ezio Mauro); luigidimaio (Italian politician Luigi Di Maio); berlinguer (Italian journalist Bianca Berlinguer); piddini (Democratic Party voters); putin (Russian president Vladimir Putin); storace (Italian politician Francesco Storace); bidelberg** (Bilderberg group)
Law and criminality	
7	legge* (law)
3	delinquere* (to commit a crime); diritti (rights); ius (+ iusgenus) (Latin = right); lobby*
2	civili (collocate of 'unioni') (civil unions); movimento (movement); repubblica (republic)
1	condanna (criminal sentence); contratto (contract); criminale (criminal); firma (signature); governo (government); illegale (illegal); leader; lista (list); mafioso (mafia man); ministro (minister); rubacchiavano (they were sneaking); segretario (secretary); toghe (robes – figurative use for magistracy); vietati (forbidden)
Freedom and fundamental rights	
1	agevolazioni (benefits); cambia* ([someone/something] changes); comandare (to command); contrapporre (to counterpose); controlleranno (they will control); discrimina* ([someone/something] discriminates); diffondere (to spread); dittatura (dictatorship); giornalai (newsagents); ideologia (ideology); imporre (to impose); lavoro (job); libertà (freedom); lotta ([political] struggle); mantenere (to maintain); odiano (they hate); parere (opinion); perbenismo (self-righteousness); prevaricato (overlooked); privato (private); razzista (racist); ribellarci (to rise up/rebel against); scandalizzato (shocked); segnalate [imperative mood] (report them!); sinistra (left-wing); sostenuto (endorsed)

Fre-quency	Semantic family
Geography and immigration	
6	italiani* (Italian people); roma (Rome)
3	africani* (African people); negri (niggers); zingari (gypsies)
2	nazioni (nations); popoli* (people/folk)
1	ebrei (Jewish); extracomunitari (non-EU citizens); immigrati (immigrants); invasione (invasion); islamici (Islamic); malta (Malta); musulmani (Muslims); olandesi (Dutch people); paese (country); provincia (Italian regional district); rifugiati (refugees); straniera (foreign); taranto (Taranto, city in Apulia); toscani (Tuscan people); venezia (Venice)

Second cluster: Nature

The second cluster accounts for 16 per cent of the lexical words in the subcorpus, namely those words related to what we consider as 'human experience' (Table 4.3). This cluster is generically named 'Nature' since it includes not only human-referring terms but also references to the animal world and abstract concepts about life and experience. We found two subclusters within the 'Nature' group, used to represent two typical attitudes in general and homotransphobic hate speech. First, we find many occurrences of words that refer to animals being used to dehumanise the hated target, a strategy also noticeable in other types of hate speech (De Mauro 2016; De Smedt et al. 2018). Second, we see the old-fashioned belief that homosexuality is a medical condition that modifies physical and mental traits in human beings or produces deviant social behaviours.

a) Human beings and animals: This subcluster concerns the *anthropological homophobia* paradigm (Rossi Barilli 1999) which has spread through the Western world since the twentieth century, and is based particularly on the dichotomy between the civilised, ordered, sober, new bourgeois class with their moral concerns, and brutal, savage, exotic societies. According to this paradigm, homosexuality afflicted savage, poor, and uneducated individuals, those unable to escape their tribal impulses. In our case,

the animal frame served as a dehumanising device to associate homosexual identities with the sexual practices of wild animals (Locatelli, Damo, and Nozza 2023), by metonymically reducing their sexual orientation to 'savage' anal penetration, as in (18), for example. The anthropological category of wildness also deprives homosexuals of their entitlement to fundamental cultural rituals, such as having a proper burial, as in (19).

(18) @bravimabasta @USERNAME Il tuo culo lo hai da tempo regalato al primo mandrillo che hai trovato, almeno taci, ricchione di merda! […]

'@bravimabasta @USERNAME You gave your ass long ago to the first lecher (lit. mandrill) you found, at least shut up, you bloody faggot! […]'

(19) […] I froci nei cimiteri sono vietati. Li diano in pasto agli squali, almeno servono a qualcosa alla fine della loro inutile esistenza.

'[…] Fags in cemeteries are prohibited. Feed them to the sharks, at least they'll serve some purpose at the end of their useless existence.'

b) Health and disease: In this subcluster, we find body-related terms, such as *sangue* (blood) or *scalpo* (scalp); words referring to different conditions, ranging from the hypernyms *disturbi* (disorders), and *patologie* (pathologies), to the hyponyms *anoressia* (anorexia), *psicosi* (psychosis), *ansia* (anxiety), and *disforia* (dysphoria), up to the extreme and negatively connotated *impazzire* (going crazy) or *pazzi* (fools). We also found some references to neurodivergence, with terms such as *autismo* (autism) or *ritardi mentali* (mental retardation). It is important to underline that the most frequent unit of meaning in the Nature cluster is *pedofilo** (including paedophile[s] and paedophilia), confirming the violent and ongoing stigma of purported paedophilic tendencies in homosexual subjects. This subcluster can be interpreted in the nineteenth-century *clinical homophobia* paradigm (Rossi Barilli

1999), which, as we see from the data, continues to foster hate speech in more recent times. That paradigm considered reproduction of the species to be the cornerstone of normal human psychophysical health. In the name of Darwinism, homosexuality was considered a psychopathology detectable at an objective scientific level. The striking success of Freudian psychoanalysis reignited the homophobic current in scientific thought by shifting the representation of homosexuality from a physical pathology to simple deviance in the process of individual sexual development. The extensive list of works that sought to medicalise homosexuality came to an end only in 1973, when the American Psychiatric Association, and subsequently also the World Health Organization removed homosexuality from the manual of psychopathologies.[9]

It is therefore unsurprising that biological and clinical words appear with a remarkable frequency in our study, not only as a reference strategy to qualify the queer target but also as a profitable setting for homotransphobic speech.

Finally, we highlight the dangerous potential of words such as *pulizia* (cleaning) and *(fare) igiene* (to clean up/sweep away), which have been resemanticised with a hateful connotation and have historically been employed in genocide storytelling, such as in the case of the Rwandan genocide in 1994 (Gagliardone, Patel, and Pohjonen 2014), see (20).

(20) Le famiglie gay non esistono. Bene, cominciamo a fare un po' di igiene. #Fontana.

'Gay families don't exist. Great, let's start a clean-up. #Fontana.'

9 Due to the World Health Organization's clinical distinction between ego-syntonic and ego-dystonic homosexuality, we see the definitive removal in 1990 and the actual application in 1994. Ego-syntonic homosexuality refers to the condition whereby the person lives and accepts their homosexuality with serenity. In contrast, ego-dystonic homosexuality is the homosexual's feelings of rejection and suffering toward their own condition. (Istituto A. T. Beck, n.d.).

These words could be more harmful than the others in the subcluster since terms connected to cleaning serve as euphemistic reformulations of words for extermination, and in addition cannot be detected by automatic systems. Nevertheless, the non-explicit harmful connotation of those meanings allows the hateful content to spread freely, and contributes to what has been called the 'trivialization of evil', a process that can lead to a

> hypertrophy of the insensibility to evil, which means there appears a systematic substitution of the good for the worse and the worse for the bad … Such a substitution is so easy because of the assistance of language which trivializes evil, for example, the word 'to kill' is replaced by the expression 'to cause death out of compassion'. (Drożdż 2016: 7)

Table 4.3: List of words in the Nature cluster.

Frequency	Semantic family
Human beings and animals	
4	uomo (man)
3	donna (woman); vivere* (to live)
2	razza (race)
1	*pescie [pesce] (fish); animali (animals); diventare (to become); esistenza (existence); mandrillo (mandrill); nasce (to be born); ragazzo (kid, boy); squali (sharks); umani (humans); zecche (ticks)
Health and disease	
4	pedofilo* (paedophiles)
2	farmaco (drug, medication)
1	aborto (abortion); anoressia (anorexia); ansia (anxiety); autismo (autism); autolesionismo (self-harm); disturbi (disorders); (fare) igiene (to clean up/sweep away); impazzire (to go crazy); patologie (pathologies); pazzi (crazy people); psicosi (psychosis); pubertà (puberty); pulizia (cleaning); ritardi mentali (mental retardation); sangue (blood); scalpo (scalp); suicidio (suicide); vizietto (bad habit) patologie (pathologies)

Third cluster: Values and Customs

Although clinical and anthropological homophobia paradigms are still relevant for the production of hate speech, current homo-transphobia mostly relies on what Queer Theory (Arfini and Lo Iacono 2012) has called the *heteronormative paradigm*. The third cluster (10 per cent of total semantic families) is hence informed by Queer Studies, according to which heteronormative everyday discursive practices contribute to setting the perception of non-heterosexual and non-cisgender identities in an undesirable position of 'otherness'. We therefore identify three subclusters, namely Morality, Family, and Religion (Table 4.4), which function as regulatory devices to distinguish ethical values from unworthy and tribal values, as we see in (21).

(21) Razza malvagia e depravata, priva di valori umani e cristiani. Scompaia pure in fretta. Sono peggio delle lobby lgbt e pro-gender. Disgustosamente amorali e immorali. Veramente vil razza dannata che lotta per distruggere l'uomo e i suoi figli in nome di una falsa libertà. Pazzi.

'Evil and depraved race, devoid of human and Christian values. Go away quickly. They are worse than the LGBT and pro-gender lobbies. Disgustingly amoral and immoral. Truly vile damned race struggling to destroy man and his children in the name of false freedom. Crazy people.'

The 'family' unit of meaning is the most frequent in the third cluster, with five references, and is thus also one of the most frequent in the entire corpus. It is important to note that the term *famiglia** (family, families) frequently co-occurs with the concept of 'non-existence', textually realised by single words like *finte* (fake) or *inesistenti* (non-existent), by verb phrases, such as *non esistono* (they do not exist), or by implicit meanings inherent to heterosexual and traditional families, depicted as the 'only true families', as in (22):

(22) @USERNAME basta con i diritti e agevolazioni a gay e finte famiglie gay…W LA NATURA E STOP AL FINTO PERBENISMO DEI CONTRONATURA…LEGGI A FAVORE DELLA VERA E UNICA FAMIGLIA(uomo e donna)……STOP AI CONTRONATURA….NO assoluto 'll'adozione di bambini a gay e cop[p]ie gay…….

'@USERNAME Enough with the rights and benefits for gay people and fake gay families…LONG LIVE NATURE AND STOP THE FALSE SELF-RIGHTEOUSNESS OF PEOPLE WHO GO AGAINST NATURE…[we want] LAWS IN FAVOUR OF THE TRUE AND ONLY FAMILY (man and woman)……NO MORE PEOPLE WHO GO AGAINST NATURE….Absolute NO to the adoption of children by gays and gay couples……'

Unlike other types of hate speech, a large number of homotransphobic offensive words come from the religious semantic sphere, as we see from the terms *dannata* (damned), *maledetta* (cursed), and *abominio* (abomination), confirming the reference system of values in homotransphobic prejudice.

Table 4.4: List of words in the Values and Customs cluster.

Frequency	Semantic family
Morality	
2	etici (ethical); immorali (immoral)
1	indegne (unworthy); servire (to serve); tribali (tribal); valori (values)
Family	
7	bambini (children)
5	famiglia* (families)
2	madre (mother)
1	adozione (adoption); casa (home); figli (children); sposati (married); tetto (roof)

Fre-quency	Semantic family
Religion	
2	depravata* (depraved) [collocate for 'race']
1	abominio (abomination); dannata (damned); glorificato (glorified); maledetta (cursed); malvagia (wicked); orrore (horror)

Fourth cluster: Affectivity and Sexuality

This cluster represents the last queer-related topic found in hate speech expression in our corpus. As can be seen from the data in Table 4.5, the cluster contains only ten semantic families, representing 5 per cent of the total. In addition, four semantic families include expressions that functioned as keywords in the TWEER construction phase. If we exclude those keywords, affectivity and sexuality words actually account for only 3 per cent of the entire corpus. This is hence not a cluster that naturally emerged from the data, but is the result of our specific search that was designed to detect how many and which words were chosen by haters to talk about the primary LGBTQIA+-related topic. Once again, in accordance with the results found by Motschenbacher (2018) and Locatelli et al. (2023), a very narrow list of words concerns sexual identity in Italian hate speech, specifically *amare* (to love), *co[p]pie* (couples), *effusioni* (displays of affection), *emotività* (emotionality), and *eros** (eros, erotically). Moreover, we found no occurrences of the lexeme *lesbic** (lesbian, lesbians), nor of *trans** (transgender, transgenderism, transexual). The previous quantitative analysis of the TWEER corpus confirms this data: we found 2815 occurrences of *gay*, but only 486 of *lesbic** and 364 of *trans**. Along with frequency scores, even NLP studies reported a clear prevalence of the gay target, rather than lesbian, as the main recipient of derogatory language (Locatelli, Damo, and Nozza 2023). The significant difference in the frequency of occurrence of the word *gay* versus the other two target words is in part due to the use of the word 'gay' as generic, unmarked masculine (Thornton 2016), including male and female referents. There remains,

however, an important difference in frequency between references to non-heterosexual orientations (e.g. 'gay', 'lesbian', 'bisexual') and non-cisgender identities (e.g. 'transgender', 'transsexual', 'genderfluid', 'bi-gender').

Although the narrow dimension of our subcorpus could limit the significance of this result, it seems appropriate to interpret this data in the light of the notion of 'corpus notable absences' (Partington 2014), namely infrequent or absent usage types in the corpus: 'for example, certain grammatical constructions or lexical combinations that are in principle possible but do not or only infrequently occur in a data set [which] may instantiate discourses that are perceived to be marked or non-normative' (Motschenbacher 2018: 11). In our case, the problem-oriented qualitative analysis allowed us to note the importance of 'what gets left out' (Kulick 2005) of the discourse, such as hateful references to lesbians and trans* people.

According to Borrillo (2009), the absence of a term in the corpus may not be due to an intentional selection of hated targets, but to the heterosexist matrix that affects even hateful discursive practices. Borrillo discussed a *regime or hierarchy of sexuality* where the homo/hetero dichotomy represents an exacerbation of the male/female dichotomy, which not only determines the direction of normative sexual desire and practices based on biological predisposition (sex) but also regulates socially expected behaviours in the masculine/feminine binary (gender). According to this view, lesbian identities have become subordinate to gay (male) identities, and trans* identities have been marginalised still further, respecting a precise hierarchical order, even in hate speech production.

Table 4.5: List of words in the Affectivity and Sexuality cluster.

Frequency	Semantic family
25	gay* (keyword in TWEER)
8	lgbt (keyword in TWEER)
5	sesso (sex) (keyword in TWEER)
4	gender (keyword in TWEER)
2	eros*
1	amare (to love); *copie [coppie] (couples); desiderarli (to desire them); effusioni (displays of affection); emotività (emotionality)

Offensive and hate words

Finally, we identify two further clusters: offensive words and hate words. These clusters (20 per cent of total semantic families) were intended to represent some typical linguistic features of hate speech, such as slurs or foul language, but also to provide an account of which negative qualities or images are generally associated with the target of hate. Starting with offensive words (Table 4.6), we firstly see slurs—that is, specific words to harm a certain target, such as 'bottana' (whore), and more terms indicating the meaning 'faggot(s)': *checche, rottoinculo, froci*, ricchione*.

Table 4.6: List of offensive words.

Frequency	Semantic family
Slurs	
6	froci* (faggot[s])
2	checche (faggots); rottoinculo ([literal] broken ass, wreck); ricchione (faggot)
1	bottana (whore)

Frequency	Semantic family
Foul language	
2	coglione* (asshole); merda (shit); palle (balls)
1	cacca (poop); incazzare (to piss off); rutto (burp); sculettare (to shake one's ass)
Negative qualities	
7	inesistenti* (non-existent)
2	inutile (useless)
1	antipatici (obnoxious); arroganza (arrogance); ciecato ([vulgar] blind); cocainomani (cocaine addicts); fastidiosi (annoying); ingordo (greedy); insostenibile (unsustainable); perditempo (time waster); pistolino ([literal] little pistol – penis); rifiuti (waste); schifo (disgust); sporca (dirty); taci (shut up); vil (coward)

It is important to note that Italian dictionaries very often provide numerous alternative terms to designate homosexuality, such as *androfilia* (androphilia), *omofilia* (homophilia), *androgamia* (androgamy), *lesbismo* (lesbianism), *saffismo* (sapphism), *uranismo* (uranism). It is much rarer to find alternatives for the term 'heterosexuality'. This imbalance is clear in the Italian dictionary Treccani (online version): while *eterosessuale* (heterosexual) shows only one alternative, namely the short form *etero*, several lexical alternatives are listed for *omosessuale*. The dictionary first defines 'gay', 'lesbian', and 'bisexual', after which we find a great degree of variation by register, such as the more informal and offensive *frocio* (faggot) or *finocchio* ([literal] fennel, for faggot), and the more formal and obsolete *sodomita* (sodomite) or *pederasta* (pederast), not to mention the countless regional variants (see also Chapter 5 in this volume).[10]

In her discussion of the representation of minorities in language use, the German scholar, activist, and writer Kübra Gümüşay

10 Lexical entry 'eterosessuale' (heterosexual) in the online Italian dictionary 'Sinonimi e contrari' by Treccani, accessed 31 March 2022, https://www.treccani.it/vocabolario/eterosessuale_%28Sinonimi-e-Contrari%29/.

(2021) perfectly describes this specific linguistic imbalance by portraying the language system as a museum. She divides people into two opposing camps: the museum visitors, also called 'named people', who in our case equate to heterosexual or cisgender individuals, and people on display, or 'unnamed people', here the non-heterosexual and non-cisgender individuals. Given this dichotomy, visitors are not referred to by multiple labels for two reasons: first, because they embody the social standard, and second, because nobody would be particularly interested in their (sexual) behaviour or external characteristics. In contrast, the unnamed people on display represent all those *weird identities* that are for some reasons 'deviant' from the social norm, whose behaviour is constantly in the spotlight and needs to be classified by others.

Along with this long-standing social and linguistic division, a wide variety of labels have been used to describe sexual minorities, all of which are informed by the dominant group's heteronormative perspective. The significant number of synonyms for 'homosexual' with a negative connotation range from those indicating an association with child abuse (e.g. 'pederast'), to those denoting pseudoscientific beliefs, such as the reference to the theory of sexual inversion (e.g. *invertito*, 'sexual invert'), alongside reduction to the synecdoche of a sexualised body part (e.g. *rottoinculo*, [literally] 'broken ass, wreck') and connection with recurring sins in biblical tradition (e.g. *sodomita*, 'sodomite').[11]

The rich and varied set of offensive words also includes generic foul language, such as *cacca* (poop), *incazzare* (to piss off), *coglione*

11 The theory of sexual inversion gained success between the 19th and 20th centuries. The theory made no distinction between biological sex and gender identity. In fact, sex inversion was first applied in biology to describe a particular behaviour of fish and amphibians. In sexology, the phenomenon concerned both transsexual and homosexual people, as people who developed 'inverse' sexual desires and behaviours compared to those normally expected based on their biological sex. The term, first proposed in German as 'konträre Sexualempfindung', was also successful in Italian translated as 'invertito', and remained in use as an insult to designate homosexuals (see Wikipedia, n.d.).

(asshole, jerk), and some words describing negative qualities, such as *antipatici* (obnoxious), *perditempo* (time waster), *vil* (coward). Once again, we can observe the process of assimilation of minorities (see section on the first cluster) within the cluster: queer individuals are associated with *cocainomani* (cokeheads) and *ciecati* ([vulgar] blind people) by the juxtaposition of words in the tweet in (23):

(23) Tutti della stessa linea di sangue. Tribali, pedofili, checche sfrenate e cocainomani. Tutti con il vizietto. Vi manderei in Africa compreso il Vostro inutile palazzo a Roma.

'[You are] All of the same bloodline. Tribals, paedophiles, wild faggots, and cokeheads. All with the vice. I would send you to Africa including your useless palace in Rome.'

With regard to hate words (Figure 4.4), we found it more useful to schematise the cluster within the 'distance' conceptual space, instead of merely presenting a list of words as we did previously. Figure 4.4 shows three separate but communicating rectangles of meanings. The first includes all the *boundary* terms, namely all those words that express an uncrossable border, such as *stop*, *basta* ([hortative] enough), *ritiriamo* (let's withdraw), and *fine* ([the] end).

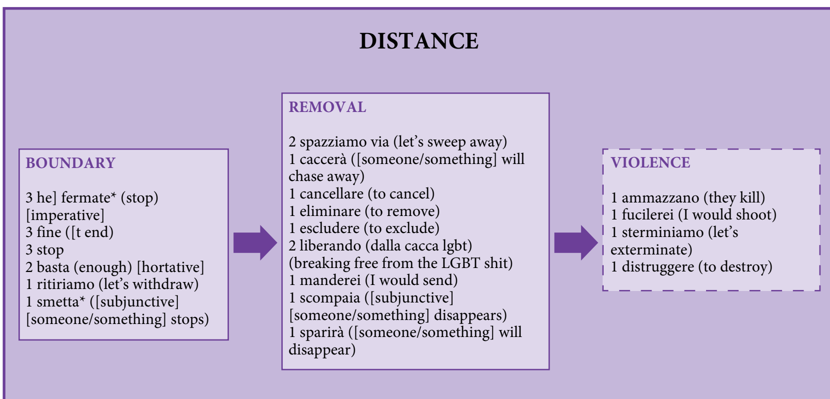


Figure 4.4: Cluster of hate words in the 'distance' conceptual space.

The meaning of 'boundary' presupposes a separation between what is included and preserved, versus what is to be left out and excluded; the same conceptual dichotomy divides in-groups from out-groups in social discursive practices. A division of this sort becomes increasingly dangerous if it has been crystallised in societies, especially in those communities that have experienced holocausts and genocide. Indeed, hate speech plays a significant and instrumental role in the perpetuation of discrimination of minorities (Gagliardone, Patel, and Pohjonen 2014): by making use of a single discursive practice, the in-group succeeds both in breaking down the minority and in reinforcing a sense of identity and community among the perpetrators. The perpetrators therefore reverse the direction of the actual hate speech: by producing fake accounts of an explicit attack waged by the minority against the in-group, haters wield hatred as the last dutiful defence available to them, as a way of defending the borders, the traditions, and even the safety of their own people, thus, transforming the ordinary social opinion gap into a dramatic 'us vs them' polarisation.

The immediate consequence of this dangerous polarisation can be found in the second rectangle of meanings, which is populated by *removal* terms, such as *liberando da* (breaking free from), *cancellare* (to cancel), *spazziamo via* (let's sweep away), *caccerà* ([someone/something] will chase away), *sparirà* ([someone/something] will disappear]. As the data in Figure 4.4 show, the semantic space hosts all the terms that refer to a forced distancing of the out-group by the in-group; in fact, with the exception of the intransitive verbs *sparirà* ([someone/something] will disappear) and *scompaia* ([subjunctive] [someone/something] disappear), all the forms are transitive and imply significant intentionality of the part of the agent of the verb.

The last rectangle of meanings includes terms relating to *true violence*, which could relate to murder or have been historically used in context of genocide, such as *ammazzano* (they kill), *fucilerei* (I would shoot), *sterminiamo* (let's exterminate), *distruggere* (to destroy).

Starting from the left side of Figure 4.4, the first two rectangles are drawn with continuous lines, which is intended to highlight the significant connection between the groups of meanings. The third rectangle, on the other hand, has a dotted line because, as we are still dealing with speech acts, we wanted to emphasise the non-deterministic consequentiality between hate speech and hate crimes (Article 19 2015). In fact, both the social polarisation and the desire for estrangement are highly pervasive traits in hate narratives, whereas explicit incitement to extermination and explicitly violent references, especially those expressed in the first person, are fortunately very rare elements in the TWEER corpus (see Section 4.2.3). If we take the second rectangle of meanings (removal) to be an effective consequence of the first (boundary), the third rectangle (violence) can be understood as a serious and dangerous degeneration of the previous two. With this in mind, we suggest that these linguistic acts should be interpreted in light of the data regarding the distribution of stereotype labels, which accounted for 82 per cent of tweets labelled as hate speech in TWEER (see Section 4.2.3). Hence, although we recognise the significant distance between the cluster of hate words and all the previous clusters in terms of intensity of hatred, the violence rectangle can only be understood as a superficial and explicit manifestation of a dense underlying network of stereotypical narratives.

4.4 Evaluation of methodology and results

The first aim of this research was to quantify the prevalence of homotransphobic speech in an Italian Twitter corpus using the opinion mining and sentiment analysis methodology. This analysis revealed that hate speech was present in 13 per cent of the corpus, mostly conveyed by stereotypes and moderate language.

The analysis of computer-mediated texts proved to be highly effective in this context, as it allowed the analysis of specific groups of meanings in a narrow text space (see Chapter 5 in this volume for similar considerations). Indeed, the binding length of tweets (280 characters in 2019) forced perpetrators of hate to be

more direct and to lay out clear argumentation strategies. At the same time, however, the impossibility of providing arguments in support of an opinion on the social network could have encouraged the production of a high number of stereotypes.

The fine-grained annotation scheme allowed us to describe hate speech behaviour on Twitter in detail, but it did have some drawbacks. First, the fact that the descriptive categories are independent of the actual presence of hate speech has caused considerable data dispersion in this study; to be precise, 20 per cent of the labels for aggressiveness, offensiveness, irony, and stereotype have been attached to harmless tweets. This occurred because the annotators had to assign six different labels to every tweet instead of analysing only harmful messages along with these descriptive features. In addition, two descriptive labels have three available values while the other two have only binary values. This value shift had the potential to cause confusion and misunderstandings during the annotation of individual tweets. The independence of the labels could be useful for general linguistic studies on computer-mediated communication, or for studies looking to explore phenomena adjacent to hate speech, such as verbal aggression (Rösner and Krämer 2016), linguistic triviality, or abusive language (Waseem et al. 2017). Yet, this kind of complexity was challenging for the goal of improving automatic hate speech detection systems. In order to achieve better performance in various detection tasks, the manual annotation should be validated by other expert or non-expert annotators by calculating the score of inter-annotator agreement. This measurement serves to establish whether the subjective opinion of the first annotator can be widely shared, therefore considered objective enough to train a machine. It is usually difficult to get a high agreement score with complex tagsets such as the one described in this chapter. In our study, we measured the inter-annotator agreement by submitting a questionnaire to non-expert annotators on Facebook. The results were far from an acceptable level of agreement: only the hate speech and irony categories showed a substantial level of agreement, while there was significant disagreement with regard

to the other four labels with two or more values; this is especially true for the intensity label, for which we observed a high level of misunderstanding. Even if expert annotators are used, opinion mining and sentiment analysis tasks always carry some degree of uncertainty and ambiguity because of the subjective interpretation and the cultural or linguistic background of the annotator.

It is also important to discuss the other two disambiguation issues involved in the hate speech detection task. Firstly, the scheme considers cases of reported speech as in (24), such as newspaper titles or statements by third parties, inevitably leading to ambiguity in the annotation. In these cases, we chose to ignore the reported hateful content by labelling the tweets as harmless.

(24) Mi fa ridere perché secondo mia mamma i bisessuali sono i peggiori loro vanno con tutti.

'It makes me laugh that according to my mum bisexuals are the worst, they screw ([literal] go with) anyone.'

We then dealt with the problem of offensive tweets that contain homotransphobic slurs but do not display any semantic link with the actual LGBTQIA+ target, as in (25). It means that the author uses a homotransphobic lexicon, but does not address actual homosexual referents.

(25) MACRON finirà di ammazzare il suo popolo, sto gay infame…

'Macron will finish killing his people, this infamous gay….'

This use of derogative words could lead to negative consequences in two ways: firstly, from a poststructuralist point of view (Motschenbacher and Stegu 2013), everyday discursive practices, and particularly idiomatic offensive language (Pinker 2007), could contribute to the social construction of the public image of the LGBTQIA+ target. Even if we do not share the desire to hurt a gay person, we still have to make the inference 'being gay is something that negatively qualifies people' to understand the

meaning of the sentence, thus we share, even in a passive way, the stereotype (Pistolesi 2007).

Secondly, racist and homotransphobic lexical items often co-occur in trivial messages without specific reference to any minority target, leading to false positives in automatic detection systems, as these are not sensitive to the pragmatic meaning of the sentence (Davidson et al. 2017; Malmasi and Zampieri 2017; Pamungkas, Cignarella, and Basile 2018). More broadly, negative sentiments do not necessarily coincide with hate speech: concerning homotransphobia, false positive detections may be due to users' indignation against posted discriminatory episodes or haters' comments. In these cases, it is likely that the system incorrectly recognises hate speech relying on the co-occurrence of users' anger and the queer references in the text (Locatelli, Damo, and Nozza 2023).

Looking at the findings of the lexical analysis (Section 4.3), the Politics, Nature, and Values and Customs clusters may serve as useful starting points to build full glossaries that could enrich the training phase in automatic homotransphobic detection tasks, moving beyond the detection of simple slurs and offensive language.

Specifically, we propose that attention should be paid in particular to those cases of co-occurrences between target words and apparently 'harmless words', such as *igiene* (hygiene) or *pulizia* (cleaning) (see section on the second cluster), especially in the construction '*fare* (to do) + Noun'. These constructions cannot be blocked by automatic systems because of their denotative meanings but still spread highly dangerous messages when they co-occur with a reference to a hated target.

With regard to the limits of the lexical and semantic analysis, we identified a problem in the subcorpus dimension as a result of the original goal of detecting homotransphobia. Because the group of explicit hate speech tweets was small, it was impossible to measure keyness and frequency correlations, such as collocates, which could have helped to create a better picture of the reference target; we thus believe that it would be useful to replicate the lexical and semantic analysis on a larger corpus of tweets.

Based on methodology used in corpus analysis and queer linguistics (Baker et al. 2008; Baker 2016; Motschenbacher 2018, 2019), it could be interesting to distinguish between lexical and semantic representations of single target words (e.g. 'gay', 'bisexual', 'lesbian', 'genderfluid', 'transgender'), with the aim of exploring different collocates and co-occurrences to define sexual orientations or gender identities. It could additionally be useful to apply the same methodology to the study of queer-related noun phrases and hashtags, such as 'rainbow families', 'LGBTQIA+ rights', 'surrogacy', and 'love is love', in order to understand how the hateful storytelling changes in accordance with specific aspects of queer life.

From a sociocultural perspective, our work demonstrated that hate speech production is dependent on the heteronormative social matrix. Indeed, a large portion of the subcorpus referred to the maintenance of the status quo, by denying access to family rights for LGBTQIA+ people, by contesting a possible introduction of a sort of queer literacy in the educational system, or by simply rejecting *unnatural* non-heterosexual behaviours.

We consequently consider a further analysis of homotransphobic speech compared to the *neutral* heteronormative discursive model to a valuable addition. First, implicit hate speech is more widespread than its explicit counterpart, and it often refers to a superior natural or normative order that is presented as the only possible and right one. Second, those prejudices could also be spread within the LGBTQIA+ community through the reproduction of a dangerous top-down hierarchy of sexualities (see section on the fourth cluster).

Queer Studies have widely discussed the concept of 'homonormativity' (Duggan 2002; Hermann-Wilmarth and Ryan 2016; Motschenbacher 2020) by describing the homosexual adoption of heteronormative sociocultural categories in referring to or interacting with other members of the community. That is why studies of this type seek to deconstruct the sexual and gender binary categories in order to reduce discrimination against sexual minorities. Some of the consequences of 'binary thinking'

and homonormativity could be found in the underrepresentation of lesbian and trans* individuals in activism, as mentioned previously, but also in biphobic and transphobic discrimination, which are still common within LGBTQIA+ communities. Both bisexual and trans* people, especially those who identify as non-binary instead of trans men or trans women, are often perceived as a non-existent third way, outside of the binary distinctions of homosexual/straight or male/female.

In line with the frequent references to 'non-existent gay love and families' in our hate speech corpus, the same kind of rejection of 'minorities among minority' can also be found in many other online spaces (e.g. as shown in Table 4.7); further work is therefore needed on abusive language relating to LGBTQIA+ individuals, particularly in the Italian context.

It is important to analyse homonormative speech in order to detect analogies and differences with homophobic speech produced by heterosexual and cisgender people. It would also be interesting, however, to investigate the out-group's response to homophobic speech, both for reproduction of hate speech, such as the public call to violent actions against people who commit homotransphobic acts, and in the reappropriation of slurs by the out-group (see Chapter 2 in this volume). For example, the Italian slur *frocio* is currently used as a pride device by the LGBTQIA+ community, as also happened in late 1980s with the English 'queer' (Perlman 2019), and it has gradually become linguistically productive with more inflected and derivational forms (e.g. *frocia*, *froce*, *frociarola*, *frociaggine*).[12]

12 The slur *frocio* commonly refers to male individuals, while in the out-group usage it could often designate female referents. Moreover, some Italian transfeminist groups often use the feminine form *froce* in a wider sense, to refer to all those individuals perceived as deviant by the social norm, regardless of their sexual orientation, thus regardless the original meaning of 'gay man'.

Table 4.7: Biphobic examples on Facebook page Gay.it

Post on Gay.it public Facebook page	Author's translation
Il bisex è fondamentalmente uno omosex mascherato per esigenze sociali.	The bisexual is basically a homosexual disguised for social needs.
Ma esistono davvero i bisessuali?	But do bisexuals really exist?
Mi piacerebbe esserlo. L'avrei vissuto con grazia: matrimonio con erede e, consensualmente, un bel maschio come amante. Sarebbe stata una vita molto meno incasinata. Però adoro i maschi...	I would like to be. I would have lived it with grace: marriage with an heir and, by consensus, a handsome male as a lover. It would have been a much less messed-up life. But I love males...
Io non credo all'esistenza del bisex.	I don't believe in the existence of bisexuals.
I bisex esistono come patologia... anch'io ero bisex.....ma poi sono guarito e ho scelto il c...o	Bisexuals exist as a pathology...I was also bisexual.....but then I recovered, and I chose the d...k
Praticamente dovremmo festeggiare un omosessuale che si nasconde per ragioni sociali? Ah ok! (Non sono leghista prima che cominciate a sparare cavolate).	Should we basically celebrate a homosexual who is hiding for social reasons? Ah OK! (I'm not a Lega Nord voter, before you start shooting bullshit).
ah ecco la coppia aperta......certo. Caro Davide io vivo in Inghilterra, dove a confronto l'Italia è medioevo. E sarà magari per l'opportunità che abbiamo qui di poter esprimere noi stessi senza che nulla accada, ma di bisessuali, ce ne sono ben pochi eh. Siamo tutti piuttosto convinti qui	ah here is the open couple......sure. Dear Davide, I live in England, which makes Italy seem medieval in comparison. And maybe it will be because of the opportunity we have here to be able to express ourselves without anything happening, but there are very few bisexuals, eh. We are all pretty convinced here

4.5 Conclusion and future developments

In conclusion, our work may represent a starting point for new guidelines against online homotransphobia, and to inform future campaigns for protection laws that encompass sexual minorities, which have not yet taken hold in Italy, after the rejection of the Zan bill by the Italian Parliament in 2021 (De Carli 2021). Indeed, the

amount of homotransphobic texts circulating online is greater in religious and conservative cultural contexts such as Italy or France compared to Germany or Norway (Locatelli, Damo, and Nozza 2023: 20), and the use of overtly abusive language is highly more likely in countries with less LGBTQIA+ safety legislation. Furthermore, the clusters analysis may help to increase awareness in educational, healthcare, and corporate contexts on the heteronormative arguments underlying explicit hate, such as the denial of the existence and marginalisation of queer people (Leonard et al. 2022).

From a philosophical point of view, one of the most significant aspects of anti-homotransphobic discourse is in the treatment of hateful speech acts as inherently dangerous for collectivity and immediately harmful to the out-group. According to Jonathan Seglow (2016: 7), hate speech acts cause direct damage to the out-group in terms of agency and entitlement to self-respect. The first term refers to the ability of an individual to pursue their own goals and to affirm a personal belief within society. These goals are more respectable the more they are endorsed by other members of the social community; in this regard, Seglow stresses how self-respect is configured as a collective construct and not a strictly individual status. The concept of entitlement to self-respect refers to the respectability that individuals perceive based on their abilities, merits, titles, and rights acquired from birth and throughout their lives.

By dehumanising the out-group, hate speech flattens the individual features of the group and diminishes its merits, projecting the image of a homogeneous social category whose goals are no longer perceived as rights but as privileges. From this perspective, not only does the LGBTQIA+ community have less entitlement regarding equal marriage and parenthood, but some studies have demonstrated that the demand for these rights and further access to legal protection from hate crimes as a social minority will be perceived as an undeserved extra (Leonard et al. 2022). Further studies must therefore focus on implicit and moderate hate speech, and particularly on the concepts of political priority and privilege while analysing discrimination against social minorities.

References

Arduini, Stefano, and Roberto Fabbri. 2013. *Che cos'*è la linguistica cognitiva. Rome: Carocci.

Arfini, Elia A. G., and Cristian Lo Iacono. 2012. *Canone inverso: antologia di teoria queer*. Pisa: Edizioni ETS.

Article 19. 2015. *'Hate Speech' Explained – A Toolkit*. London: Article 19. https://www.article19.org/data/files/medialibrary/38231/'Hate-Speech'-Explained---A-Toolkit-%282015-Edition%29.pdf.

Assimakopoulos, Stavros, Fabienne H. Baider, and Sharon Millar, eds. 2017. *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*. Berlin: Springer. https://doi.org/10.1007/978-3-319-72604-5.

Attanasio, Giuseppe, and Eliana Pastor. 2020. 'PoliTeam @ AMI: improving sentence embedding similarity with misogyny lexicons for automatic misogyny identification in Italian tweets'. In *EVALITA Evaluation of NLP and Speech Tools for Italian – 17 December 2020*, edited by Valerio Basile, Danilo Croce, Maria Maro, and Lucia C. Passaro, 48–54. Turin: Accademia University Press. https://doi.org/10.4000/books.aaccademia.6807.

Baker, Paul. 2016. 'Gendered discourses'. In *Triangulating Methodological Approaches in Corpus Linguistic Research*, edited by Paul Baker and Jesse Egbert, 138–151. London: Routledge.

Baker, Paul, Costas Gabrielatos, Majid Khosravinik, Michał Krzyżanowski, Tony McEnery, and Ruth Wodak. 2008. 'A useful methodological synergy? Combining Critical Discourse Analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press'. *Discourse and Society* 19 (3): 273–306. https://doi.org/10.1177/0957926508088962.

Basile, Valerio, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, eds. 2020. *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, 2765. Turin: Academia University Press. http://ceur-ws.org/Vol-2765/.

Basile, Valerio, Mirko Lai, and Manuela Sanguinetti. 2018. 'Long-term social media data collection at the University of Turin'. In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-It 2018*, edited by Elena Cabrio, Alessandro Mazzei, and Fabio Tamburini, 40–45. Collana Dell'Associazione Italiana Di Linguistica Computazionale. Turin: Accademia University Press. http://books.openedition.org/aaccademia/3075.

Bianchi, Claudia. 2017. 'Linguaggio d'odio, autorità e ingiustizia discorsiva'. *Rivista di estetica* 64: 18–34. https://doi.org/10.4000/estetica.2059.

Borrillo, Daniel Ángel. 2009. *Omofobia: storia e critica di un pregiudizio.* Bari: Edizioni Dedalo.

Bucholtz, Mary, and Kira Hall. 2004. 'Theorizing identity in language and sexuality research'. *Language in Society* 33 (October): 469–515. https://doi.org/10.1017/S0047404504334020.

Caselli, Tommaso, Valerio Basile, Jelena Mitrović, and Michael Granitzer. 2021. 'HateBERT: retraining BERT for abusive language detection in English'. In *Proceedings of the Fifth Workshop on Online Abuse and Harms (WOAH 2021)*, edited by Aida Mostafazadeh Davani, Douwe Kiela, Mathias Lambert, Bertie Vidgen, Vinodkumar Prabhakaran, and Zeerak Waseem, 17–25. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.woah-1.3.

Chakravarthi, Bharathi Raja, Ruba Priyadharshini, Thenmozhi Durairaj, John McCrae, Paul Buitelaar, Prasanna Kumaresan, and Rahul Ponnusamy. 2022. 'Overview of the shared task on homophobia and transphobia detection in social media comments'. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, edited by Bharathi R. Chakravarthi, B Bharathi, John P McCrae, Manel Zarrouk, Kalika Bali, and Paul Buitelaar, 369–377. Dublin: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.ltedi-1.57.

Coates, Jennifer. 2013. 'The discursive production of everyday heterosexualities'. *Discourse and Society* 24 (5): 536–552. https://doi.org/10.1177/0957926513486070.

Davidson, Thomas, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. 'Automated hate speech detection and the problem of offensive language'. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)*, 512–515. Montreal: AAAI Press.

De Carli, Mirko. 2021. 'Ddl Zan, non è prioritario I manifesti lo dimostrano'. *Il Resto del Carlino*, 13 August 2021. https://www.ilrestodelcarlino.it/ravenna/cronaca/ddl-zan-non-e-prioritario-i-manifesti-lo-dimostrano-1.6691834.

De Mauro, Tullio. 2016. 'Le parole per ferire'. *Internazionale*, 27 September 2016. https://www.internazionale.it/opinione/tullio-de-mauro/2016/09/27/razzismo-parole-ferire.

De Smedt, Tom, Sylvia Jaki, Eduan Kotzé, Leïla Saoud, and Maja Gwóźdź, Guy De Pauw, and Walter Daelemans. 2018. 'Multilingual cross-domain perspectives on online hate speech'. In *CLiPS Technical Report* 8: 1–24. https://doi.org/10.48550/arXiv.1809.03944.

Drożdż, Michał. 2016. 'Hate speech in media discourse'. *Acta Universitatis Lodziensis: Folia Litteraria Polonica* 35 (5): 19–30. https://doi.org/10.18778/1505-9057.35.02.

Duggan, Lisa. 2002. 'The new homonormativity: the sexual politics of neoliberalism'. In *Materializing Democracy: Toward a Revitalized Cultural*

Politics, edited by Russ Castronovo and Dana D. Nelson, 175–194. Durham, NC: Duke University Press.

Fersini, Elisabetta, Debora Nozza, and Paolo Rosso. 2020. 'AMI @ EVALITA2020: automatic misogyny identification'. In *EVALITA Evaluation of NLP and Speech Tools for Italian – 17 December 2020*, edited by Valerio Basile, Danilo Croce, Maria Maro, and Lucia C. Passaro, 21–28. Turin: Accademia University Press. https://doi.org/10.4000/books.aaccademia.6764.

Fillmore, Charles J. 1985. 'Frames and the semantic of understanding'. *Quaderni di Semantica: Rivista Internazionale id Semantica Teorica e Applicata* 6: 222–254.

Gagliardone, Iginio, Alisha Patel, and Matti Pohjonen. 2014. 'Mapping and analysing hate speech online: opportunities and challenges for Ethiopia'. *SSRN Electronic Journal*, 1 May 2014. https://dx.doi.org/10.2139/ssrn.2601792.

Gümüşay, Kübra. 2021. *Lingua e essere*, translated by Lavinia Azzone. Rome: Fandango Libri.

Hermann-Wilmarth, Jill M., and Caitlin L. Ryan. 2016. 'Queering chapter books with LGBT characters for young readers: recognizing and complicating representations of homonormativity'. *Discourse: Studies in the Cultural Politics of Education* 37 (6): 846–66. https://doi.org/10.1080/01596306.2014.940234.

Istituto A. T. Beck, Centro Studi Omosessualità. n.d. 'L'omosessualità è un disturbo mentale?'. Accessed 31 March 2022. https://www.istitutobeck.com/omosessualita.

Kulick, Don. 2005. 'The importance of what gets left out'. *Discourse Studies* 7 (4/5): 615–624. https://doi.org/10.1177/1461445605054408.

Leonard, William, Debra Parkinson, Alyssa Duncan, Frank Archer, and Catherine Orian Weiss. 2022. 'Under pressure: developing lesbian, gay, bisexual, transgender and intersex (LGBTI) inclusive emergency services.' *Australian Journal of Emergency Management* 37 (1): 52–58. https://doi.org/10.47389/37.1.52.

Locatelli, Davide, Greta Damo, and Debora Nozza. 2023. 'A cross-lingual study of homotransphobia on Twitter.' In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, edited by Sunipa Dev, Vinodkumar Prabhakaran, David Adelani, Dirk Hovy, and Luciana Benotti, 16–24. Dubrovnik: Association for Computational Linguistics. https://aclanthology.org/2023.c3nlp-1.3.

Liddicoat, Anthony J. 2009. 'Sexual identity as linguistic failure: trajectories of interaction in the heteronormative language classroom'. *Journal of Language, Identity and Education* 8 (2–3): 191–202. https://doi.org/10.1080/15348450902848825.

Malmasi, Shervin, and Marcos Zampieri. 2017. 'Detecting hate speech in social media'. In *Proceedings of the International Conference Recent*

Advances in Natural Language Processing, RANLP 2017, edited by Ruslan Mitkov and Galia Angelova, 467–472. Varna: INCOMA Ltd. https://doi.org/10.26615/978-954-452-049-6_062.

McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Mishra, Pushkar, and Marco Del Tredici. 2018. 'Author profiling for hate speech detection'. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, edited by Emily M. Bender, Leon Derczysnki, and Pierre Isabelle. Santa Fe, NM: Association of Computational Linguistics.

Mongie, Lauren. 2016. 'Critical Discourse Analysis as queer linguistics: religious pro- and anti-LGBT equality framing and counterframing in two letters to the editor in the City Press'. *Stellenbosch Papers in Linguistics Plus* 49: 23–43. https://doi.org/10.5842/49-0-664.

Motschenbacher, Heiko. 2020. 'Walking on Wilton Drive: a linguistic landscape analysis of a homonormative space'. *Language and Communication* 72: 25–43. https://doi.org/10.1016/j.langcom.2020.02.002.

Motschenbacher, Heiko. 2019. 'Methods in language, gender and sexuality studies: an overview'. *Wiener Slawistischer Almanach* 84: 43-70.

Motschenbacher, Heiko. 2018. 'Corpus linguistics in language and sexuality studies: taking stock and future directions'. *Journal of Language and Sexuality* 7 (2): 145–174. https://doi.org/10.1075/jls.17019.mot.

Motschenbacher, Heiko, and Martin Stegu. 2013. 'Queer Linguistic approaches to discourse'. *Discourse and Society* 24 (5): 519–535. https://doi.org/10.1177/0957926513486069.

Musto, Cataldo, Angelo P. Sansonetti, Marco Polignano, Giovanni Semeraro, and Marco Stranisci. 2019. 'HATECHECKER: a tool to automatically detect hater users in online social networks'. In *Proceedings of the Sixth Italian Conference on Computational Linguistics* (CLiC-it 2019), edited by Raffaella Bernardi, Roberto Navigli, and Giovanni Semeraro. Bari: CEUR-WS.

Muti, Arianna, and Alberto Barrón-Cedeño. 2020. 'UniBO @ AMI: a multiclass approach to misogyny and aggressiveness identification on Twitter posts using AlBERTo'. In *EVALITA Evaluation of NLP and Speech Tools for Italian – 17 December 2020*, edited by Valerio Basile, Danilo Croce, Maria Maro, and Lucia C. Passaro, 29–34. Turin: Accademia University Press. https://doi.org/10.4000/books.aaccademia.6769.

Norocel, Ov Cristian. 2011. 'Heteronormative constructions of Romanianness: a genealogy of gendered metaphors in Romanian radical-right populism 2000–2009'. *Debatte: Journal of Contemporary Central and Eastern Europe* 19 (1–2): 453–70. https://doi.org/10.1080/0965156x.2011.626121.

Nozza, Debora. 2022. 'Nozza@LT-EDI-ACL2022: ensemble modeling for homophobia and transphobia detection'. In *Proceedings of the Second*

Workshop on Language Technology for Equality, Diversity and Inclusion, edited by Bharathi R. Chakravarthi, B Bharathi, John P McCrae, Manel Zarrouk, Kalika Bali, and Paul Buitelaar, 258–264. Dublin: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.ltedi-1.37.

Orrù, Paolo. 2017. *Il discorso sulle migrazioni nell'Italia contemporanea*. Milan: Edizioni Franco Angeli.

Pamungkas, Endang Wahyu, Alessandra Teresa Cignarella, and Valerio Basile. 2018. 'Automatic identification of misogyny in English and Italian tweets at EVALITA 2018 with a multilingual hate lexicon'. In *EVALITA Evaluation of NLP and Speech Tools for Italian: Proceedings of the Final Workshop 12–13 December 2018, Naples*, edited by Tommaso Caselli, Nicole Novielli, Viviana Patti, and Paolo Rosso. Turin: Accademia University Press. https://doi.org/10.4000/books.aaccademia.4421.

Partington, Alan. 2014. 'Mind the gaps: the role of corpus linguistics in researching absences'. *International Journal of Corpus Linguistics* 19 (1): 118–146. https://doi.org/10.1075/ijcl.19.1.05par.

Patti, Viviana, Cristina Bosco, and Rossana Damiano. 2017. 'Implicazioni etiche dell'analisi di contenuti, emozioni e interazioni nei social media'. *Sistemi intelligenti* 3: 579–600. https://doi.org/10.1422/88510.

Perlman, Merrill. 2019. 'How the word 'queer' was adopted by the LGBTQ community'. *Columbia Journalism Review*, 22 January 2019. https://www.cjr.org/language_corner/queer.php.

Peterson, Cassie. 2013. 'The lies that bind: heteronormative constructions of "family" in social work discourse'. *Journal of Gay and Lesbian Social Services* 25 (4): 486–508. https://doi.org/10.1080/10538720.2013.829394.

Pinker, Steven. 2007. *The Stuff of Thought: Language as a Window into Human Nature*. London: Penguin.

Pistolesi, Elena. 2007. 'Identità e stereotipi nel discorso conflittuale'. In *Vicini/lontani: identità e alterità nella/della lingua*, edited by Elena Pistolesi and Sabine Schwarze, 115–130. Berlin: Peter Lang Verlag.

Poletto, Fabio, Marco Stranisci, Manuela Sanguinetti, Viviana Patti, and Cristina Bosco. 2017. 'Hate speech annotation: analysis of an Italian Twitter corpus'. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-It 2017)*, edited by Roberto Basili, Malvina Nissim, and Giorgio Satta, 263–268. Aachen: CEUR-WS. https://doi.org/10.4000/books.aaccademia.2448.

Reddy, Vasu. 2002. 'Perverts and sodomites: homophobia as hate speech in Africa'. *Southern African Linguistics and Applied Language Studies* 20 (3): 163–175. https://doi.org/10.2989/16073610209486308.

Rösner, Leonie, and Nicole C. Krämer. 2016. 'Verbal venting in the social web: effects of anonymity and group norms on aggressive language use in online comments'. *Social Media + Society* 2 (3). https://doi.org/10.1177/2056305116664220.

Rossi Barilli, Gianni. 1999. *Il movimento gay in Italia*. Milan: Universale Economica Feltrinelli.

Sanguinetti, Manuela, Fabio Poletto, Cristina Bosco, Viviana Patti, and Marco Stranisci. 2018. 'An Italian Twitter corpus of hate speech against immigrants'. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2798–2805. Miyazaki: European Language Resources Association.

Seglow, Jonathan. 2016. 'Hate speech, dignity and self-respect'. *Ethical Theory and Moral Practice* 19 (5): 1103–1116. https://doi.org/10.1007/s10677-016-9744-3.

Silva, Leandro, Mainack Mondal, Denzil Correa, Fabricio Benevenuto, and Ingmar Weber. 2016. 'Analyzing the targets of hate in online social media'. In *Proceedings of the Tenth International Conference on Web and Social Media (ICWSM 2016)*, edited by Markus Strohmaier and Krishna Gummadi, 687–690. Cologne: AAAI Press.

Socciarelli, Matteo. 2019. 'The "synonymy" of gay and omosessuale in Italian: a corpus linguistic analysis of two news websites'. In *TESOL Working Paper Series* 17, edited by Hanh Thi Nguyen, 46–68. Honolulu, HI: Hawaii Pacific University TESOL.

Thornton, Anna Maria. 2016. 'Designare le donne: preferenze, raccomandazioni e grammatica'. In *Genere e linguaggio: i segni dell'uguaglianza e della diversità*, edited by Fabio Corbisero, Pietro Maturi, and Emanuela Ruspini, 15–33. Milan: Franco Angeli.

Van Dijk, Teun A. 2004. *Ideologie, discorso e costruzione sociale del pregiudizio*, translated by Paola Villano. Rome: Carocci.

Warner, William, and Julia Hirschberg. 2012. 'Detecting hate speech on the World Wide Web'. In *Proceedings of the Second Workshop on Language in Social Media (LSM 2012)*, edited by Sara Owsley Sood, Meena Nagarajan, and Michael Gamon, 19–26. Montreal: Association for Computational Linguistics.

Waseem, Zeerak, Thomas Davidson, Dana Warmsley, and Ingmar Weber. 2017. 'Understanding abuse: a typology of abusive language detection subtasks'. In *Proceedings of the First Workshop on Abusive Language Online (ACL 2017)*, edited by Tommaso Caselli, Nicole Novielli, and Viviana Patti, 78–84. Vancouver: Association for Computational Linguistics.

WCF (World Congress of Families XIII). 2019. 'Il congresso'. Accessed 2 March 2022. https://wcfverona.org/it/about-the-congress/.

Wikipedia. n.d. 'Sexual inversion (sexology)'. Accessed 3 January 2023. https://en.wikipedia.org/wiki/Sexual_inversion_(sexology).

CHAPTER 5

Homophobic hate speech in Italian tweets

Contextual cues of offensiveness

Stefano De Pascale
Vrije Universiteit Brussel, FWO & KU Leuven

Edoardo Cavirani
KU Leuven

Stefania Marzo
KU Leuven

Abstract

Social media channels have become omnipresent tools for communicating, sharing knowledge, and establishing communities, but also places where all sorts of hate speech comes to the surface. This chapter contributes to the body of work on online hate speech towards sexual orientation minorities and takes some initial steps towards a quantitative variationist sociolinguistic study of homophobic language in Italian social media. By exploring the different lexicalisations (i.e. near-synonyms) available in Italian

How to cite this book chapter:
De Pascale, Stefano, Edoardo Cavirani, and Stefania Marzo. 2024. 'Homophobic hate speech in Italian tweets: Contextual cues of offensiveness'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 159–191. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-5.

for expressing or referring to the concept of HOMOSEXUAL MAN on X (formerly known as Twitter), this study provides both a descriptive and a methodological contribution. For this purpose, we created a dataset of 3000 manually annotated tweets in which at least one of the 33 lexicalisations of the concept HOMOSEXUAL MAN was recorded, and which we annotated for linguistic and stylistic factors that favour the perception of an expression as offensive. Our study shows that the interaction of lexical variation and explicit and implicit contextual cues, such as irony and dialect use, is significant in determining the degree of offensiveness of tweets. Furthermore, this study provides further evidence in favour of using social media as a laboratory for mapping language variation on a large scale, and for reflecting on the refinement of semi-automatic annotation of linguistic, stylistic, and social variables in written social media language.

Keywords: homophobia, lexical variation, contextual cues, conditional inference tree

5.1 Introduction

Since the massive rise of social media networks, online hate speech has been a widespread phenomenon, often targeting vulnerable minorities on the basis of their social, religious, or ethnic background, their gender identity, or their sexual orientation. This chapter focuses on hate speech towards the last of these groups and takes some initial steps towards a quantitative variationist sociolinguistic study of homophobic language in Italian social media. By exploring lexical variation in a large-scale corpus of terms referring to homosexuality on the social network X (formerly known as Twitter), this study provides both a descriptive and a methodological contribution. Given that there is increased evidence of the effects of hate speech on offline behaviour (see, among others, Soral, Bilewicz, and Winiewski 2018; Müller and Schwarz 2021), it is necessary to strive for a multidata approach to these issues. The current study, with its combination of corpus and attitudinal data, attempts to provide a contribution to such an endeavour.

On a descriptive level, the chapter offers an initial insight into lexical variation in the common terminology used on Italian social media to refer to homosexual people. For this purpose, a large-scale X corpus was compiled that includes expressions featuring the 33 most commonly used terms (henceforth called 'near-synonyms') referring to homosexual men. The aim of the study is not only to identify variation *per se*, but to examine to what extent the degree of offensiveness towards homosexuals correlates with the variation in terminology. To this end, we examine which linguistic and stylistic factors favour the perception of an expression as insulting.

In addition to providing a quantitative analysis of an X corpus, this chapter addresses a number of methodological issues. Although we are not the first to work with X data for the analysis of hate speech (Poletto et al. 2017; Sanguinetti et al. 2018; Chapter 4 in this volume), our analysis aims to provide further support for the value of X data (and social media data in general) as a research tool for studying language variation (Bohmann 2016; Coats 2016; Grondelaers and Stuart-Smith 2021; Grondelaers and Marzo 2023). Specifically, this study provides further evidence in favour of using X as a laboratory for mapping language variation on a large scale, and for reflecting on the annotation of linguistic, stylistic, and social variables in written social media language.[1] In what follows, we will discuss the possibility of annotating stylistic aspects of spoken language that can also be found in social media language, and that seem to be highly relevant for explaining the construction of homophobic language.

In the remainder of this introduction, we will outline our research against the background of the public debate on hate speech in general; specifically, we will focus on the complexity of

1 Since the collection of the data in March 2021 and the drafting of the chapter in 2022, Twitter has changed name and owner (now being called X), and has severely limited the free functionalities of its API and the availability of its data. As a result of these restrictions, the microblogging service has partially lost its appeal as a provider of large amounts of informal written language.

pinpointing hate speech due to the multiplicity of contextual factors that play a role in the construction of offensive speech. In Section 5.2, we describe our dataset and the methodological tools that we have used to study lexical variation in our data. We will also explain the annotation parameters for the detection of the stylistic and linguistic cues of offensiveness, as well as the automatic annotation of the sociodemographic characteristics of the X users. Section 5.3 presents the results of our statistical analysis, which are organised in terms of the descriptive and methodological goals of the study. In the discussion and conclusion (Section 5.4), we flesh out the most striking results while reflecting on the limitations of this preliminary study but also on the possibilities that it has opened up for further research on the use of homophobic language in Italy.

In recent years, interest in the detection of online hate speech has steadily increased, due to the societal impact of the phenomenon since the rise of web content. In particular, the automatisation of hate speech detection has grown significantly. Natural language processing has become a primary method for detecting hate speech since it became clear that simple word queries did not provide sufficient insight into such a complex phenomenon. Indeed, hate messages appear to be determined not only by the use of explicit hate words and slurs but also by multiple contextual aspects such as 'the domain of an utterance, its discourse context, as well as […] co-occurring media objects (e.g. images, videos, audio), the exact time of posting and world events at this moment, identity of author and targeted recipient' (Schmidt and Wiegand 2017: 1), in addition to more abstract semantic and discursive frames (for an example of those, see the discussion on heteronormative matrices which go beyond sexual identity in Chapter 4 in this volume). Although it is controversial and generally condemned, hate speech is particularly multilayered and may not be directly observable at first sight. As Federico Faloppa puts it, observing and studying hate speech is therefore like 'chasing the panther to which Dante, in *De vulgari eloquentia*, compares the illustrious vernacular: although you can smell it, it cannot be

grasped' (Faloppa 2020: 22, our translation). The boundaries of what can be called 'hate' are far from clear, and relatively subjective.

According to research since the mid-2010s (Bianchi 2017, 2021; Gao, Kuppersmith, and Huang 2017; Caselli et al. 2020; Faloppa 2020), there are two types of hate speech. The most obvious type involves utterances that contain explicitly derogatory terms—that is, terms referring to stereotypes, insults, or threats, or that clearly incite hatred or discrimination. In some cases, it contains explicitly the word 'hate' or other words that evoke hate (Orlando and Saab 2021). The degree of offensiveness of these words is widely recognised, and is almost objectively measurable, as confirmed by the extensive literature on swearing and slurs (see Jay and Janschewitz 2008, Beers Fägersten and Stapleton 2017).

Much more complex and less studied are the implicit verbal manifestations of hate, conveyed by words that are not necessarily derogatory in and of themselves (see chapters 4 and 7 in this volume). This type of hate speech often implicitly contains, suggests, or builds on offensive associations which are difficult to pinpoint (see, e.g., Chapter 4 in this volume for the role played by a set of semantic frames in homotransphobic speech in an X corpus that does not contain explicit incitement of hatred). There is generally less consensus about the impact of this form of hate speech, and it is the context in which it occurs that seems to determine its degree of offensiveness.

As these implicit forms of hate speech are less evident, they are also much more complex to investigate. Natural language processing has been investigating the role of contextual factors in the construction of hate speech for years. Many aspects were found to play a role in the development of offensive or aggressive language use, for example linguistic features such as imperatives or specific syntactic constructions. However, a detailed study of the role of finer-grained linguistic features in hate speech has not yet been carried out (whereas the role that some of these features—specifically, 'deviations' from a canonical grammatical structure in phonology, morphology, and syntax—play in the encoding of

emotion/expressivity has been given some attention; see Corver 2016; Saab 2021, 2022; see also Trotzke and Villalba 2021).

This chapter explores the difference between explicit and implicit hate speech in reference to homosexual men (see Chapter 1 for a comparison with the distinction between 'hard' and 'soft' hate speech, particularly in a legal context). It investigates the role of the linguistic context in the creation of homophobic hate speech and aims to better understand the grey area in which apparently neutral terms nevertheless lead to hate messages. We therefore use quantitative methods to investigate which semantic, structural, and stylistic elements operate together in order to construct hate speech towards homosexuals.

Our work can thus be seen as complementary to Safina's study (Chapter 4) in this volume. Safina uses X data on homotransphobia and aims to go beyond lexical evidence, but, unlike our study, she carries out a qualitative study that tries to understand 'which semantic spheres are involved in stereotypical representations of the queer community' and to detect which word clusters possibly construct these semantic spheres. Crucially, she shows that the most significant clusters are not related to words expressing desire, sexuality, or erotic practices, but rather to Politics (i.e. law and criminality, freedom and fundamental rights, and immigration), Nature (i.e. health, diseases, and dehumanising references to animals), and Values and Customs (i.e. morality, family, and religion).

5.2 Materials, data, and methods

5.2.1 Data collection and corpus

In order to investigate online homophobic language, we decided to look at X data.[2] X represents a relevant type of digital social space, which people across the world use to communicate, share

2 We have made the instructions to annotate the tweets, the sample of annotated tweets, and the data for the perception survey publicly available through the OSF-project 'Homophobic hate speech in Italian tweets: contextual cues of offensiveness', retrievable via https://osf.io/4ftq8/.

knowledge, and establish communities (Fuchs 2014), but also where all sorts of anti-social behaviour comes to the surface relatively unfiltered. The verbal manifestation of this behaviour, hate speech, has been studied in detail by computational linguists (Schmidt and Wiegand 2017). One of the reasons for its popularity in this field is that X has for a long time made the data extraction process very easy through its API (Application Programming Interface; but see note 1 for the current state of affairs), which we also take advantage of in our study. However, tweets are also popular in variationist research, not only because they are available in enormous quantities but also because they are an extremely rich source of informal language use with features mirroring orality through several compensation techniques, such as variant spellings or expressive compensation strategies (repetition of characters, excessive use of interpunctions, etc.). This non-standard language use and these orthographic compensations are seen as a 'form of identity, signaling authenticity, solidarity of resistance to norms imposed from above' (Eisenstein 2013: 362, cited in Grondelaers and Marzo 2023). For this reason, X has been found to be highly suitable for the investigation of language variation and the indexical meanings attached to it, including, as in this case, expressions of hate or offensiveness.

For this study, we first determined the lexical field consisting of all the near-synonymous lexicalisations of the concept HOMOSEXUAL MAN. For two main reasons, we have opted to focus exclusively on terms referring to homosexual men, and not on women. This is partially for reasons of feasibility, but we also have a theoretical-methodological justification for this choice. While it might be arguable, we found that homosexual men and women do not belong to the same conceptual category from a linguistic point of view. Methodologically, it is therefore not completely appropriate to bundle them together under one unique conceptual 'flag': it would have been odd to have a concept HOMOSEXUAL PERSON with the subordinate concepts of HOMOSEXUAL MAN and HOMOSEXUAL WOMAN. We have therefore opted to consider them as distinct concepts.

Table 5.1: Near-synonyms of the concept HOMOSEXUAL MAN.

Near-synonyms (Treccani)
bardassa, buco, checca, cinedo, culattone, culorotto, cupio, finocchio, frocio, gay, invertito, omo, omofilo, omosessuale, omosex, paraculo, pederesta, recchione/ricchione, sodomita, uranista
Near-synonyms (Wikipedia)
bucaiolo, buggerone, buliccio, caghinero, ciucciacazzo, ciucciapisello, garrusu, matellu, piglianculo, pivellu, puppu, rottinculo, succhiacazzo

In order to obtain the most exhaustive list possible, thereby adhering to the principle of accountability (a cornerstone of variationist sociolinguistics; see Tagliamonte 2006), we primarily consulted two types of resource. First, we skimmed through the online Treccani synonyms and antonyms dictionary.[3] The procedure for finding all the near-synonyms was carried out recursively: we i) started out with the lemma *gay*, ii) recorded all the synonyms listed under that entry, and then iii) searched for each synonym lemma individually, so as to retrieve the synonyms of the synonyms of *gay*. This search continued until no more new synonyms could be retrieved. We then consulted the Italian Wikipedia (n.d.) entry 'Lessico dell'omofobia' (The lexicon of homophobia) as a complementary resource, in order to retrieve regionally marked words that were not included in the Treccani dictionary. In total, we collected 33 near-synonym nouns for this concept, which are listed in Table 5.1.

A quick glance at the two lists shows, unsurprisingly, the different focus and coverage of the two resources. Whereas the Treccani list covers a diverse set of properties, the Wikipedia list provides near-synonyms that are either regional or with a very negative connotation. Based on the stylistic and sociolinguistic labels in the Treccani, it is possible to identify roughly four qualitatively

3 The dictionary can be accessed online at https://www.treccani.it/enciclopedia/elenco-opere/Sinonimi_e_Contrari.

different groups of near-synonyms: vulgar and derogatory terms; regional or dialectal terms; archaic and uncommon terms; literary terms. As some near-synonyms receive multiple labels, we have summarised the relationships in the full lexical field in Figure 5.1. Given that the geosynonyms taken from Wikipedia are all consistently considered derogatory, they are shown as nested within the pool of vulgar terms. There is also wide variation in the origin of the regional terms: *matellu*, *garrusu*, and *puppu* are found in Sicily; *caghinero* and *pivellu* in Sardinia; *checca*, *bucaiolo*, and *finocchio* are considered to have originated in Tuscany; *buliccio* in Liguria; *frocio* in Rome; *ricchione* in Southern Italy; whereas *culattone* and *cupio* seem to come from northern Italian varieties. Some of these terms (e.g. *checca*, *finocchio*, and *frocio*) no longer have a specific regional distribution and might thus be considered pan-Italian.

Some terms, which are also the most frequently used, have no label and are classified in Figure 5.1 as a superordinate category of 'neutral' terms: *gay*, *omosessuale*, *omosex*, *pederasta*, and *sodomita*. The term 'neutral' refers in the first place to those near-synonyms that do not receive any sociolinguistic or stylistic label in the given source (Treccani or Wikipedia). For the first three terms, we can indeed expect that the lack of such a label genuinely represents the unmarked nature of these near-synonyms. Furthermore, *gay*, *omosessuale*, and *omosex* will be particularly relevant for our upcoming analyses: as they are not inherently negatively connotated, in order to use them in a derogatory way, speakers have to resort to strategies involving the manipulation of their sentential context. However, with respect to the latter two words in this category, *pederasta* and *sodomita*, the lack of label is probably more due to specific lexicographic choices than to a truthful representation of the sociolinguistic status of these near-synonyms.

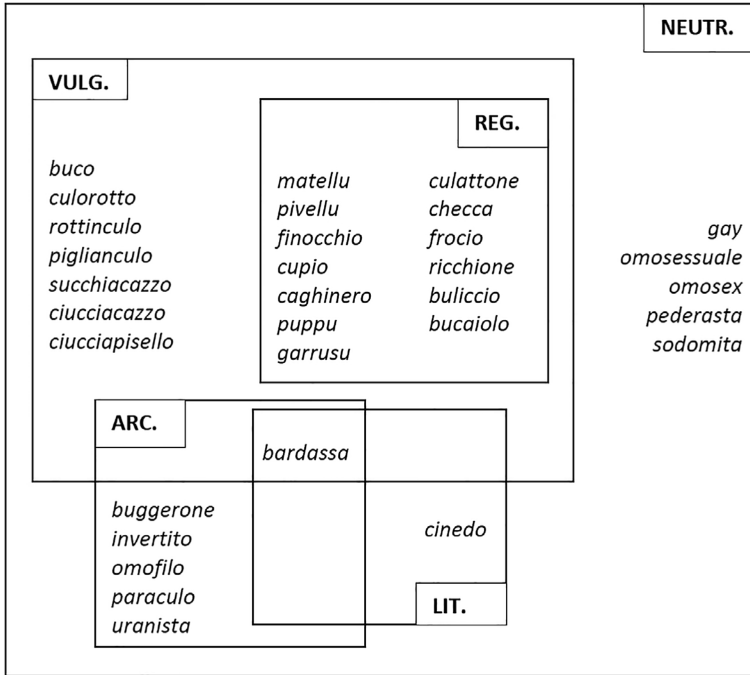


Figure 5.1: Qualitative classification of the near-synonyms, in terms of vulgarity (VULG.), regionality (REG.), archaism (ARC.), literary use (LIT.), and neutrality (NEUTR.).

For each of these terms we set up a search making use of X's API. Given the lack of elaborate wildcard possibilities in the API, or basic lemmatisation, we retrieved both plural and singular forms of the terms separately. We searched for tweets in which all of the 33 near-synonyms appeared from 2006 (when the site now known as X was launched) up to 2020, making a total of 14 years of tweets. This resulted in a database of more than 2.6 million tweets in which one of the lexicalisations of HOMOSEXUAL MAN was used. From this collection we randomly sampled 3044 tokens—occurrences of each of the near-synonyms—for further inspection and annotation, such that two conditions were met: i) the relative proportion (in terms of frequency) of the near-synonyms in the dataset sampled for annotation would mirror the

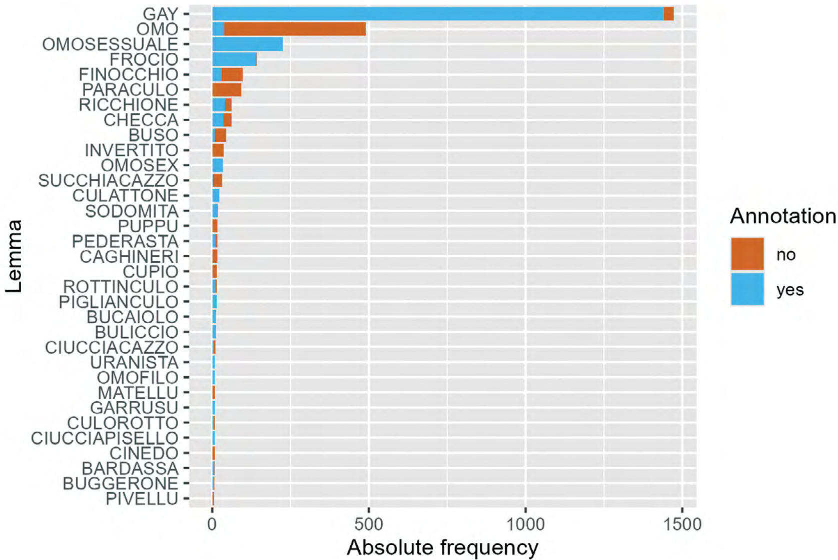


Figure 5.2: Absolute frequencies of near-synonyms for the concept HOMOSEXUAL MAN.

one of the larger database as much as possible, and ii) at least five tokens per near-synonym would be sampled. Figure 5.2 plots the set of near-synonyms for which tokens were found, in descending order of frequency. Almost half of the sampled tokens are occurrences of *gay* (1442), which is therefore by far the most frequent lexicalisation of the concept. Not taking into account *omo* (which is mainly used in a different sense; see below), the second most frequent lexicalisation is also a connotationally neutral term, namely *omosessuale* (225). The third most frequent is the derogatory *frocio* (141), which appears more than three times as often as the next most frequent terms, *ricchione* (41) and *checca* (35). Unsurprisingly, the whole lexical field of HOMOSEXUAL MAN follows a Zipfian word frequency distribution.[4]

4 When generating frequency lists of word types in a corpus, one typically encounters a so-called Zipfian frequency distribution (named after George K. Zipf [1949]). The fundamental property of such a distribu-

5.2.2 Neutral and offensive near-synonyms and the role of context

Before we began our quantitative analysis of the impact of linguistic and stylistic factors on the degree of offensiveness, we explored the degree of offensiveness of these near-synonyms in a preliminary perception test. The aim of this pilot test was to better understand the difference between supposedly neutral terms and offensive near-synonyms and to assess the specific role of linguistic and stylistic context among native Italians. We therefore administered an online perceptual test (in Qualtrics) to 49 Italian respondents, all native speakers (with 20 men, 26 women, and 3 respondents who indicated no specific gender). Each was presented with a set of the five most common near-synonyms referring to homosexuals, those that occurred most frequently in the corpus: *gay*, *omosessuale*, *ricchione*, *culattone*, *frocio*. The first two near-synonyms were supposed to be less offensive (and more commonly used), whereas the other three were supposedly more directly and inherently offensive. The five near-synonyms were alternately presented separately (i.e. without context) and in a context containing linguistic (morphological, dialectical) or stylistic (ironic) elements that contributed to making the term more expressive and possibly offensive. For each tweet, respondents had to indicate the degree of offensiveness on a scale from 0 to 10.

Figure 5.3 shows the relative (average) values given by respondents for each term, both in a context without linguistically or stylistically expressive cues and in one with those cues. The results of this perception experiment clearly showed that *gay* and *omosessuale*, the supposedly unmarked terms, can also be perceived

tion is its skewness: a few word types are extremely frequent (usually such lists are topped by closed class types such as prepositions, pronouns, and articles) while the vast majority of the remaining word types are very infrequent, mainly occurring just once or twice (mainly open class items). The lexical field under investigation here also follows this type of distribution, having one very frequent lexicalisation (*gay*) and a long tail of infrequent near-synonyms (occurring just a few times). Figure 5.2 clearly displays the skewness in the frequency counts.

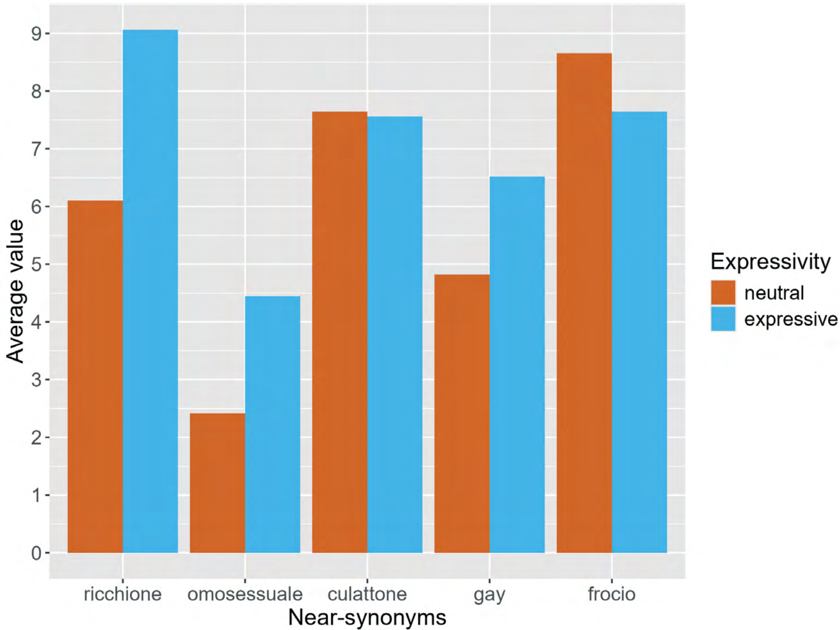


Figure 5.3: Average scores of offensiveness of five near-synonyms for homosexual uses in neutral or expressive contexts.

offensively in a linguistically and stylistically charged tweet: the averages for these terms were clearly different depending on the context. For the other three terms—the supposedly inherently offensive *ricchione*, *culattone*, and *frocio*—there was less noticeable difference between the two contexts, although there still was some difference noticed for *ricchione*. This means that these terms might be perceived as inherently offensive regardless of the contextual elements in which they are used. This pilot test served as preparatory work for the annotation and analyses we conducted on a larger scale with the X corpus (see following sections) and as a validation of our starting hypothesis that hate speech, and specifically homophobic terminology in this case, is strongly dependent on the linguistic context in which it occurs.

5.2.3 Annotation of the context

The next step involves the annotation of the tokens. Two native speakers of Italian were recruited for this task and both annotated the set of sampled tokens. Each tweet was annotated for linguistic and sociolinguistic parameters. The primary goal of the annotation was to provide grammatical and semantic disambiguation for the usage of the near-synonyms. This data-cleaning step principally involved three filters.

First, as we were only interested in uses of the words that refer to a person, which largely corresponds to nominal uses of the terms, we discarded all non-noun tokens, particularly adjectival uses that refer to the quality of being homosexual. Some of the lexemes in the list, such as the most frequent *gay* and *omosessuale* (e.g. *un atteggiamento gay/omosessuale*, 'a gay/homosexual attitude'), can in fact be used both as nouns and as adjectives. However, the majority of the terms, especially the more derogatory ones, can only be used as nouns (e.g. *un atteggiamento *checca/*frocio/da checca/da frocio*, where asterisks indicate ill-formedness). The removal of adjectival use is particularly important for the term *invertito*, which is often found as a past participle of the verb *invertire,* 'to turn, to exchange'. Similarly, we had to discard some hits of *checche*, as they turned out to be accent-less instances of the indefinite pronoun *checché*, 'whatever' (social media orthography is often characterised by such spelling). The second cleaning step involves cases of homonymy, such as the case of *finocchio*, which mostly refers to the 'vegetable fennel'. A lexeme such as *paraculo* never refers to a homosexual man in our sample, but rather to 'an opportunist, someone who can skilfully and casually turn a situation in their favour'. Therefore, this lexeme was also discarded. The majority of *omo* tokens are not instances of the meaning of the concept we are interested in here, but are occurrences of the Roman dialectal variant of *uomo* 'man, mankind'. In this step we also removed instances of lexemes that refer to women rather than to men. A third class of discarded tokens are those in which it was impossible, within the context of the tweet, to ascertain what the

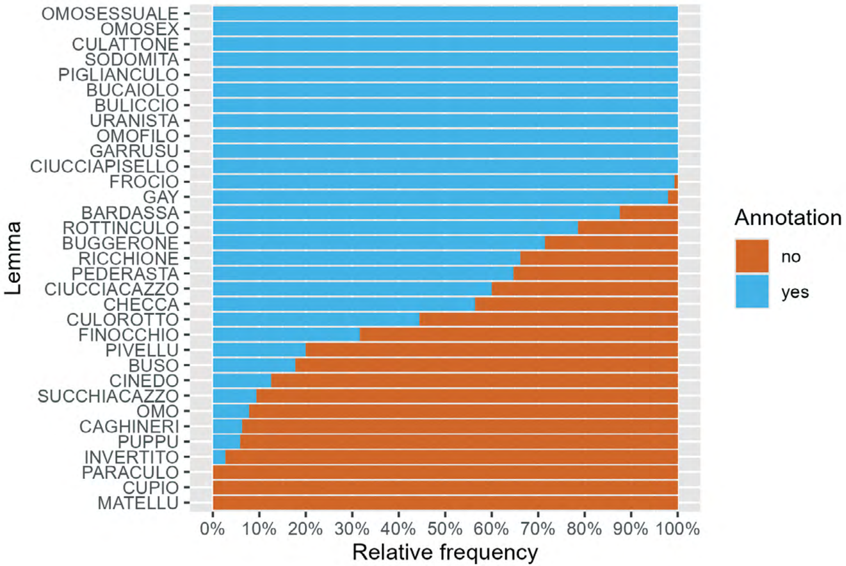


Figure 5.4: Proportions of 'in-concept' (blue) vs 'out-of-concept' (orange) tokens, ordered in decreasing order of 'in-concept' tokens.

word referred to. Given that we had not filtered out tweets shorter than a specific number of words, nor tweets that do not have well-formed syntactic structure, some tweets did not provide enough clues to allow us to disambiguate them.

At the end of this cleaning and disambiguation procedure, we were left with 2159 tokens (around 70 per cent of the dataset). Figure 5.4 summarises the proportions of 'out-of-concept' occurrences (i.e. the tokens that did not refer to HOMOSEXUAL MAN) vs 'in-concept' occurrences for each near-synonym (i.e. all tokens that did refer to HOMOSEXUAL MAN and that were retained in the corpus). This plot shows the importance of a disambiguation and data-cleaning step for many of the lexemes in our dataset, without which the quantitative results would very likely be distorted.

The second part of the annotation consisted in the identification of other characteristics of the tweet that are considered relevant for the analysis of hate speech in social media. The annotation

scheme was modelled on that in Poletto et al. (2017), who also annotated a dataset of Italian tweets (for computational-linguistic purposes; see Chapter 4 in this volume for a similar annotation scheme). In the following we describe the annotation scheme we provided to our two annotators. The first three parameters are taken from Poletto et al. (2017). The first is 'degree of offensiveness'. In contrast to aggressiveness, this parameter focuses more on the potentially offensive effect of the content of the tweet. According to this parameter, a tweet is considered weakly offensive in most cases. These are cases in which, for example, the target is associated with a certain characteristic (biological, sociological, behavioural, etc.) that emphasises the target's status as a disadvantaged or discriminated minority, or cases in which a description of the target is proposed that qualifies the target as an unpleasant person; on the other hand, if overtly offensive language is used, or if the target is addressed with outrageous or degrading expressions, the tweet is considered highly offensive. The coding of the degree of offensiveness consisted of three values: 0 (neutral content), 1 (mildly offensive content), and 2 (highly offensive content). For all the other features, the coding is binary: either the presence or the absence of the feature.

(1) ma che cacchio dice?? vuol attirare l'attenzione con argomenti da invasato finocchio culattone!! sparisci. [highly offensive]

'what the heck is he saying?? he wants to attract attention with possessed gay faggot arguments!! get lost.'

The second parameter is 'irony'. This parameter determines whether the tweet is ironic or sarcastic rather than being based on the literal meaning of words. The third is 'stereotype', which determines whether the tweet contains implicit or explicit references to (mostly false) beliefs about a given target.[5]

5 It is notoriously hard to generate reliable and consistent annotations for the abovementioned variables, due to the high degree of subjectivity involved and the difficulty in reaching an operational definition

(2) Capisco gli uomini che diventano gay perché si rendono conto che è meglio prenderlo al culo da cialtroni frustrati come loro. Le riflessioni geniali della domenica sera. Buonanotte Twitter.

‘I understand men who turn gay because they realise it is better to take it up the ass from frustrated wafflers like them. Sunday night genius musings. Goodnight Twitter.’

The following parameters are original additions for our specific study. The fourth is ‘unconventional spelling’, and refers to the presence of unconventional punctuation or unexpected spelling arguably translating unexpected/emphatic pronunciations (e.g. vowel or consonant lengthening or use of capital letters).

(3) GAY SI NASCE di solito forse boh chi lo sa ED IO NON LO NAQQQUIII.

‘ONE IS BORN GAY usually maybe eh who knows AND I WASN’T BORN LIKE THAT.’

The fifth concerns ‘dialectal words’, namely the use of words that are not part of (neo)standard Italian and can be considered dialectal or markedly regional, as in

(4) @USERNAME ‘c’è la crisi e te stai a pensa’ ai matrimoni de’ li froci’

‘@USERNAME “there is crisis and you are thinking about gay weddings” ’

The next three parameters are related to morphosyntax. The first concerns ‘dislocations’, namely the presence or absence of right- or left-dislocation of the nominal phrase containing the near-synonym, or the pronoun referring to it. The relevance of this parameter

of those notions (Sanguinetti et al. 2018). For degree of offensiveness, irony, and stereotype, we find inter-annotator agreement values, measured with weighted Cohen’s K, between 0.14 and 0.42, which indicate only fair agreement overall.

for the detection of hate speech builds on: i) the assumption that 'dislocations' are 'deviations' from the (pragmatically) unmarked constituent order that bring about a change in information structure and have a marked pragmatic function (Lambrecht 1994), and ii) the hypothesis that such a function might be that of signalling a negative emotive attitude on the part of the speaker (Fónagy 1995; Oliveira 2013). The second is 'derivational morphology', namely the presence or absence of derivational morphemes (diminutives, pejoratives, reduplication) on nouns, adjectives, and adverbs. The final is the presence or absence of deictics such as pronouns and demonstratives, which can be employed to emphasise social distinctions and, crucially, the attitude of a speaker towards a specific referent, such as a person or a group of people (Hart 2010, 2014; Da Milano 2016).

(5) Bene è l'ora della buona notte, etero gay trans che voi siate vi amo quasi tutt*! #noeterosessualità ahaahah

'Well it's night time, straight gay trans whatever you are I love almost all of you! #noheterosexuality hahaha'

Finally, two demographic parameters are reliant on the information provided by the X API. One is the gender of the user, which is not given by X users as such in their profile but can be inferred from their username or user screen name. The assumption was that the username would be a good indicator of the gender of the user, which was coded as either male, female, or 'not available' (when the username did not allow the recognition of a gender). The other demographic parameter is 'geographical origin', which was based on the information available in the 'place_country', 'place_name', or 'user_location' tags of the output. If a real, concrete place was mentioned, the tweet was coded with the Italian region the place was located in. Together with the semantic and grammatical disambiguation, a total of 12 properties were annotated.

Based on this dataset, we carried out two sets of analyses. In the first analysis, we explored the impact of the annotated parameters—linguistic, stylistic, and demographic—of the tweets on the

distribution of our near-synonyms (Section 5.3.1). The second analysis adopted the opposite perspective and investigated which of the properties of the tweet create 'offensive' tweets (Section 5.3.2). In variationist and statistical parlance, in the first analysis, 'degree of offensiveness' corresponds to the 'predictor variable' of the analysis, while the near-synonyms are the 'response variable'. Both the predictor and response variables are categorial, with the response variable following a multinomial distribution. In the second analysis, 'degree of offensiveness' corresponds to the response variable, and the distribution of near-synonyms is part of the set of predictors.

In both cases, the statistical technique we adopt is *conditional inference tree*. This technique is part of the family of classification tree methods, and offers a few advantages over similar inferential methods: first, its output is mostly visual (in the form of a tree structure), which makes an analysis of the patterns easier to interpret for users with a more limited knowledge of statistics; second, it is especially useful for reporting on how multiple features cooperate in the selection of the near-synonyms (or, in statistical parlance, how the predictors interact); and third, it is in principle more robust against the expected correlation between the features and the unbalanced distribution of both our response variable and the predictors. As our near-synonyms follow a skewed, Zipfian distribution, and the majority of the tweets received a '0' or 'no' coding for many features, this is a particularly relevant property of conditional inference trees.

In brief terms, the technique works as follows. Firstly, independence tests between each predictor and the response variable are performed and the predictor that covaries most strongly with the response variable gets selected. In the following step, the technique splits the data into two subsets according to the levels of this selected predictor and then once again tests all predictors as before on these separate partitions of the data. This procedure is repeated until no further splits are justified by the independence tests (in other words, until no statistically significant patterns remain). The result is a flowchart-like decision tree, with

the strongest discriminative predictor at the top of the tree, and recursive splitting by other predictors generates a hierarchy of interacting predictors. The leaves at the end of the branches of the tree are barplots that show the distribution of the response variable values for that subset of predictor variables selected by means of the interactions (see Strobl, Malley, and Tutz 2009; Tagliamonte and Baayen 2012; and Gries 2020 for an introduction, but also for the drawbacks of this type of technique). For the analysis of our dataset, we made use of the implementation of inference trees available in the `party` and `partykit` packages in R.

5.3 Analyses

5.3.1 Analysis with 'near-synonyms' as response variable

We start with an analysis of the conditional inference tree that has the set of near-synonyms as the response variable. This should help us to understand which properties of the tweet (i.e. the predictor variables) drive the distribution of those synonyms. This analysis has two components: the first focuses on the 9 most frequent near-synonyms, while the second focuses on the 3 most frequent ones: *gay*, *omosessuale*, and *frocio*. There are two reasons why it is reasonable not to model the full set of near-synonyms. The first is methodological, and comes as a consequence of the Zipfian distribution of the lexical field: with most of the words having very low frequencies, it is impossible to carry out a statistical analysis that involves some type of interaction. Recall that the conditional inference tree algorithm functions with increasingly smaller subsets. As a consequence, the infrequent items would quickly fall out of the picture. We maintain that looking at 9 near-synonyms instead of 32 strikes a good balance between variation and feasibility.

For the analysis of the three most frequent items, the analysis becomes statistically even more robust. Moreover, we consider each of the three items to have a specific 'sociolinguistic' profile. *Gay*, the most frequent item, is a loanword; *omosessuale* is the

standard Italian counterpart with the same neutral connotational value as *gay*; and *frocio* is the most frequent regionally marked and derogatory item. In a sense, *frocio* can be seen as the exponent of the range of regional offensive terms and can be argued to cover the other terms in the dataset.

We start with the conditional inference tree of the nine most frequent items: *gay*, *omosessuale*, *frocio*, *ricchione*, *omo*, *checca*, *omosex*, *finocchio*, and *culattone* (Figure 5.5). The sociolinguistic predictors 'gender' and 'region' do not seem to have a significant impact, nor do the presence of stereotypes, dialectal terms, unconventional orthography, derivational morphology, or the use of dislocations. All these predictors are excluded from the plotted conditional inference tree.

Unsurprisingly, the most determinant predictor is 'degree of offensiveness', which sets apart non-offensive tweets from offensive tweets (regardless of the magnitude). The second most important predictor is 'irony', which only has an impact on the non-offensive tweets. And within the group of ironic tweets, the use of deictics also has an influence on the distribution of the near-synonyms. When looking at the four barplot figures (i.e. the 'leaves' of the tree) from left to right, a natural cline can be observed from the most neutral and least expressive tweets (i.e. the inoffensive and unironic tweets) to the most charged and expressive tweets (i.e. the most offensive ones and those with an ironic undertone and/or deictic forms). With this cline of 'expressivity' comes a marked change in the distribution of the nine near-synonyms: neutral tweets follow the expected overall distribution, with a large predominance of *gay*. The more a tweet becomes expressive (and here 'offensive'), the smaller the proportion of *gay* tweets (i.e. the bars become smaller) and the higher the lexical diversity.

This relation between affect and lexical variation has been attested in previous lexicological research (see, among others, Franco et al. 2019), showing that negative connotations boost lexical variation and the use of less frequent and more marked alternative words. However, in the absence of a clear negative connotation (in the left branches of inoffensive tweets), X users resort

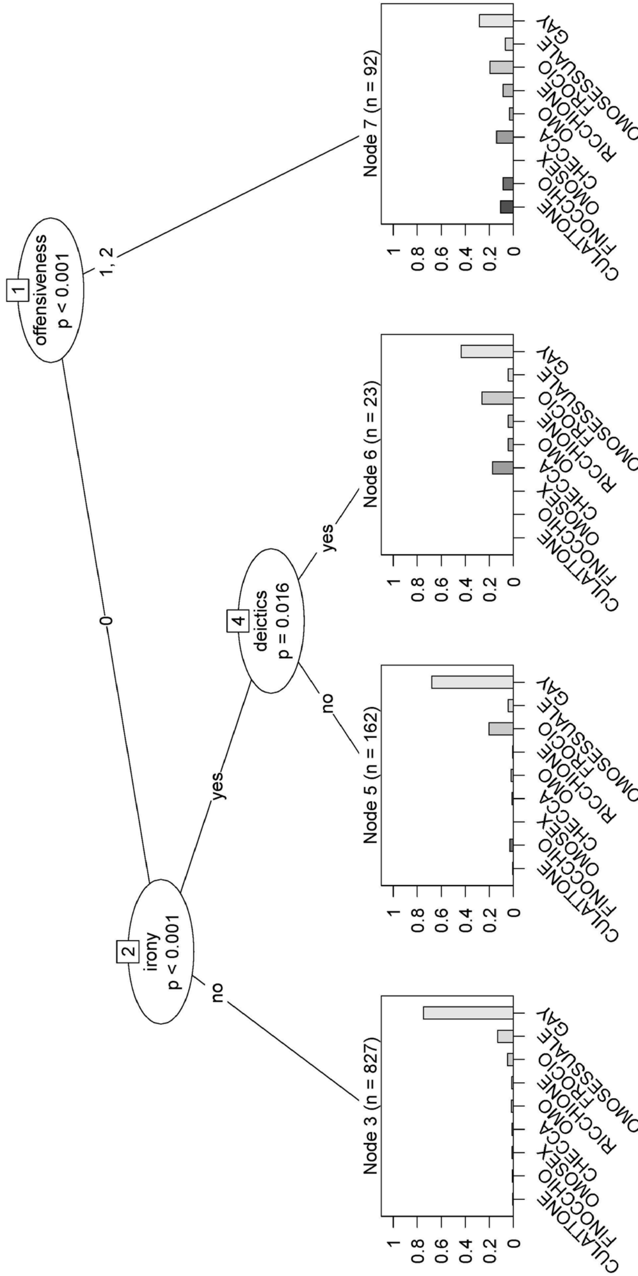


Figure 5.5: Conditional inference tree plot of the impact of the features on the distribution of the nine most frequent near-synonyms.

to other stylistic strategies that trigger the use of marked lexicalisations, such as irony and deictics, which tend to reinforce one another (in the interaction, Node 6).

After the nine near-synonyms, we focus on the three most frequent ones (*gay*, *omosessuale*, and *frocio)* in the conditional inference tree shown in Figure 5.6. Restricting the lexical variation brings about a reordering of the importance of predictors, and the addition of new ones. While the presence of irony and offensive language remains influential, we see an increased relevance of the gender assigned to the X user and the presence of dialectal terms. In the left branches of the tree, where we find the unironic tweets, we observe a similar pattern as in Figure 5.5: in neutral tweets, speakers largely favour the use of *gay*, while in more offensive but unironic tweets, there is a decrease in *gay* which is coupled with an expected increase in the derogatory term *frocio*. The right branching of the tree reveals new information. The three rightmost barplot groups (nodes 6, 8, and 9) show a similar cline as that which was attested in the previous analysis, but this time structured along completely different predictors. Ironic tweets written by female users (Node 6) follow the general pattern of stark preference for *gay*, although there is a noticeable increase in the derogatory *frocio*. On the other hand, ironic tweets written by male users largely deviate from the overall distribution. When those tweets do not contain dialectal terms, *gay* is still the majority term, but *frocio* accounts for a third of occurrences. However, when male X users insert dialectal terms into their tweets, *gay* becomes the minority lexicalisation, and instances of *frocio* and *omosessuale* are greater. This is also the first time *omosessuale*, which has a similar neutral connotational value as *gay*, is more frequently attested than the loanword. The majority lexicalisation is yet another term, *frocio*. In sum, ironic tweets containing dialectal features written by men, as opposed to women, use the derogatory term twice as much as the overall most frequent near-synonym *gay*. Two observations should be made with regard to these results. First, the preponderance of the indigenous Italian words, and especially *omosessuale*, with respect to the loanword

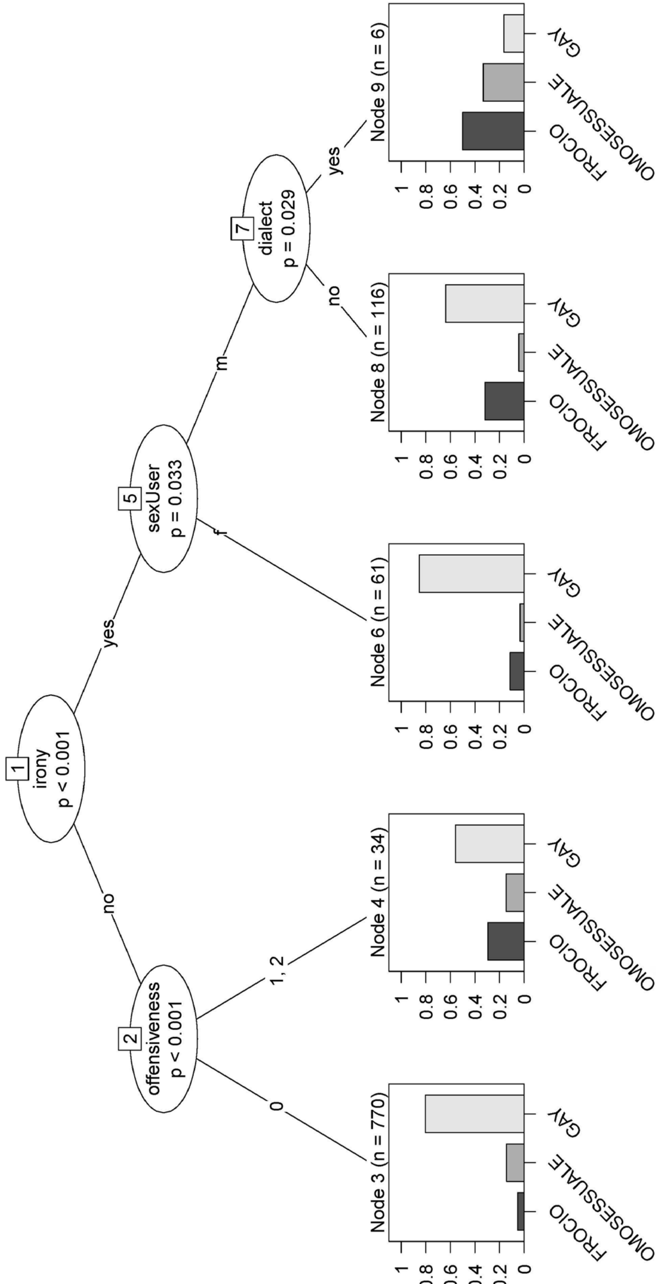


Figure 5.6: Conditional inference tree plot of the impact of the features on the distribution of the three most frequent near-synonyms.

gay, in tweets that contain some dialectal material, might indicate that users tend to be linguistically coherent in their production. *Gay* might be perceived as the more modern, or perhaps more written-language alternative, whereas *omosessuale* could be more easily integrated into an utterance which is either closer to the regional or dialectal speech of the user, or is embedded in a more informal conversation or expression on X. Further research into the nuances between those two terms will hopefully shed more light on this division of labour. Second, in the right branching of the tree (i.e. in all the ironic tweets) 'degree of offensiveness' no longer plays a major role. This means that among the abovementioned tweets one can find both offensive and inoffensive tweets, and that the dialectal tweets written by men are not necessarily offensive. This leads to the observation that the relative preference for the Italian terms over the loanword *gay* is not only a matter of contextual and communicative expressivity, but rather a stable property of the subcode of this sociolinguistic group of writers/speakers—that is, men writing in dialect. We have to stress at this point that the number of tweets being ironic, written by men, and with the inclusion of dialectal words, is very small: only six (i.e. Node 9 on the very right of the tree only contains six tweets). These last observations therefore cannot be generalised with the same confidence as those made on the other nodes of the conditional inference tree.

5.3.2 Analysis with 'degree of offensiveness' as response variable

The second analysis adopts the opposite perspective and looks at the factors that influence the 'degree of offensiveness' of a tweet, thus taking offensiveness as the response variable. This time the three-way distinction in 'not offensive', 'mildly offensive', and 'very offensive' (see Section 5.2.3) has been reduced to a binary distinction—offensive (labelled 'yes' in the conditional inference tree) vs inoffensive (labelled 'no')—in order to determine what contributes

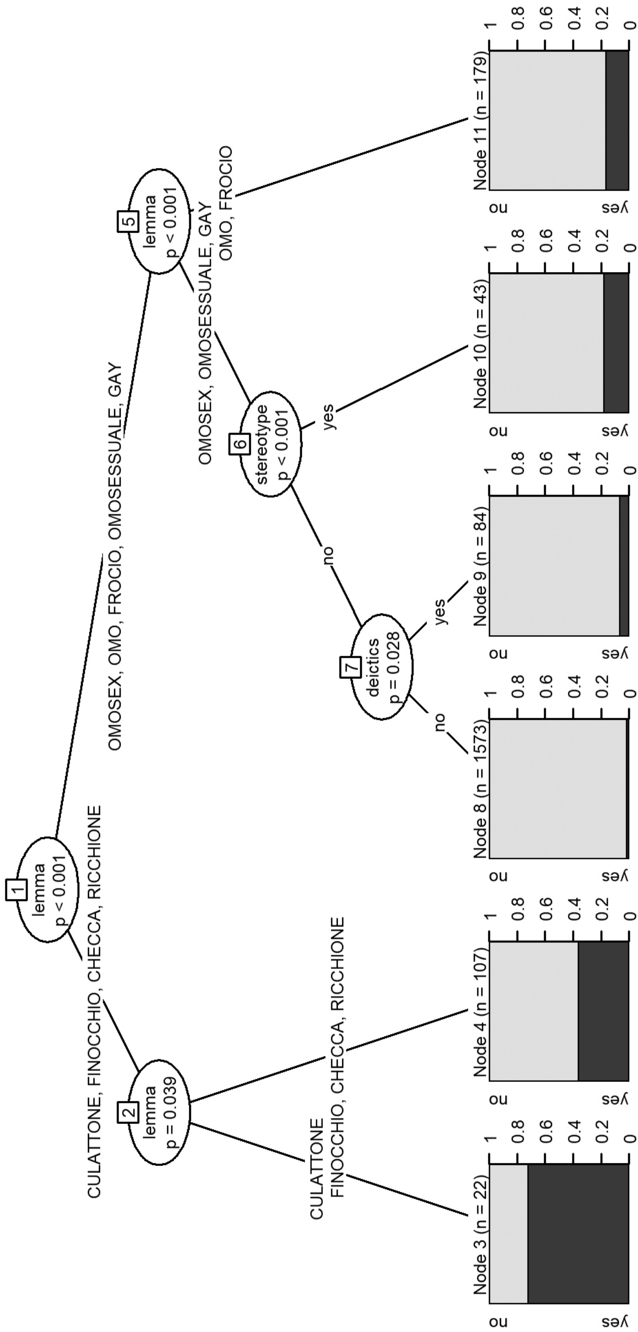


Figure 5.7: Conditional inference tree plot with offensiveness as a response variable.

to the offensive character of a tweet about a HOMOSEXUAL MAN or in which a HOMOSEXUAL MAN is mentioned.

In this case, we constructed a conditional inference tree with the near-synonym set as a predictor variable and offensiveness as a response variable. The same series of predictors as used in the previous analysis is included. What could we learn from turning a response variable into a predictor variable, and vice versa, with all the other predictors being retained? An interesting feature of conditional inference trees is that they can split predictors with many levels into groups of levels that behave homogeneously with respect to their impact on the response variable. In our case, 'near-synonym set' is such a multilevel predictor (with nine levels). So, while in the first analysis (in which the near-synonym set was a response variable) every level of that response was treated individually and mutually independently, we can now allow levels to be grouped together to assess their commonalities. In other words, we will be able to identify which near-synonyms are used in a similar way in the dataset and which of them lead to offensive tweets and under which contextual conditions.

In Figure 5.7, we plotted the conditional inference tree with 'degree of offensiveness' as the response variable. This tree clearly shows that the nine different near-synonyms form several groups, as 'lemma' is now also the most important predictor. At the top of the tree, the first split takes place between the tweets containing *culattone*, *finocchio*, *checca*, and *ricchione* on the one hand, and *omosex*, *omo*, *frocio*, *omosessuale*, and *gay* on other hand. Before examining the interaction structure of the predictors in the branches, we can observe that in the stacked barplots of the leaves there is a non-trivial amount of variation between offensive and inoffensive tweets. This means that not all inherently derogatory terms in the first group automatically lead to or belong to offensive tweets, and not all neutral near-synonyms in the second group appear in inoffensive tweets. The context in which those words function will likely employ other compensatory or boosting strategies to either decrease or increase the expressive (and negative) effect of the tweet.

The barplots reveal the following overall pattern: *culattone*, *finocchio*, *checca*, and *ricchione* are associated with a relative majority of offensive tweets (nodes 3 and 4), while *omosex*, *omo*, *frocio*, *omosessuale*, and *gay* (nodes 8 to 11) are associated with far less offensive tweets. As the conditional inference tree further splits and diversifies, we can observe further granular distinctions within these two higher-order groups. In the most offensive group, *culattone* splits from *finocchio*, *checca*, and *ricchione*. The vast majority (73 per cent) of the tweets in which *culattone* appears are offensive, versus a sizeable minority (36 per cent) of the tweets with *finocchio*, *checca*, or *ricchione*. This means that the remaining 64 per cent of tweets containing the negatively connotated words *finocchio*, *checca*, or *ricchione* were not considered offensive, leaving open the question of what makes this possible. It is also remarkable that a word like *frocio* clusters with the more neutral group. Just as there was a split between extreme and very negative terms in the group of derogatory terms, there is also a split within the neutral group in Node 5: on the one hand, *omosex*, *omosessuale*, and *gay* (only 3 per cent of the tweets containing these words were considered offensive), and on the other hand, *omo* and *frocio* (18 per cent of the tweets containing these words were considered offensive). In summary, we again find a cline in 'degree of offensiveness', with at one extreme the tweets with *culattone*, followed by the group including *finocchio*, *checca*, and *ricchione*, then *omo* and *frocio*, and then the least offensive ones: *omosex*, *omosessuale*, and *gay*.

How does this interact with the other linguistic predictors in our tree? As expected, within the group of least offensive words (*omosex*, *omosessuale*, and *gay*), stylistic and indirect strategies play a substantial role and reach statistical significance: it is only when these terms are accompanied by deictics and stereotypes that the percentage of offensive tweets goes up (2 per cent for offensive tweets when no deictic or stereotypes are used, 19 per cent when a stereotype is mentioned, and 8 per cent when a deictic is used). Conversely, for the group of derogatory terms, these strategies do not seem to have an effect: the terms carry enough expressive

power themselves to communicate hate and do not need to be boosted. Compensatory strategies, which would help to decrease the expressive load of a tweet, were not included in the annotation scheme.

5.4 General discussion and conclusions

In this section, we discuss the main results of our study and explore future applications of this methodology in studies on hate speech, and on social media in general.

This preliminary analysis of the dataset allowed us to show that homophobic terminology in social media is heavily dependent on the linguistic context in which it appears. Linguistic and stylistic cues such as ironic and dialectal language can endow harmless terms with an offensive connotation. This was clearly shown by the second set of inference trees (Figure 5.7). The other analyses showed, among other things, that these same cues were used more often by male X users and, when combined, also lead to: i) an increased use of certain derogatory terms, including *frocio* and *culattone*, and ii) a decrease in the use of the more recent term *gay*. When the tweet was perceived as globally offensive, without contextual cues, more lexical variation in usage occurred. These results confirm that the role of the linguistic context in which terms are used is fundamental to the construction of homophobic language and possibly hate speech in general. This also confirms that natural language processing research should pay significant attention to these microlinguistic aspects in further refining automatic recognition of hate language, but also to other, more stylistic cues, such as ironic or dialectal language use.

On a methodological level, we demonstrate the usefulness of X and social media in general for the large-scale investigation of language variation and change (Grondelaers and Stuart-Smith 2021). We were able to show that X data presents a promising research tool for exploring contextual cues in the construction of hate speech. Not only does it open up a gigantic data source for lexical variation research but it also allows for microlinguistic analyses

of the context in which the relevant words or variables occur. The tweets feature not only morphological and syntactic cues, but also other complex and layered cues, such as ironic and dialectal usage.

Of course, this initial exploration does have some limitations. One is that, at this stage, this fairly large dataset was annotated by only two people. This is too limited to arrive at a broad and valid inter-annotator agreement. For annotations without an obvious measure, such as the degree of offensiveness, this poses a number of problems related to the subjectivity of the annotators. Thus, in future research, more annotators will be needed; they should vary in gender, age, and possibly regional origin. Furthermore, we could cluster among the predictors a number of (perhaps correlating) variables to create larger parameters that could tell us something about the language use in question. Following the example of Grondelaers and Marzo (2023), non-standard spelling and repetition, along with other parameters, could be a manifestation of reinforcement or expressive language use. These parameters could then be used as broader stylistic cues (stylisation) in the regression analyses. Finally, the study should be extended to hate speech towards homosexual women, with a similar dataset of near-synonymous lexicalisations of the concept of HOMOSEXUAL WOMAN. This expansion would not only offer further insight into how homophobic language is used on social media but would also provide a new parameter in the research. It would allow us to explore not only lexical differences in the conceptualisation of HOMOSEXUAL WOMAN (compared to men) but also variation in the use of conceptual cues in the construction of homophobic language towards men and women.

In other words, much work remains to be done and despite the limitations of the present study, we hope that these initial analytical steps have paved the way for further research into the rich variation in the lexical field of homophobic language use in Italy.

References

Beers Fägersten, Kristy, and Karyn Stapleton, eds. 2017. *Advances in Swearing Research: New Languages and New Contexts*. Amsterdam: John Benjamins.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.

Bianchi, Claudia. 2017. 'Linguaggio d'odio, autorità e ingiustizia discorsiva'. *Rivista di estetica* 64: 18–34. https://doi.org/10.4000/estetica.2059.

Bohmann, Axel. 2016. 'Language change because Twitter? Factors motivating innovative uses of because across the English-speaking Twittersphere'. In Squires, *English in Computer-Mediated Communication*, 149–178. https://doi.org/10.1515/9783110490817.

Caselli, Tommaso, Valerio Basile, Jelena Mitrović, Inga Kartoziya, and Michael Granitzer. 2020. 'I feel offended, don't be abusive! Implicit/explicit messages in offensive and abusive language'. Paper presented at *LREC 2020 – 12th International Conference on Language Resources and Evaluation*, Conference Proceedings, 6193–6202. Marseille: European Language Resources Association.

Coats, Steven. 2016. 'Grammatical feature frequencies of English on Twitter in Finland'. In Squires, *English in Computer-Mediated Communication*, 179–210. https://doi.org/10.1515/9783110490817-009.

Corver, Norbert. 2016. 'Emotion in the build of Dutch: deviation, augmentation and duplication'. *Tijdschrift voor Nederlandse Taal-en Letterkunde* 132 (4): 232–275.

Da Milano, Federica. 2016. 'Deictic strategies as expression of identity'. In *Beyond Language Boundaries: Multimodal Use in Multilingual Contexts*, edited by Marta Fernández-Villanueva and Konstanze Jungbluth, 153–161. Berlin and Boston, MA: De Gruyter.

Eisenstein, Jacob. 2013. 'What to do about bad language on the internet'. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, edited by Lucy Vanderwende, Hal Daumé III, and Katrin Kirchhoff: 359–369. Atlanta, GA: Association for Computational Linguistics.

Faloppa, Federico. 2020. *#ODIO: manuale di resistenza alla violenza delle parole*. Turin. UTET.

Fónagy, Ivan. 1995 'Iconicity of expressive syntactic transformations'. In *Syntactic Iconicity and Linguistic Freezes*, edited by Marge E. Landsberg, 285–304. Berlin and New York: De Gruyter Mouton.

Franco, Karlien, Dirk Geeraerts, Dirk Speelman, and Roeland Van Hout. 2019. 'Concept characteristics and variation in lexical diversity in two Dutch dialect areas'. *Cognitive Linguistics* 30 (1): 205–242. https://doi.org/10.1515/cog-2017-0136.

Fuchs, Christian. 2014. *Social Media: A Critical Introduction*. London: Sage.

Gao, Lei, Alexis Kuppersmith, and Ruihong Huang. 2017. 'Recognizing explicit and implicit hate speech using a weakly supervised two-path bootstrapping'. *Approach Computation and Language* (cs.CL). https://doi.org/10.48550/arXiv.1710.07394.

Gries, Stefan Th. 2020 'On classification trees and random forests in corpus linguistics: some words of caution and suggestions for improvement'. *Corpus Linguistics and Linguistic Theory* 16 (3): 617–647. https://doi.org/10.1515/cllt-2018-0078.

Grondelaers, Stefan, and Stefania Marzo. 2023. 'Why does the *shtyle* spread? Street prestige boosts the diffusion of urban vernacular features'. *Language in Society*: 52(2): 295–320. https://doi.org/10.1017/S0047404521001202.

Grondelaers, Stefan, and Jane Stuart-Smith. 2021. 'Twitter as a laboratory for language variation and change. New opportunities for social-media sociolinguistic research'. Panel organised at *New Ways of Analyzing Variation (NWAV) 49*. Austin, TX, 19–24 October 2021.

Hart, Christopher. 2014. *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. London: Bloomsbury.

Hart, Christopher. 2010. *Critical Discourse Analysis and Cognitive Science: New Perspectives on Immigration Discourse*. Basingstoke: Palgrave.

Jay, Timothy, and Kristin Janschewitz. 2008. 'The pragmatics of swearing'. *Journal of Politeness Research* 4 (2): 267–288. https://doi.org/10.1515/JPLR.2008.013.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Müller, Karsten, and Carlo Schwarz. 2021. 'Fanning the flames of hate: social media and hate crime'. *Journal of the European Economic Association* 19 (4): 2131–2167. https://doi.org/10.1093/jeea/jvaa045.

Oliveira, Ruth de. 2013. 'Manifestations émotionnelles de la dislocation: le cas de l'agacement'. In *Cartographie des émotions: propositions linguistiques et sociolinguistiques*, edited by Fabienne Baider and Georgeta Cislaru, 211–222. Paris: Presses Sorbonne Nouvelle.

Orlando, Eleonora, and Andrés Saab, eds. 2021. *Slurs and Expressivity: Semantics and Beyond*. Lanham, MD: Lexington.

Poletto, Fabio, Marco Stranisci, Manuela Sanguinetti, Viviana Patti, and Cristina Bosco. 2017. 'Hate speech annotation: analysis of an Italian Twitter corpus'. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-It 2017)*, edited by Roberto Basili, Malvina Nissim, and Giorgio Satta, 263–268. Aachen: CEUR-WS. https://doi.org/10.4000/books.aaccademia.2448.

Saab, Andrés. 2022. 'Introducing expressives through equations: implications for the theory of nominal predication in romance'. In *Proceeding of SALT 32*, 356–383. https://doi.org/10.3765/salt.v1i0.5330.

Saab, Andrés. 2021. 'On the locus of expressivity. deriving parallel meaning dimensions from architectural considerations'. In Orlando and Saab, *Slurs and Expressivity*, 17–44.

Sanguinetti, Manuela, Fabio Poletto, Cristina Bosco, Viviana Patti, and Marco Stranisci. 2018. 'An Italian Twitter corpus of hate speech against immigrants'. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2798–2805. Miyazaki: European Language Resources Association.

Schmidt, Anna, and Michael Wiegand. 2017. 'A survey on hate speech detection using natural language processing'. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. Valencia: Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-1101.

Soral, Wiktor, Bilewicz Michał, and Winiewski Mikołaj. 2018. 'Exposure to hate speech increases prejudice through desensitization'. *Aggressive Behavior* 44 (2): 136–146. https://doi.org/10.1002/ab.21737.

Squires, Lauren, ed. 2016. *English in Computer-Mediated Communication: Variation, Representation, and Change*. Berlin and Boston, MA: De Gruyter Mouton.

Strobl, Carolin, James Malley, and Gerhard Tutz. 2009. 'An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests'. *Psychological Methods* 14 (4): 323–348. https://doi.org/10.1037/a0016973.

Tagliamonte, Sali A. 2006. *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tagliamonte, Sali A., and Harald R. Baayen. 2012. 'Models, forests and trees of York English: was/were variation as a case study for statistical practice'. *Language Variation and Change* 24 (2): 135–178. https://doi.org/10.1017/S0954394512000129.

Trotzke, Andreas, and Xavier Villalba. 2021. *Expressive Meaning Across Linguistic Levels and Frameworks*. Oxford: Oxford University Press.

Wikipedia. n.d. 'Lessico dell'omofobia'. Accessed 23 November 2022. https://it.wikipedia.org/wiki/Lessico_dell%27omofobia.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. New York: Hafner.

CHAPTER 6

The impact of bilingualism on hate speech perception and slur appropriation

An initial study of Italian UK residents

Mattia Zingaretti
York St John University

Maria Garraffa
University of East Anglia

Antonella Sorace
University of Edinburgh

Abstract

The complex relationship between bilingualism and emotions has been extensively studied since the early 2000s, but the potential impact of bilingualism on speakers' perceptions and reactions to an emotionally loaded topic such as hate speech has been overlooked. This chapter reports the first investigation of this kind,

How to cite this book chapter:
Zingaretti, Mattia, Maria Garraffa, and Antonella Sorace. 2024. 'The impact of bilingualism on hate speech perception and slur appropriation: An initial study of Italian UK residents'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perceptio*n, edited by Silvio Cruschina and Chiara Gianollo, 193–230. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-6.

examining whether hate speech perception differs for late bilinguals in their first language (L1) versus their second language (L2), and how bilingual experience factors such as length of residence in the L2 country and language dominance may predict these perceptions. This research also explores whether the same factors, along with identifying with a sexual or ethnic minority, may predict bilinguals' perception of appropriateness in using slurs to react to hate speech, and whether bilinguals would appropriate slurs themselves. The bilingual group surveyed consists of 43 highly proficient L1 Italian speakers of L2 English, who grew up in Italy until at least the age of 16 and have been in the UK for an average of 5 years. The results indicate that the participants perceive hate speech rather similarly in their L1 and L2. Importantly, despite the overall higher familiarity with L1 hate words, a longer period of residence in the UK is associated with L1 hate words becoming less accessible in terms of familiarity, use, and imageability, while L2 words become less offensive. Moreover, slur appropriation is not predicted by any of the bilingual experience variables, but only by whether participants identify as part of a minority. The findings are discussed with reference to bilingualism research on L1 attrition and emotion, and by highlighting the implications of considering bilinguals' unique perceptions of hate speech from both linguistic and interdisciplinary perspectives.

Keywords: hate speech perception, slur appropriation, bilingualism, second language acquisition, first language attrition

6.1 Introduction

The proliferation of content across various media has allowed hate speech to spread more widely, consequently exacerbating concerns as to how to best define and identify it (see MacAvaney et al. 2019; Kovács, Alonso, and Saini 2021; see also chapters 1, 4, 5, and 7 in this volume). Importantly, these concerns are not confined to the online realm, but affect society and individuals more broadly, highlighting the pervasive nature of hate speech and its impact on social interactions and public discourse (see Chapter 1

for an introduction to these general issues). In addition to this, more than half of the world's population today is bilingual (Grosjean 2010), and the questions of how bilingualism may impact the perception of hate speech, as well as bilingual speakers' reactions to hate speech, have been overlooked.[1] Interestingly, while the literature on how bilinguals express emotions is rich and outlines a complex interplay of factors behind bilingual speakers' language choice when expressing emotions (see, e.g., Dewaele 2010), there is a definite lack of research that explicitly examines the way that speakers of more than one language perceive the emotions conveyed by the use of hate words in their different languages, or how bilingual speakers may react in situations where they encounter, or are the target of, hate speech. The research contained in this chapter thus represents a first attempt at filling this gap.

We begin by highlighting the main research findings on bilingualism and emotions (Altarriba 2003; Ramírez-Esparza et al. 2006; Pavlenko 2006, 2008; Kim and Starks 2008; Wilson 2008, 2013; Dewaele 2010; Gawinkowska, Paradowski, and Bilewicz 2013; Costa et al. 2014), as well as hate speech and emotions (Calvert 1997; Boeckmann and Liew 2002; Gelber and McNamara 2016; Brown 2017a,b; Chiril et al. 2022), and then proceed to review some pragmatic accounts of slur appropriation employed as a way of reacting to hate speech (Hornsby 2003; Hom 2008; Bianchi 2014). After explaining the motivations for our investigation, we present our methodology and report hate speech perceptions and reactions of a group of late bilinguals (namely, 43 Italian people resident in the UK). In particular, we analyse their perception of hate words via ratings of word pairs as well as their reactions to hate speech scenarios with and without slur appropriation, in both their first language (i.e. Italian, henceforth 'L1') and their second language (i.e. English, henceforth 'L2'). After discussing the results gathered through our online study, we high-

1 Throughout this chapter, we use terms such as 'bilingualism' and 'bilingual' to refer to the use of more than one language, synonymously with 'multilingualism' and 'multilingual'.

light the important implications of the current and future research for other disciplines (such as philosophy and ethics, as well as law and politics), acknowledging the limitations of this study and providing suggestions for future research into bilingualism and hate speech.

6.1.1 Research questions and hypotheses

This study analyses the potential impact of bilingualism on hate speech perception and slur appropriation, by seeking to answer the following research questions (RQs):

RQ1. Do (late) bilingual speakers perceive hate speech similarly in their two languages, and is their perception in the L1/L2 predicted by bilingual experience factors—specifically, by the length of residence (LoR) in the L2 country and their language dominance?

RQ2. Do the same factors (i.e. L1/L2, LoR, and dominance), as well as identifying as part of a (sexual or ethnic) minority, predict the degree to which bilingual speakers: a) find it appropriate to use slurs to react in situations where someone is the target of hate speech, and b) would appropriate slurs themselves in such situations?

Our main hypotheses (Hs) are the following:

H1. Firstly, bilingual speakers' perception of slurs may be predicted by increasing length of residence in the L2 country and a switch to L2 dominance, with hate speech being perceived more emotionally in the L2 and less in the L1.

H2. Secondly, bilinguals may find it more appropriate to use slurs as a response to hate speech in the L2 and less appropriate in the L1, the longer they reside in the L2 country and the more L2 dominant they are. Identifying with a minority may predict the degree of appropriateness perceived, irrespective of other factors. Further, identifying with a minority may also predict the degree

> to which bilingual speakers appropriate slurs themselves, and may interact with bilingual experience factors: sexual and ethnic minorities may appropriate slurs more in the L2 and less in the L1, the longer they reside in the L2 country and the more dominant they are in the L2.

The main findings of existing research on the links between bilingualism and emotions as well as the links between hate speech and emotions are reviewed below, laying the groundwork for the present investigation.

6.2 Background

6.2.1 Bilingualism and emotions

A large amount of the research into bilingualism and emotions since the turn of the century has highlighted the fact that bilingual speakers feel different when speaking different languages. For instance, Ramírez-Esparza et al. (2006) investigate whether bilinguals show different personalities with regard to the Big Five personality traits in English and in Spanish (i.e. extraversion, agreeableness, openness, conscientiousness, and neuroticism) and whether these differences are consistent with differences between English- and Spanish-speaking cultures. The bilinguals in their study are indeed found to be more extroverted, agreeable, and conscientious (but not more open or more neurotic) in English than in Spanish, and these differences are consistent with the personality displayed in each culture.

Similarly, Pavlenko (2006) explores whether multilingual speakers feel different when changing languages. After analysing 1039 responses to an open question about feeling different in a foreign language in her Bilingualism and Emotion Questionnaire (BEQ), Pavlenko finds that 65 per cent of her participants report feeling different when using another language compared to only a quarter of participants who reported not feeling different, with 10 per cent giving ambiguous responses. Pavlenko further observes that there are four main sources of perception of a

different self in a foreign language for her participants: linguistic and cultural differences, distinct learning contexts, different levels of language emotionality, and different levels of language proficiency. What is interesting to note here is that feeling different in a foreign language is not restricted to late or immigrant bilinguals, but, as Pavlenko (2006: 27) states, it 'is a more general part of bi- and multilingual experience'. Her participants also describe their bilingual experiences as enjoyable and unique and refer to the integration of their different identities.

Following Pavlenko, Wilson (2008) further investigates the issue of feeling different in a foreign language, and categorises and quantifies the responses of 1414 participants in the BEQ. The feelings reported by Wilson's participants when interacting in their foreign language are overall positive, with most participants feeling as if it were somebody else speaking, hence reporting feeling more confident and more outgoing in their foreign language, as well as highlighting changes in body language, mannerisms, and voice, and deeper levels of disguise. Wilson (2013) later correlates the BEQ scores of 108 adult participants with scores on the Big Five personality traits: while gender and age have no effect, higher BEQ scores of feeling different in a foreign language are reported by introverts who rate their L2 proficiency at intermediate level or above, participants with lower educational levels, participants with higher levels of perceived L2 proficiency and who started learning the L2 at a younger age, as well as mixed and naturalistic learners as opposed to instructed learners.

The evidence reviewed so far shows that feeling different in an L2 is incredibly common among most bilinguals, and that the degree to which bilinguals feel different in their L2 can be affected by factors such as L2 learning contexts and age, as well as proficiency and educational levels. More recently, research into the factors influencing the use of one of the languages (LX) of a multilingual speaker to express emotions finds similar results. In particular, Dewaele (2010) notes that late learners tend to use the LX less frequently to communicate emotions, rating positive characteristics of the LX lower and reporting higher levels of anxiety

when communicating in the LX. Moreover, similarly to Pavlenko (2006) and Wilson (2008, 2013), Dewaele finds that languages learnt only through formal instruction are less frequently used to communicate emotions, and that naturalistic and mixed learners feel more competent and less anxious than instructed learners. Something not noted in previous research but reported in Dewaele is that frequency of use of the LX, LX socialisation, and networks of interlocutors, along with the number of languages spoken, are also all factors that significantly influence the use of an LX to express emotion: specifically, more frequent use of an LX, higher levels of socialisation in the LX, and larger networks of interlocutors in the language, as well as a higher number of LXs spoken, all correspond to more frequent use of an LX to express emotions. Lastly, as also reported in Pavlenko (2006), Dewaele observes that communicating emotions covers a range of speech acts that are often culture specific: for example, raising one's voice when angry may be acceptable in some parts of the Western world, such as Southern Europe, but it is considered taboo in Asia. Importantly, multilingual speakers are able to exploit their multicompetence to develop speech acts and emotion scripts that are entirely unique to them or their peers, thus revealing an incredibly dynamic aspect of language choice to express emotions, and showing a growing awareness of sociocultural and sociopragmatic LX norms is accompanied by an evolution of the LX user's repertoire to express emotions in the LX. While cultural and social norms need to be learnt in the L2, these kinds of norms are already acquired in the L1—and they are not the only aspects that a late bilingual speaker is already acquainted with in their L1.

Indeed, some research has demonstrated the existence of a greater emotional overtone of the L1 connected with first emotion experience (Altarriba 2003): in other words, because certain emotions may be experienced first in a bilingual's L1, the expression of those emotions may come more easily for them in their L1, as that context presumably carries more connotations and associations, at least for late bilinguals. The fact that words in the L1 may seem more natural for late bilinguals, while words in the L2 may

be considered 'disembodied', as some further research suggests (Pavlenko 2008), leads Gawinkowska, Paradowski, and Bilewicz (2013) to point out that two opposite conclusions could be drawn: either it may be easier for bilinguals to talk about emotional topics in their 'more natural' L1, or they may prefer to do so in their L2 when social and cultural norms of their L1 may be 'too burdensome'. In their research, Gawinkowska, Paradowski, and Bilewicz report evidence that supports the latter conclusion. Specifically, they analyse the offensiveness ratings given by 61 Polish–English bilingual students to two texts they are asked to translate in both of their languages (from the L1 to their L2 and from the L2 to their L1). Gawinkowska, Paradowski, and Bilewicz find an effect only when analysing the results of offensiveness in target ethnophaulisms (i.e. words targeting social groups, which, as such, are subject to greater norms of political correctness) as opposed to general swear words (which are subject to lower norms of political correctness). Specifically, ethnophaulisms in the L2 translations are found to be significantly more offensive than source words in the L1, and vice versa: ethnophaulisms in the L1 translations are significantly less offensive than source words in the L2. They thus conclude that the main factor triggering 'emotion-related language choice' (ERLC; Kim and Starks 2008), with students feeling freer to swear in their L2, are social and cultural norms. According to Gawinkowska, Paradowski, and Bilewicz (2003: 5), 'the foreign language exempts us from our own socially imposed norms and limitations and makes us more prone to swearing and offending others'. Although the authors note that these findings may be extended to all people who know a foreign language to a 'communicative level', the extent to which this is true remains to be seen empirically.

This is also true of the conclusions drawn by other research that highlights an increase in psychological distance inducing utilitarianism when bilingual speakers are asked to make moral judgements in their L2—for instance, when answering a question such as: 'Would you sacrifice a man to save five?' (Costa et al. 2014). Further research is required to empirically verify the extent

to which these conclusions can be applied, because bilingualism is a complex and dynamic process that sees both languages always being active and interacting in the bilingual mind (for an overview, see Bialystok, Craik, and Luk 2012). The interaction of the two languages results in significant linguistic changes for bilingual speakers which set them apart from monolingual speakers of either language—or, as Grosjean (1989) famously observed, bilingual speakers are not the sum of two monolinguals in one. Crucial to the present discussion are the changes experienced by late bilingual speakers in the understanding and use of their L1, usually referred to as L1 'attrition' (for an overview of works in the field from different perspectives, see, among others, Schmid 2016; Sorace 2020; Gallo et al. 2021; Zingaretti 2022; Zingaretti et al. forthcoming). In particular, despite the general consensus on the importance of emotions in L1 attrition, the links between emotions and L1 attrition remain largely understudied, and neither Gawinkowska, Paradowski, and Bilewicz (2003) nor Costa et al. (2014) take L1 attrition into account in their investigations.

One of the few studies attempting to bridge this gap in the field is research by Kim and Starks (2008), which investigates emotions in L1 attrition and L2 acquisition in a group of 30 Korean–English L1-dominant late bilinguals in New Zealand. The results of a story-retelling task, a questionnaire, and a follow-up interview with their participants show a shift from the L1 to the L2 related to an increase in L2 fluency and a decrease in L1 accuracy. Indeed, despite the overall preference for ERLC to be in the L1 rather than the L2, there is a considerable amount of L2 use, particularly when anger-related emotion is involved, as well as correlations between most measures of ERLC with decreasing L1 accuracy and increasing L2 fluency. Ultimately, Kim and Starks note that these results call for an urgent need for dual language support for young Korean immigrants in New Zealand: for these immigrants, the increase in socialisation in the L2 may come with delayed or 'primitive' socialisation in the L1, which may ultimately result in the attrition (here intended as 'loss', due to disuse) of the language in later stages of their lives.

Overall, while there seem to be few differences in terms of the emotional resonance of L1 and L2 words for early and simultaneous bilinguals (see Harris 2004), the same cannot be said for late bilinguals, whose relationships with emotions in the different languages they speak are shaped by a variety of different factors, and ultimately give rise to highly individual preferences when expressing emotions in either language, as seen thus far. To better understand why bilingualism may also influence speakers' perception of hate speech, we now review some of the literature on the relationship between emotions and hate speech.

6.2.2 Hate speech and emotions

The relationship between hate speech and emotions is similarly complex, starting from the very emotion and feeling that is evoked by the terminology itself (i.e. hate) or, as some scholars somewhat drastically put it, 'the myth of hate' (Brown 2017a, 2017b). Brown (2017a) observes that hate speech, in its ordinary (rather than legal) meaning, does not correspond to a single monolithic phenomenon, but to a diverse set of expressive phenomena that may not necessarily involve hate in its most distinctive quality of intense or extreme dislike. Indeed, according to Brown, there are many occasions on which the terms 'hate speech' or 'hate crime' may be used where no hate or hatred is involved. One of the examples proposed is that of a fundamentalist or evangelical Christian directing the following words at people entering a gay club on the street: 'You homosexuals and lesbians are sinners in the eyes of God, you disobey the teachings of the Bible, and for this reason you will go to hell if you do not repent' (Brown 2017a: 450). Rather than expressing emotions, feelings or attitudes of hate or hatred, these words serve as an expression of religious belief or feelings of disgust or repulsion learnt from parents or community leaders; however, expressions like these, as well as others, can still be classified as hate speech on the basis that they are forms of speech that carry prejudiced messages (here, as homosexuals are portrayed as

morally inferior beings) or hate speech acts that rank, degrade, harass, or persecute someone (homosexuals in this case).

Brown (2017b: 562) goes further, saying that hate speech should not be interpreted as a compositional concept made up of the literal meanings of the words 'hate' and 'speech', but as an equivocal idiom, or rather 'a family of meanings, for which there is no one overarching precise definition available'. Although the aim of this chapter is not to discuss the multifarious ways in which hate speech can manifest, it seems important to emphasise that the phenomenon—or rather 'phenomena', following Brown's logic—under investigation here entails a complexity of emotions and feelings that go beyond mere hate. Acknowledging that hate speech entails complex emotions and feelings is ultimately crucial in identifying it successfully, as more recent research in natural language processing shows that the emotions encoded in sentic computing sources and semantically structured hate lexicons help to detect forms of online hate speech more accurately (Chiril et al. 2022).

Importantly, emotions are not only part of the source—they are also generated in the listener as a result of being targeted by hate speech. In this respect, borrowing from Carey's (1989) transmission model of communication, Calvert (1997) discusses the emotional and physical harms of hate speech:

> The question of harm caused by hate speech, when considered from the perspective of the transmission model, boils down to this: Did communication of a particular message, X, cause a change, Y, in the attitude or behavior of the recipient of the message? Does a bigot's calling the African-American standing next to him in line at the movie theater a 'nigger' cause the African-American's *pulse rate to increase* or his stomach to *feel nauseated*? Does it cause him *mental pain and anguish* or make him feel *angry*? Does it cause him to strike the speaker? Each affirmative response is a direct physical or emotional change caused by a particular message. (Calvert 1997: 10; emphasis added).

The important point highlighted by Calvert (1997: 10) is that law courts deem emotional harms to be intangible unless accompanied

by proof of physical harms, and that the physical symptoms experienced by victims of hate speech (e.g. increased pulse rate or breathing difficulties) do not 'lend themselves to proof of harm at trial', while, in fact, these harms are very much real.

The ways in which hate speech impacts targets emotionally are significant and widely documented. For instance, Gelber and McNamara (2016) investigate the harms of hate speech as evidenced by the experiences of 101 members of indigenous and minority ethnic communities in Australia. Their research demonstrates multiple types of harms reported by the participants, among which are hurt and upset feelings that are sometimes deep enough to be perceived as 'existential' pain. Hate speech also has the power to silence targets, rendering them unable to respond directly, and at times causing them to withdraw from the situation altogether. Feelings of exclusion, dehumanisation, anger, and frustration, alongside the inability to identify with one's own ethnicity or religion, are also reported as consequences of being targeted by hate speech, and, at times, also as the consequence of the perpetuation of negative stereotypes by the media (Gelber and McNamara 2016: 334–335). Similar results are also reported in research analysing Asian American students' responses after reading second-hand accounts of hate speech: not only do the students suffer what Boeckmann and Liew (2002: 377) define as a '(presumably) temporary reduction in collective self-esteem', but, according to the researchers, being the direct target of similar speech in real-life scenarios would surely result in 'more extreme and enduring consequences'. Both Gelber and McNamara (2016) and Boeckmann and Liew (2002) agree that the significant emotional and psychological effects of hate speech constitute evidence in favour of sanctions against it due to its extremely harmful impact.

Clearly, then, the relationship between hate speech and emotions is complex, and the impact that hate speech has on its targets can result in significant emotional and psychological harms: feelings of exclusion, dehumanisation, anger, frustration, reduced identification with one's own ethnicity or identity, and reduced self-esteem are all reported in the literature, alongside silence and withdrawal from

the situation as common reactions among victims of hate speech. We will now explore some accounts that seek to explain another possible reaction to hate speech—namely, slur appropriation.

6.2.3 Slur appropriation as a pragmatic response to hate speech

From a linguistic perspective, specific pragmatic strategies related to hate speech reactions have been investigated. One of these strategies is 'slur appropriation'—that is, the use of a slur usually targeting a specific group by that very same targeted group, for non-derogatory purposes within the group (Hom 2008). According to Potts (2007: 10): 'When lesbian and gay activists use the word "queer", its meaning (its expressive content) differs dramatically from when it is used on conservative talk radio'. According to Hom (2008) there are multiple reasons why in-group members may wish to appropriate a slur. Not only does slur appropriation allow in-group members to take back the instrument of discrimination from the discriminators, but in doing so the appropriators also soothe and neutralise the originally offensive effect of the slur. Importantly, slur appropriation allows members to demarcate their (in-)group and show solidarity among other members, while highlighting that they are still objects of discrimination. Hornsby (2003) further explains that in-groups critically position themselves against normal (i.e. derogatory) uses of the slur, and adds to Hom's 'soothing' and 'neutralising' objectives that of subverting the old, non-descriptive meanings of the slur being appropriated.

Appropriation, reversal, and subversion are conceived by Bianchi (2014) as 'echoic uses'. To elaborate further, in echoic uses speakers not only report the utterances or thoughts of others, but also convey their own attitude in regard to those utterances or thoughts: by appropriating slurs, then, in-groups echo derogatory uses in ways that make explicit the dissociation from the offensive content originally conveyed by the slur. The effect, Bianchi suggests, is ironic: slur appropriators mock those who make use of the slurs in a derogatory way, by only mentioning part or some of

the constituents of the attributed utterance or thought, with a tacit echo and dissociative attitude that has to be understood from the context, facial expression, tone of voice, or other paralinguistic cues. Lastly, Bianchi notes that over time, appropriated uses may extend from in-group use only to out-group use as well, which is what has happened to words such as 'queer' or 'gay' today, as they have lost their hints of echo or irony, and are openly used without connotation by people who do not identify as such.

The relevance of slur appropriation to the present study becomes clear when we consider that most of our knowledge about how slur appropriation works, and about the links between hate speech and emotions more generally, comes from monolingual research. Crucially, since bilingual speakers are part of two linguistic and cultural communities, they may be better able to manipulate the polarity of a term, particularly due to their reportedly enhanced ability to understand behaviours that differ from their own (see the work on Theory of Mind in adult bilinguals by Navarro and Conway 2021). Moreover, given the heavy emotional load associated with L1 slurs due to first emotion experience (see Section 6.2.1 on bilingualism and emotions), it may be easier for bilinguals to appropriate slurs in the L2 than in the L1. The extent to which this is true requires empirical observation, which the research contained within the present chapter carries out.

6.2.4 Focus of the previous literature

In brief, the research presented so far demonstrates clear links and complex relationships between i) bilingualism and emotions, as well as ii) hate speech and emotions. On the one hand, bilinguals commonly report feeling different in the different languages they speak for a variety of reasons (e.g. L2 learning contexts and age, L2 proficiency and educational levels) and they often choose to express their emotions in a language due to a range of factors (e.g. medium of instruction in the language, socialisation, number of interlocutors, and cultural specificity). L1 attrition also contributes to shaping bilingual speakers' language choice when

expressing emotions, although the extent of its role requires further scrutiny. On the other hand, the impact of hate speech on the individual can be extreme in terms of emotions, with feelings of exclusion, dehumanisation, anger, frustration, reduced identification with one's own ethnicity or identity, and reduced self-esteem all being reported in the literature, together with silence and withdrawal from hate speech situations.

However, the evidence we have regarding the emotional and psychological impact of hate speech, as well as what we know regarding the possible reactions to hate speech through slur appropriation, has only been gathered within monolingual contexts. Indeed, no study to date has sought to investigate the ways in which bilingual speakers perceive an emotionally loaded topic such as hate speech in their different languages, or the possibly different reactions that bilingual speakers may have when they are the target of hate speech in either of their languages. This research thus represents an initial attempt at addressing this gap.

6.3 Methodology of the present study

6.3.1 Materials and design

This study gathered data through an online questionnaire administered on the Qualtrics XM platform (https://www.qualtrics.com) after obtaining ethics approval from the School of Philosophy, Psychology and Language Sciences Research Ethics Committee at the University of Edinburgh (ref. 377-2021/7). The online questionnaire employed in the study required participants to obligatorily answer three main blocks of questions and it gathered participants' background information through optional final questions. The first part of the questionnaire collected qualitative data regarding participants' experiences with hate speech during their adolescence and in the preceding five years. The second and third parts constitute the experimental block of the study; they gathered participants' quantitative responses on ratings of word pairs and reactions to hate speech scenarios. The materials used and data

gathered in all three parts of the study can be found on the project page on the Open Science Framework website (https://osf.io/fbtmy/). In this chapter, we focus on the design, results, and discussion of the data gathered in the experimental block of the study.

The word pair ratings included the ten word pairs shown in Table 6.1, which were presented to our participants first in Italian and then in English in randomised order. Following what has already been carried out by other studies on monolingual speakers (see, for English, Janschewitz 2008; for Italian, Sulpizio et al. 2020), we asked participants to evaluate each of the words on a Likert scale from 1 to 9, for the following seven criteria in each language: *familiarity* (i.e. how familiar they were with the term presented, where 1 = never having encountered the word, and 9 = encountering it all the time); *personal use* (i.e. how often they used the term, where 1 = never having used the word, and 9 = using it all the time); *imageability* (i.e. how easily the term evoked an image for the participants, where 1 = no image being evoked at all, and 9 = image being evoked easily); *arousal* (i.e. how much the term stimulated their attention, where 1 = not finding the term very stimulating, and 9 = finding it very stimulating); *offensiveness* (i.e. how offensive the term was perceived to be, where 1 = not finding the word offensive, and 9 = finding it extremely offensive); *tabooness* (i.e. how taboo the term was perceived to be, where 1 = not recognising the word as taboo, and 9 = recognising it as a major taboo); and finally *valence* (i.e. how positive or negative they found the term, where 1 = finding the word very unpleasant, and 9 = finding it very pleasant).

With regard to the reactions elicited by reading scenarios containing hate speech, inspired by similar work in the field with monolingual speakers (see Boeckmann and Liew 2002), we designed four different scenarios in each language, for a total of eight scenarios presented to our participants. Further, in each language, two of the scenarios contained the metalinguistic use of the same slur used to target the victim by the very same victim in response to hate speech (in a similar though not identical way to how slur appropriation works, as described in Section 6.2.3).

Table 6.1: Italian–English word pairs rated by the participants of the study.

Italian	English
Testa di cazzo	Dickhead
Stronzo/a	Asshole
Spastico/a	Spastic
Ritardato/a	Retard
Troia	Whore
Puttana	Slut
Cagna	Bitch
Negro/a	Nigger
Frocio	Faggot
Lella	Dyke

Note: The last word pair (*lella*, dyke) was removed from the analysis because many respondents reported never having heard the Italian word (probably due it being mainly used within the Rome area).

This is illustrated, for instance, in (1), where the female victim of hate speech responds to the slur uttered by the person sitting next to her reclaiming the same slur. This affects the person who insulted her, who leaves, embarrassed:

(1) An African American lady is talking on the phone with her son on the tram. When the tram stops, an old woman gets on board and sits down next to her while she keeps on speaking to her son on the phone. As the old woman starts to get irritated, she looks at the lady and says: 'You niggers just know how to shout, don't you?' As soon as the lady hears the woman's comment she looks away from her phone, staring straight into the woman's eyes, and shouts: '*Of course us niggers know how to shout*. We've got to make sure people hear us when there's some racist around us!' The old woman then stands up and leaves, feeling completely embarrassed. (Emphasis added)

The remaining two scenarios did not include slur appropriation—as in (2), where a young woman subject to hate speech keeps quiet as a result:

(2) A young woman is heading to her best friend's wedding on the train. She is wearing a low-cut dress which shows her cleavage. When she gets up to go to the bathroom, she sees two women staring at her. One of them says to the other: 'Have you seen her?! With that dress, she looks like such a slut'. The young woman notices the two. *She pretends not to have overheard anything and proceeds to walk towards the bathroom, feeling humiliated*. (Emphasis added)

After reading each scenario, participants were prompted to answer the following question: 'Do you find it appropriate to use the word X to respond to an insult (as in the example above),[2] and would you ever use a slur which refers to sexuality or ethnicity to respond to an insult yourself?' As we explain in Section 6.4, we analysed the responses that deemed (in)appropriate the use of an insult in response to an insult (i.e. the dependent variable we named 'appropriateness') separately from the responses relating to the reappropriation of the slur by the participants themselves (i.e. the dependent variable named 'appropriation').

6.3.2 Participants

For the purposes of this study, we aimed to recruit a group of late bilingual speakers in the UK—specifically, Italian speakers of English with high intermediate/advanced L2 proficiency, who had to have grown up in Italy and emigrated to the UK after 16 years of age. The participants were recruited online on social media platforms (e.g. on Facebook pages for Italian expatriates in the UK, Twitter, and Instagram) and by word of mouth. After discarding

2 'As in the example above' only appeared in questions relating to scenarios with slur appropriation.

Table 6.2: Participants' background information variables with either Mean (M) and Standard Deviation (SD) or percentage (%) for each variable.

Background variable	M and SD or %
Age	M = 31.2, SD = 7.94
Years	
Length of residence	
Years	M = 5.05, SD = 2.86
Language dominance	
Italian	83.72%
English	16.28%
Gender	
Woman	67.44%
Man	27.91%
Non-binary	4.65%
Sexual orientation	
Heterosexual	69.77%
Homosexual	11.63%
Bisexual	11.63%
Pansexual	2.33%
Other	4.65%
Ethnicity	
White	90.70%
Hispanic/Latinx	4.65%
Asian	2.33%
Middle Eastern	2.33%
Disability	
No	83.72%
Yes	13.95%
Rather not say	2.33%
Educational level	
University	79.06%
Secondary school	20.93%

6 participants who did not fit our initial requirements, our final sample consisted of 43 bilingual speakers, whose background information is summarised in Table 6.2.

Participants were instructed to complete the questionnaire in their own time as no time limitations were set, although they were not allowed to go back and change answers that they had

previously given. Upon completion, they were entered into a prize draw with a chance to win one of three different monetary prizes.

6.4 Results

Below we report our results for the word pair ratings and reactions to hate speech scenarios. All of our analyses were carried out using jamovi (The jamovi project 2021), an open-source statistical software based on the R programming language (R Core Team 2021). For optimal understanding, our results are presented in two complementary formats. First, we provide an accessible explanation, facilitating comprehension for a broad audience irrespective of their statistical proficiency. The initial, non-technical explanation is then followed by detailed statistical results, catering to those keen on methodological specifics. Corresponding figures and tables are provided to visually aid the interpretation of our results.

6.4.1 Word pair ratings

To begin, our investigation into the various word characteristics rated by our participants revealed that participants were significantly more familiar with the Italian words than they were with the corresponding words in English. In terms of other characteristics (personal use, imageability, arousal, offensiveness, tabooness, and valence), there were no notable differences in ratings between languages.

Precisely, paired samples Student's *t*-tests (or Wilcoxon signed-ranks tests where data failed to meet normality assumptions) were run to compare ratings for the same pairs of criteria between languages (e.g. familiarity in Italian and familiarity in English, personal use in Italian and personal use in English). The results are reported in Table 6.3 for familiarity (Fam.), personal use (P.use), imageability (Ima.), arousal (Aro.), offensiveness (Off.), tabooness (Tab.) and valence (Val.) in Italian and English with Mean (M), Median (Mdn), Standard Deviation (SD), degrees of freedom (df) and *p* values.

Table 6.3: Student's t-tests and Wilcoxon signed-ranks tests for word pair ratings.

	Italian			English						
	M	**Mdn**	**SD**	**M**	**Mdn**	**SD**	**Test type**	**Stat**	**df**	***p***
Fam.	5.58	—	1.73	4.67	—	1.52	Stud.'s *t*	3.89	42	<.001
P. use	2.49	—	0.85	2.28	—	0.66	Stud.'s *t*	1.54	35	0.132
Ima.	4.82	—	1.99	4.68	—	2.12	Stud.'s *t*	0.49	42	0.621
Aro.	—	5.11	—	—	4.56	—	Wilcox. W	473	—	0.249
Off.	—	6.44	—	—	6.22	—	Wilcox. W	580	—	0.198
Tab.	—	4.67	—	—	4.89	—	Wilcox. W	339	—	0.648
Val.	—	2.78	—	—	3.11	—	Wilcox. W	395	—	0.950

As can be seen, the only between-language rating comparison that showed a statistically significant difference was familiarity: that is, participants were significantly more familiar with words in Italian (M = 5.58, SD = 1.73) than in English (M = 4.67, SD = 1.52), $t(42) = 3.89$, p <.001. The remaining comparisons—personal use, imageability, arousal, offensiveness, tabooness, and valence—were not significant between languages (p >.05).

In addition to comparing the ratings in both languages, we also examined how each criterion related to the other between languages. In other words, we checked whether individuals who rate a word highly for a particular characteristic in one language, such as familiarity, also rate it highly for the same characteristic in the other language. We found that this is generally the case. For instance, if a participant was familiar with a word in Italian, they also tended to be familiar with the corresponding word in English. The same was true for all the other characteristics examined: personal use, imageability, arousal, offensiveness, tabooness, and valence.

Elaborating further, Pearson correlation analyses (or Spearman rank-order correlations where data failed to meet normality

assumptions) were run to explore the relationship of each criterion between languages (e.g. the relationship between familiarity in Italian and familiarity in English, personal use in Italian and personal use in English), as well as across criteria within languages (e.g. familiarity in Italian and personal use in Italian, familiarity in English and personal use in English). The results are reported in Table 6.4. As can be seen, we found positive moderate-to-strong correlations for all criteria between languages, indicating that high ratings for each criterion in one language corresponded to high ratings for the same criterion in the other language (e.g. high ratings for familiarity in Italian corresponded to high ratings for familiarity in English, high ratings for personal use in Italian corresponded to high ratings for personal use in English, and so on for each pair of criteria).

Moreover, we examined how different criteria related to each other within a single language, either Italian or English, discovering interesting patterns. For instance, in both languages, the words people were more familiar with were also the ones they used more often, and the words deemed more taboo were also viewed as more offensive. Additionally, again in both languages, words that grabbed more attention (arousal) were seen as more offensive and were also more likely to bring a clear mental picture (imageability). Specific to English, participants tended to use offensive words less frequently, and words that grabbed attention (arousal) were more often labelled as taboo. On the other hand, specific to Italian, words that easily brought a mental picture (imageability) were seen as more familiar and more offensive.

To unpack the specifics, we found correlations of different strengths across criteria within a language. Among the within-language correlations reported in Table 6.4, we found: positive correlations between familiarity and personal use in both Italian, $r_s(41) = .435$, $p = .004$, and English, $r_s(41) = .415$, $p = .006$, meaning that in both languages words rated as more familiar were also rated as being more used; positive correlations between tabooness and offensiveness in both Italian, $r_s(41) = .431$, $p = .004$, and English, $r(41) = .507$, $p < .001$, meaning that words rated as being

Table 6.4: Correlation matrix for all variables in Italian (I) and English (E) with Pearson's r (r) and Spearman's rho (ρ) values.

		Fam. (I)	Fam. (E)	P. use (I)	P. use (E)	Ima. (I)	Ima. (E)	Aro. (I)	Aro. (E)	Off. (I)	Off. (E)	Tab. (I)	Tab. (E)	Val. (I)	Val. (E)
Familiarity (I)	r	—													
	ρ	—													
Familiarity (E)	r	.553***	—												
	ρ	—	—												
Personal use (I)	r	—	—	—											
	ρ	.435***	.192	—											
Personal use (E)	r	—	—	—	—										
	ρ	.383*	.415*	.465**	—										
Imageability (I)	r	.512***	.308*	—	—	—									
	ρ	—	—	.149	.041	—									
Imageability (E)	r	.208	.189	—	—	.644***	—								
	ρ	—	—	.128	-.021	—	—								
Arousal (I)	r	.352*	.134	—	—	.495***	.479**	—							
	ρ	—	—	.039	-.153	—	—	—							
Arousal (E)	r	.109	-.049	—	—	.467**	.611***	.641***	—						
	ρ	—	—	.064	-.238	—	—	—	—						
Offensiveness (I)	r	—	—	—	—	—	—	—	—	—					
	ρ	.019	-.072	-.053	-.107	.346*	.27	.363*	.336*	—					
Offensiveness (E)	r	.028	-.142	—	—	.306*	.251	.32*	.384*	—	—				
	ρ	—	—	-.028	-.34*	—	—	—	—	.639***	—				
Tabooness (I)	r	-.001	-.182	—	—	.047	.021	.192	.036	—	.466**	—			
	ρ	—	—	-.06	-.216	—	—	—	—	.431**	—	—			
Tabooness (E)	r	.072	-.119	—	—	.034	.129	.358*	.361*	—	.507***	.584***	—		
	ρ	—	—	.55	-.221	—	—	—	—	.319*	—	—	—		
Valence (I)	r	—	—	—	—	—	—	—	—	—	—	—	—	—	
	ρ	.068	-.106	.251	.009	-.121	-.008	.133	.161	-.046	-.061	-.094	.026	—	
Valence (E)	r	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	ρ	.084	-.000	.207	.017	-.286	-.042	.048	.06	-.133	-.183	-.087	.085	.626***	—

Note: * p <.05, ** p <.01, *** p <.001

more taboo were also rated as being more offensive in both languages; positive correlations between arousal and offensiveness in both Italian, $r_s(41) = .363$, $p = .017$, and English, $r(41) = .384$, $p = .011$, and also between arousal and imageability in both Italian, $r(41) = .495$, $p < .001$, and English, $r(41) = .611$, $p < .001$, indicating that words that stimulated our participants' attention more were also rated as being more offensive and more easily evoked in their minds in both languages; in English only, a negative correlation between offensiveness and use in English, $r_s(41) = -.340$, $p = .026$, indicating that high ratings for offensiveness in our participants' L2 corresponded to low ratings for personal use of the same words in the same language, but not in their L1; also only in English, a positive correlation between arousal and tabooness, $r(41) = .361$, $p = .018$, meaning that more stimulating words were also rated as being more taboo in the L2; finally, in Italian only, positive correlations between imageability and familiarity, $r(41) = .512$, $p < .001$, and imageability and offensiveness, $r_s(41) = .346$, $p = .023$, indicating that words that more easily evoked images in our participants' minds were also rated as being more familiar to them and more offensive in their L1.

Lastly, we wanted to understand if LoR in the UK and language dominance could predict the ratings for each criterion in each language. To summarise, we found that the length of time participants had lived in the UK—not their language dominance—predicted their familiarity, their personal use, and how vividly they could picture the meanings of Italian words. Specifically, for Italian, a longer LoR predicted lower familiarity, lower personal use, and lower imageability. On the other hand, for English, neither LoR nor language dominance could predict familiarity, personal use, or imageability. However, the LoR in the UK did predict how offensive participants found English words, with longer LoR predicting lower perceived offensiveness. None of the factors we tested could significantly predict arousal, tabooness, and valence in either language.

Going into further detail, multiple linear regressions were run to investigate whether each of the seven criteria rated by our

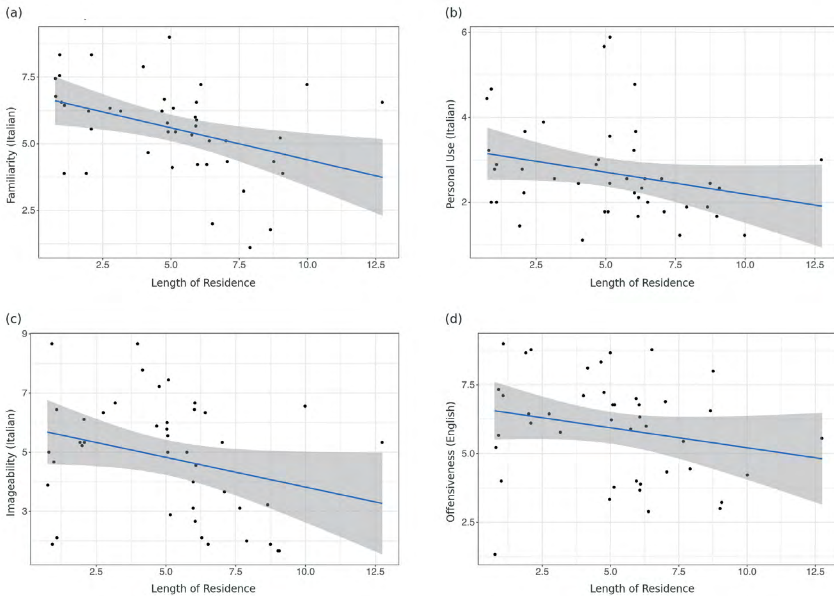


Figure 6.1: Scatter plots with regression lines and shaded standard errors, showing how length of residence predicts familiarity in Italian (a), personal use in Italian (b), imageability in Italian (c) and offensiveness in English (d).

participants in each language (e.g. familiarity in Italian, familiarity in English, personal use in Italian, personal use in English) could be predicted by LoR, language dominance (i.e. Italian/English), and their interaction. The effects of significant predictors are shown in Figure 6.1.

First, familiarity in Italian was significantly predicted by our model, $F(2, 40) = 3.84$, $p = .030$, $R^2 = .161$. While LoR ($\beta = -.379$, $p = .015$) added significantly to the prediction—see Figure 6.1(a)—dominance ($\beta = .183$, $p = .650$) was not a significant predictor in the model. In English, instead, familiarity was not significantly predicted by our model and none of our predictors added significantly to the prediction ($p > .05$). Second, personal use in Italian was not significantly predicted by our model, $F(2, 33) = 3.06$, $p = .060$, $R^2 = .157$; however, LoR ($\beta = -.392$, $p = .021$), unlike

dominance (β = -.348, *p* =.455), did add significantly to the prediction—see Figure 6.1(b). On the other hand, personal use in English was not significantly predicted by our model and none of our predictors added significantly to the prediction (p >.05). Third, imageability in Italian was significantly predicted by our model, $F(3, 30) = 2.90$, p =.047, R^2 =.182. LoR (β = -.392, *p* =.021) added significantly to the prediction—see Figure 6.1(c); however, neither dominance (β = -.180, *p* =.142) nor LoR ∗ dominance (β = 1.20, *p* =.071) were significant predictors in the model. While imageability in English was also significantly predicted by our model, $F(3, 39) = 3.00$, p =.042, R^2 =.188, none of the variables added significantly to the prediction (p >.05). Fourth, offensiveness in Italian was not significantly predicted by our model and none of the variables added significantly to the prediction (p >.05). On the other hand, although offensiveness in English was not significantly predicted by our model either, $F(3, 39) = 1.92$, p =.143, R^2 =.128, LoR (β = -1.41, *p* =.036) did add significantly to the prediction, while dominance (β = -.619, *p* =.080) and LoR ∗ dominance (β = 1.26, *p* =.066) were not significant predictors in the model. None of our remaining models for arousal, tabooness, and valence were significant in Italian or English, with no variable adding significantly to the predictions (p >.05).

6.4.2 Reactions to hate speech scenarios

Looking at reactions to hate speech scenarios, we wished to investigate whether different factors could predict whether bilinguals would consider the appropriation of slurs as an appropriate or inappropriate choice. These factors included the language used, LoR, language dominance, ethnicity, sexuality, and a combination of these factors. Overall, our model showed that only sexuality played a significant role, with individuals identifying as a non-heterosexual being more likely to deem the appropriation of slurs as acceptable.

To delve deeper into the analysis, we ran a first binomial logistic regression to predict the likelihood that bilinguals would perceive

slur appropriation as an appropriate rather than inappropriate choice (i.e. 'appropriateness', yes/no) by language (i.e. L1/L2), LoR, dominance (i.e. Italian/English), ethnicity (i.e. white/non-white), sexuality (i.e. heterosexual/non-heterosexual), and the interaction of language with any of these variables.[3] The overall model was significant, χ^2 (9) = 17.65, p =.039, with between 9.7 per cent and 13.8 per cent of the variance in the odds of appropriateness explained by the predictor set. Across both outcome categories, 72.7 per cent of cases were accurately classified, with sensitivity lower than specificity. Answers deeming slur appropriation an appropriate choice were correctly predicted in 25.5 per cent of cases compared to 92.6 per cent of answers deeming it inappropriate. Sexuality (p =.025) was a significant predictor in the model, as mentioned above, with participants identifying with a sexual minority (i.e. non-heterosexual) being more likely to perceive slur appropriation as an appropriate compared to those not identifying with a sexual minority (OR = 3.44, 95% CI [1.16, 10.18])—see Figure 6.2. None of the other variables or interactions added significantly to the prediction (p >.05).

Lastly, we also wanted to determine the chances of our bilingual participants appropriating slurs themselves when responding to insults on the basis of the same factors used in the previous model—namely, language used, LoR, language dominance, ethnicity, sexuality, and a combination of these. Our findings suggest that, out of these factors, only participants' sexuality and ethnicity played a crucial role, as those who identified as part of a sexual or ethnic minority were more likely to appropriate slurs in their response to insults.

Expanding on this, we ran a second binomial logistic regression to predict the likelihood that bilinguals would appropriate slurs in response to an insult (i.e. 'appropriation', yes/no) by language (i.e. L1/L2), LoR, dominance (i.e. Italian/English), ethnicity

3 To reduce the number of levels of the predictor variables, we decided to group all ethnic minorities as 'non-white' and sexual minorities as 'non-heterosexual'.

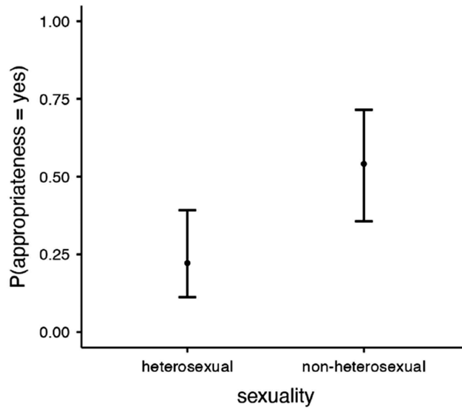


Figure 6.2: Effects plots with 95 per cent confidence interval showing how sexuality significantly predicts the likelihood of perceiving slur appropriation as appropriate.

(i.e. white/non-white), sexuality (i.e. heterosexual/non-heterosexual), and the interaction of language with any of these variables. The overall model was significant, χ^2 (9) = 32.69, p <.001, with between 17.3 per cent and 34.5 per cent of the variance in the odds of appropriateness explained by the predictor set. Across both outcome categories, 92.4 per cent of cases were accurately classified, with sensitivity lower than specificity. Answers deeming slur appropriation an appropriate choice were correctly predicted in 36.8 per cent of cases compared to 99.3 per cent of answers deeming it inappropriate. Sexuality (p =.008) and ethnicity (p =.044) were significant predictors in the model, with participants identifying with a sexual minority being more likely to appropriate slurs in response to hate speech compared to those not identifying with a sexual minority (OR = 25.71, 95% CI [2.32, 284.73]), and participants identifying with an ethnic minority being more likely to appropriate slurs in response to hate speech compared to those not identifying with an ethnic minority (OR = 9.36, 95% CI [1.05, 82.99])—see Figure 6.3. None of the other variables or interactions added significantly to the prediction (p >.05).

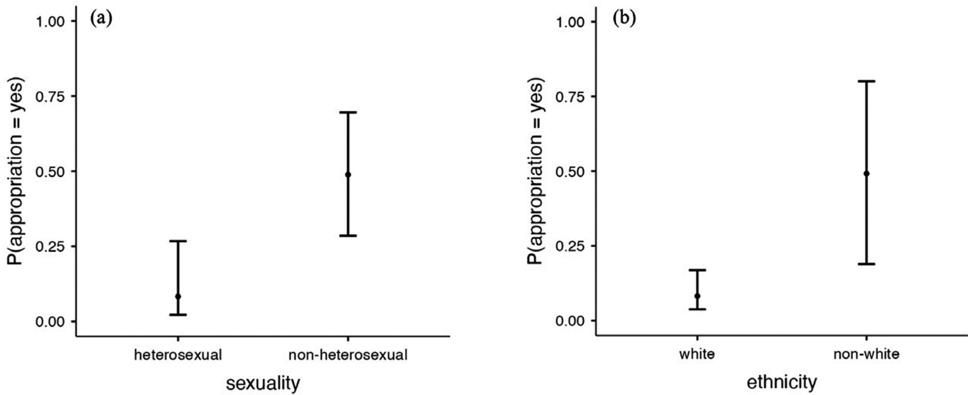


Figure 6.3: Effects plots with 95 per cent confidence interval showing how (a) sexuality and (b) ethnicity significantly predict the likelihood of slur appropriation in response to hate speech.

6.5 Discussion

Our initial analyses for word pair ratings reveal that the only significantly different criterion in the two languages is familiarity, with participants reporting greater familiarity with hate words in L1 Italian than L2 English despite their relatively long period of residence in the UK and high level of proficiency in L2 English. The remaining comparisons returned non-significant results, meaning that our participants' ratings for personal use, imageability, arousal, offensiveness, tabooness, and valence are similar between languages. These findings seem to indicate that, despite the higher familiarity with L1 words, hate words in the different languages are perceived rather similarly for our participants. This is further corroborated by the positive correlations found between all of the same criteria between languages, demonstrating that high ratings in one language for one criterion correspond to equally high ratings for the same criterion in the other language. Different criteria were also related within each language: while some highlight easily interpretable relationships (e.g. the negative correlation between perceived offensiveness and personal use in English, with

highly offensive terms being used less, and vice versa) some reveal an intricate relationship between criteria, pointing to potential difficulties for participants—thus potential confounds for similar experiments—in teasing some of these aspects apart (specifically, due to the correlations of arousal with familiarity, offensiveness, tabooness, and imageability, and of imageability with familiarity and offensiveness).

Importantly, the results of our multiple regression models indicate that LoR—but not dominance, contrary to our predictions—has a significant effect on familiarity, personal use, and imageability in Italian, as well as on offensiveness in English (see Figure 6.1). Specifically, participants residing longer in the L2 country reported lower ratings for familiarity, personal use, and imageability in the L1, suggesting that moving away from the L1 country may have an impact not only on the amount of L1 use but also on the levels of familiarity with L1 words and the ease with which L1 words evoke an image in bilingual speakers' minds. This is in line with the potential linguistic changes that happen in bilingual speakers' L1 as a result of 'the co-activation of language, crosslinguistic transfer or disuse' (Schmid and Köpke 2017: 637; see also the works on L1 attrition previously cited). At the same time, an increase in LoR also corresponds to a significant decrease in the degree of offensiveness perceived in the L2—a finding that seems to suggest that the longer speakers reside in the L2 country (and possibly, the more accustomed they become to L2 sociocultural norms), the less offensive L2 words become for them. This expands on related findings in Gawinkowska, Paradowski, and Bilewicz (2013) and Costa et al. (2014), by demonstrating that the supposedly reduced emotional response in the L2 varies as a function of LoR (and, arguably, proficiency in the L2). The overall non-significant models for both personal use in Italian and perceived offensiveness in English, however, call for further research into the impact of LoR and related variables on hate speech perception in both the L1 and the L2.

On the other hand, contrary to our predictions regarding the effects of bilingualism on slur appropriation, our analyses reveal

that, among our participants, the factors related to the bilingual experience (i.e. language, LoR, dominance, and the interaction of language with each of these factors) do not predict the likelihood of deeming slur appropriation as an appropriate choice, nor the likelihood of them appropriating those slurs (for why this may be the case, see Section 6.5.2). Instead, identifying as part of a minority has a significant effect on both of these: interestingly, while it is only non-heterosexual participants who deem slur appropriation as a significantly more appropriate choice than heterosexual participants (see Figure 6.2), participants identifying with sexual and ethnic minorities are significantly more likely to appropriate slurs themselves as a reaction in situations where they are the target of hate speech (see Figure 6.3). These findings are overall in line with our predictions regarding the effect of identifying with minorities, and the mechanism of slur appropriation (see Bianchi 2014, among other works previously cited).

6.5.1 Implications

The present study has important implications for research in both linguistics and other disciplines. From a linguistic perspective, the research within this chapter represents a first attempt at understanding how bilingualism can influence the way hate speech is perceived, and the way bilingual speakers react when targeted by it. Specifically, this study shows the effects of LoR on familiarity, use, and imageability of hate words in the L1, and on perceived offensiveness in the L2. It also reports the absence of bilingualism-related effects on slur appropriation, at least when analysing behavioural data (but see Section 6.5.2 below). From an interdisciplinary point of view, this chapter highlights the importance of considering bilingualism (and, more generally, psycholinguistic) research findings for disciplines dealing with hate speech from other angles—for instance, philosophy and ethics, as well as law and politics. Slagle (2009: 246) states that one of the difficulties reported in punishing hate speech is the fact that, with regards to emotional distress, 'such damage can never be quantified or even

proven'. Studies such as ours, as well as the evidence reported in our background section (see Section 6.2), suggest otherwise: the impact of hate speech on a psychological and emotional level is very much real (see Calvert 1997), has several consequences (see Gelber and McNamara 2016), some of which may be long-term (see Boeckmann and Liew 2002: 377), and may actually be influenced by the bilingual experience. Further, the fact that bilingual speakers may perceive hate speech as more or less emotional in their different languages has important implications for law and politics. Imagine two different scenarios: one where a bilingual speaker who no longer lives in their L1 country is insulted by a speaker of their L1 during a holiday in their home country; and another where a bilingual is summoned for jury duty in their L2 country and is asked to evaluate how offensive a hate speech scenario is. In both situations, different factors relating to the bilingual experience of the speakers will impact the perceived degree of offensiveness (and related dimensions) in each language. It is clear, then, that awareness of the bilingual experience factors that influence hate speech perceptions and reactions is crucial, as is further research into these factors.

6.5.2 Limitations and recommendations for future research

Before concluding, it is important to address some of the limitations of the present study that can inform future research on bilingualism and hate speech. Firstly, the number and type of word pairs employed here included both more 'embodied' (e.g. dickhead, asshole) and less embodied (e.g. nigger, faggot) insults; as some research points to greater recall of more embodied insults (Wellsby et al. 2010), this may have an impact on some of the variables rated by our participants. Future research could avoid potential confounds by either investigating ratings for a single term, or by employing equally large numbers of more/less embodied insults, and adding this factor into the analysis. Similarly, the number and type of scenarios employed here could be extended

to include more scenarios of different types. In particular, slur appropriation was presented here as a reaction to hate speech (i.e. in scenarios where participants were the direct target of slurs). It would be interesting to investigate whether similar ratings arise in contexts where speakers are appropriating slurs without being targeted in that specific scenario (e.g. a man referring to himself as a 'faggot' when talking with a friend).

The number of participants in the study also represents a limitation, particularly given the relatively unbalanced proportion of participants identifying with ethnic and sexual minorities compared with those who did not (see Table 6.2). Future research could include more participants, which would also allow for a larger number of possible explanatory variables in the models, as well as increasing power in the analyses. Indeed, with specific regard to the predictor variables used in this study, it must be noted that, although it did not contribute significantly to any of the models, language dominance was collected as a self-reported measure. A better measure of language dominance in the future may consider different predictors that include personal background factors and exposure, use, and attitudes, relating to both the L1 and L2 of the participants (as suggested, for instance, in Schmid and Yılmaz 2018). Importantly, given the significant effect of LoR as a predictor in both L1 familiarity, use, and imageability, and in L2 perceived offensiveness, it would be informative to gather linguistic information on participants' L1 and L2 (i.e. by collecting accurate proficiency measures in the two languages, and investigating attrition widely in more than one L1 domain).

Finally, it is important to consider the nature of the study itself: in this research, behavioural data was collected via an online questionnaire; while participants were not able to change an answer once it had been provided, they could complete the questionnaire in their own time. It would thus be interesting to see whether research collecting more spontaneous data would yield similar results—that is, through the use of technologies such as electroencephalogram (EEG), eye-tracking or pupillometry. This way,

future research could probe more deeply into the subtle changes that occur in the bilingual mind.

6.6 Conclusions

This chapter has outlined the evident links between bilingualism and the emotionally loaded topic of hate speech, and has reported an initial investigation into the understudied impact of bilingualism on hate speech perception and slur appropriation. The analyses of responses to an online questionnaire by 43 Italian residents in the UK demonstrate that despite higher degrees of familiarity with Italian hate words, the same familiarity together with personal use and imageability in Italian tend to decrease with longer LoR in the L2 country, in line with our first hypothesis. At the same time, longer LoR in the L2 country is associated with lower degrees of perceived offensiveness in the L2, which adds to findings in the literature that show lower degrees of emotionality in an L2, by suggesting that the degree of emotionality in the L2 can change according to LoR (and possibly other related variables). With regards to slur appropriation, contrary to our second hypothesis, bilingual experience factors do not seem to predict slur appropriation; on the other hand, participants identifying with sexual minorities are more likely to deem slur appropriation as an appropriate choice, and those identifying with both sexual and ethnic minorities are also more likely to appropriate slurs themselves in situations where they are targeted by hate speech. We have discussed the implications of our study as being central not only to linguistics research, which has long overlooked bilinguals' perception of hate speech specifically, but also as important for other disciplines. We have noted, for instance, what are now well-documented psychological harms of hate speech in response to ethical and philosophical accounts, as well as the need to be aware of the different hate speech perceptions and reactions of bilinguals, which are dependent on a variety of factors and may prove to be fundamental in courts of law and jurisdiction. Finally, we have acknowledged the limitations of our study, and suggested

that future research should not only increase and diversify the number of hate words, scenarios, and participants, but should also collect more linguistic data on both the L1 and L2. Ultimately, the implementation of technologies such as EEG, eye-tracking, or pupillometry may shed further light on the field, by looking at the cognitive and physiological processing of hate speech in late bilingual speakers.

References

Altarriba, Jeanette. 2003. 'Does cariño equal "liking"? A theoretical approach to conceptual nonequivalence between languages'. *International Journal of Bilingualism* 7 (3): 305–322. https://doi.org/10.1177/13670069030070030501.

Bialystok, Ellen, Fergus I. M. Craik, and Gigi Luk. 2012. 'Bilingualism: consequences for mind and brain'. *Trends in Cognitive Sciences* 16 (4): 240–250. https://doi.org/10.1016/j.tics.2012.03.001.

Bianchi, Claudia. 2014. 'Slurs and appropriation: an echoic account'. *Journal of Pragmatics* 66 (May): 35–44. https://doi.org/10.1016/j.pragma.2014.02.009.

Boeckmann, Robert J., and Jeffrey Liew. 2002. 'Hate speech: Asian American students' justice judgments and psychological responses'. *Journal of Social Issues* 58 (2): 363–381. https://doi.org/10.1111/1540-4560.00265.

Brown, Alexander. 2017a. 'What is hate speech? Part 1: The myth of hate'. *Law and Philosophy* 36 (4): 419–468. https://doi.org/10.1007/s10982-017-9297-1.

Brown, Alexander. 2017b. 'What is hate speech? Part 2: Family resemblances'. *Law and Philosophy* 36 (5): 561–613. https://doi.org/10.1007/s10982-017-9300-x.

Calvert, Clay. 1997. 'Hate speech and its harms: a communication theory perspective'. *Journal of Communication* 47 (1): 4–19. https://doi.org/10.1111/j.1460-2466.1997.tb02690.x.

Carey, James W. 1989. *Communication as Culture: Essays on Media and Society*. Boston, MA: Unwin Hyman.

Chiril, Patricia, Endang Wahyu Pamungkas, Farah Benamara, Véronique Moriceau, and Viviana Patti. 2022. 'Emotionally informed hate speech detection: a multi-target perspective'. *Cognitive Computation* 14 (1): 322–352. https://doi.org/10.1007/s12559-021-09862-5.

Costa, Albert, Alice Foucart, Sayuri Hayakawa, Melina Aparici, Jose Apesteguia, Joy Heafner, and Boaz Keysar. 2014. 'Your morals depend on language'. Edited by Mariano Sigman. *PLoS ONE* 9 (4): e94842. https://doi.org/10.1371/journal.pone.0094842.

Dewaele, Jean-Marc. 2010. *Emotions in Multiple Languages*. London: Palgrave Macmillan. https://doi.org/10.1057/9780230289505.

Gallo, Federico, Beatriz Bermudez-Margaretto, Yury Shtyrov, Jubin Abutalebi, Hamutal Kreiner, Tamara Chitaya, Anna Petrova, and Andriy Myachykov. 2021. 'First language attrition: what it is, what it isn't, and what it can be'. *Frontiers in Human Neuroscience* 15 (September): 686388. https://doi.org/10.3389/fnhum.2021.686388.

Gawinkowska, Marta, Michał B. Paradowski, and Michał Bilewicz. 2013. 'Second language as an exemptor from sociocultural norms. Emotion-related language choice revisited'. Edited by Mark Aronoff. *PLoS ONE* 8 (12): e81225. https://doi.org/10.1371/journal.pone.0081225.

Gelber, Katharine, and Luke McNamara. 2016. 'Evidencing the harms of hate speech'. *Social Identities* 22 (3): 324–341. https://doi.org/10.1080/13504630.2015.1128810.

Grosjean, François. 2010. *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.

Grosjean, François. 1989. 'Neurolinguists, beware! The bilingual is not two monolinguals in one person'. *Brain and Language* 36 (1): 3–15. https://doi.org/10.1016/0093-934X(89)90048-5.

Harris, Catherine L. 2004. 'Bilingual speakers in the lab: psychophysiological measures of emotional reactivity'. *Journal of Multilingual and Multicultural Development* 25 (2–3): 223–247. https://doi.org/10.1080/01434630408666530.

Hom, Christopher. 2008. 'The semantics of racial epithets'. *Journal of Philosophy* 105 (8): 416–440. https://doi.org/10.5840/jphil2008105834.

Hornsby, Jennifer. 2003. 'Free speech and hate speech: language and rights'. In *Normatività, fatti, valori*, edited by Rosaria Egidi, Massimo Dell'Utri, and Mario De Caro, 393. Analisi filosofiche 2. Macerata: Quodlibet.

The jamovi project. 2021. *jamovi* (version 1.6) [computer software]. https://www.jamovi.org.

Janschewitz, Kristin. 2008. 'Taboo, emotionally valenced, and emotionally neutral word norms'. *Behavior Research Methods* 40 (4): 1065–1074. https://doi.org/10.3758/BRM.40.4.1065.

Kim, Sun Hee Ok, and Donna Starks. 2008. 'The role of emotions in L1 attrition: the case of Korean-English late bilinguals in New Zealand'. *International Journal of Bilingualism* 12 (4): 303–319. https://doi.org/10.1177/1367006908098573.

Kovács, György, Pedro Alonso, and Rajkumar Saini. 2021. 'Challenges of hate speech detection in social media: data scarcity, and leveraging external resources'. *SN Computer Science* 2 (2): 95. https://doi.org/10.1007/s42979-021-00457-3.

MacAvaney, Sean, Hao-Ren Yao, Eugene Yang, Katina Russell, Nazli Goharian, and Ophir Frieder. 2019. 'Hate speech detection: challenges and

solutions'. Edited by Minlie Huang. *PLOS ONE* 14 (8): e0221152. https://doi.org/10.1371/journal.pone.0221152.

Navarro, Ester, and Andrew R. A. Conway. 2021. 'Adult bilinguals outperform monolinguals in Theory of Mind'. *Quarterly Journal of Experimental Psychology* 74 (11): 1841–1851. https://doi.org/10.1177/17470218211009159.

Pavlenko, Aneta. 2008. 'Emotion and emotion-laden words in the bilingual lexicon'. *Bilingualism: Language and Cognition* 11 (2): 147–164. https://doi.org/10.1017/S1366728908003283.

Pavlenko, Aneta. 2006. 'Bilingual selves'. In *Bilingual Minds: Emotional Experience, Expression, and Representation*, edited by Aneta Pavlenko, 1–33. Bristol: Multilingual Matters. https://doi.org/10.21832/9781853598746-003.

Potts, Christopher. 2007. 'The centrality of expressive indices'. *Theoretical Linguistics* 33 (2): 255–268. https://doi.org/10.1515/TL.2007.019.

Ramírez-Esparza, Nairán, Samuel D. Gosling, Verónica Benet-Martínez, Jeffrey P. Potter, and James W. Pennebaker. 2006. 'Do bilinguals have two personalities? A special case of cultural frame switching'. *Journal of Research in Personality* 40 (2): 99–120. https://doi.org/10.1016/j.jrp.2004.09.001.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing* (version 4.0). Vienna: R Foundation for Statistical Computing. https://cran.r-project.org.

Schmid, Monika S. 2016. 'First language attrition'. *Language Teaching* 49 (2): 186–212. https://doi.org/10.1017/S0261444815000476.

Schmid, Monika S., and Barbara Köpke. 2017. 'The relevance of first language attrition to theories of bilingual development'. *Linguistic Approaches to Bilingualism* 7 (6): 637–667. https://doi.org/10.1075/lab.17058.sch.

Schmid, Monika S., and Gülsen Yılmaz. 2018. 'Predictors of language dominance: an integrated analysis of first language attrition and second language acquisition in late bilinguals'. *Frontiers in Psychology* 9 (August): 1306. https://doi.org/10.3389/fpsyg.2018.01306.

Slagle, Mark. 2009. 'An ethical exploration of free expression and the problem of hate speech'. *Journal of Mass Media Ethics* 24 (4): 238–250. https://doi.org/10.1080/08900520903320894.

Sorace, Antonella. 2020. 'L1 attrition in a wider perspective'. *Second Language Research* 36 (2): 203–206. https://doi.org/10.1177/0267658319895571.

Sulpizio, Simone, Elisa Vassallo, Remo Job, and Jubin Abutalebi. 2020. 'ITABÙ: Dati preliminari di un database delle parole tabù per l'italiano'. *Giornale italiano di psicologia* 2: 599–614. https://doi.org/10.1421/97887.

Wellsby, Michele, Paul D. Siakaluk, Penny M. Pexman, and William J. Owen. 2010. 'Some insults are easier to detect: the embodied insult detection

effect'. *Frontiers in Psychology* 1 (November). https://doi.org/10.3389/fpsyg.2010.00198.

Wilson, Rosemary. 2008. 'Foreign languages: a guide for the inhibited'. *Language Learning Journal* 36 (1): 111–115. https://doi.org/10.1080/09571730801988538.

Wilson, Rosemary. 2013. 'Another language is another soul'. *Language and Intercultural Communication* 13 (3): 298–309. https://doi.org/10.1080/14708477.2013.804534.

Zingaretti, Mattia. 2022. 'First Language Attrition in Late Bilingualism: Lexical, Syntactic and Prosodic Changes in English-Italian Bilinguals'. PhD thesis, University of Edinburgh. *Edinburgh Research Archive*. http://dx.doi.org/10.7488/era/3187.

Zingaretti, Mattia, Vasiliki Chondrogianni, D. Robert Ladd, and Antonella Sorace. Forthcoming. 'L1 Attrition *vis-à-vis* L2 Acquisition: Lexicon, Syntax-Pragmatics Interface and Prosody in L1-English L2-Italian Late Bilinguals'. Special issue, *Languages*.

III

Grammatical and pragmatic strategies

CHAPTER 7

Decoding implicit hate speech

Italian political discourse on social media during the COVID-19 pandemic

Mattia Retta
University of Helsinki

Abstract

In this chapter, I examine the main linguistic and discursive features of anti-Chinese and anti-immigrant discourse in Italian political debate on social media. I combine a top-down approach, focusing on politicians, and a horizontal approach, which analyses the discourses produced by other social media users. The aims of this study are to identify the implicit levels of hate speech found in the corpus and to describe the intertextual and interdiscursive construction of discriminatory and stereotyping language. Implicitness is a key element of online political discourse, since a politician's goal is to induce the audience to perceive the world in the way the politician wants them to. In the study, the pragmatic analysis shows that some kinds of connectives (contrastive, cor-

How to cite this book chapter:
Retta, Mattia. 2024. 'Decoding implicit hate speech: Italian political discourse on social media during the COVID-19 pandemic'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 233–279. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-7

relative, and temporal) and certain adverbial phrases emerge as effective structures to convey such implicit messages. The vilification of out-groups takes place mainly through dehumanising and naturalising metaphors, which are more effectively unveiled by the discourse analysis. This level of analysis also confirms previously identified metaphors and stereotypes used for othering migrants; however, some *topoi* seem to be more commonly attributed to specific categories, such as unreliability and brutality being used almost exclusively in relation to the Chinese.

Keywords: implicit hate speech, implicit meanings, online political debate, in-groups and out-groups, discourse analysis, social media, COVID-19

7.1 Introduction

Hate speech, amplified and disseminated more rapidly by social media, serves a dual purpose: direct aggression against individuals and groups, as well as political propaganda (Bianchi 2021). In political settings hate speech is rarely overtly derogatory, rather it is conveyed implicitly and aims to incite discrimination and hate rather than promote direct violence (Baider 2019, 2023; Ferrini and Paris 2019; Faloppa 2020; Parvaresh 2023). An implicit communication is useful to politicians, allowing them to partially deny their responsibility, while their audience is less inclined to question the content of propaganda (Stanley 2015; Lombardi Vallauri 2019).

Hate speech, a well-established communication strategy in democratic societies (Petrilli 2019), constitutes an ordinary *practice* that is based on a supply–demand logic, where speakers fulfil their audience's expectations by attacking social or ethnic groups (Fumagalli 2019). Following the basic conceptualisation in Critical Discourse Analysis of language as a form of social practice that shapes and is shaped by other social practices and institutions (Fairclough 1989; Fairclough and Wodak 1997; Titscher et al. 2000), we can conclude that the ordinariness of hate speech can normalise hate and demeaning attitudes towards certain

groups and minorities. Hate speech represents a serious threat to democracies as it can silence its targets, restricting their ability to engage in public debate (Pöyhtäri 2015; Bianchi 2021); in addition, the combination of hate speech and the echo chambers of like-minded people encouraged by social media algorithms fuels political antagonism, polarisation, and the strengthening of in-group norms (Pöyhtäri 2015). Studies, such as that by Miller-Idriss (2018), have shown that the legitimation of right-wing populism and its naturalisation of hate speech has caused other political agents to adopt more conservative and nationalistic standpoints.

In this chapter, I will examine the main linguistic and discursive properties of derogatory speech directed against Chinese people and African migrants in the Italian political debate on Twitter and Facebook during the initial period of the COVID-19 pandemic (February to July 2020). The health and economic crisis caused by the pandemic led to a rise in hate speech and hate crimes, both in Italy and in other European countries (Caiani, Carlotti, and Padoan 2021; Dipoppa, Grossman, and Zonszein 2023).

By combining a linguistic-pragmatic analysis with the study of the discursive properties of the corpus, I aim to show how:

(i) Certain linguistic elements, which have no literal meaning of hatred, can convey implicit derogatory messages.
(ii) The meaning of implicit hate speech is co-constructed interdiscursively and generates hatred and discrimination through language that becomes progressively more aggressive in the comment sections.
(iii) Certain discursive and argumentative strategies categorise out-groups and justify the discrimination against them.

I show that the implicit dimension of hate speech is built on two linguistic levels that operate simultaneously. The first level concerns the pragmatic dimension of the implicit meanings related to implicatures, presuppositions, and topicalizations, which are triggered by specific linguistic elements within sentences that ‘function as a Trojan horse for the desired meaning’ (Ferrini and Paris 2019: 25). The second level deals with the discursive and

argumentative strategies through which the interactants—politicians and other users—construe in- and out-groups, (re)producing or strengthening discriminatory communication. For the sake of simplicity, I will call these the pragmatic and the discourse level, respectively.

The pragmatic level of analysis aims to recognise these linguistic elements by making explicit the implicit messages. The subsequent examination of user comments in response to tweets/posts by politicians (for the composition of the corpus, see Section 7.4) provides tangible evidence of the power of implicit language. The discourse analysis is based on Wodak's (2001) discourse-historical approach (DHA) to Critical Discourse Analysis, which understands discourse as a hybrid space and reflects on the interdiscursive construction of identity by detecting specific discursive strategies. In this study, a particular relevance is given to the analysis of argumentation, combining the examination of linguistic forms and the extra-linguistic factors that influence them.

The article is organised as follows. In Section 7.2 I will present the definition of hate speech adopted in this study. Section 7.3 outlines the theoretical basis of the study by defining the pragmatic and discursive dimensions of implicitness investigated here. In Section 7.4 I will briefly present the methodology adopted and will explain the composition of the corpus, also outlining the historical and political background. Section 7.5 presents an analysis of the linguistic elements that function as triggers of implicit meanings, while Section 7.6 explores the discursive and argumentative strategies that categorise out-groups and justify their vilification. Some concluding remarks about the findings and future lines of enquiry will be presented in Section 7.7.

7.2 The definition of hate speech

The very definition of hate speech is problematic, both on a legal and on a linguistic level, because of the difficulty in establishing the limits of free speech (Fish 1994; Määttä 2020) and the excessive importance given to the emotional component (Perry 2001).

Even the academic community has not yet been able to provide an unambiguous definition of this concept (see Brown 2017), and many argue that such search is futile (Anderson and Barnes 2022). Hate speech is best understood as a spectrum or a continuum that involves processes of alienation and social exclusion (Baider 2020; see also Chapter 1, Section 1.3 in this volume). For the identification of derogatory language in the corpus, I adopted the wide-ranging definition by the United Nations, as follows:

> the term hate speech is understood as any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor. This is often rooted in, and generates intolerance and hatred and, in certain contexts, can be demeaning and divisive. (United Nations 2019)

7.3 Implicit meanings

7.3.1 Implicatures, presuppositions, and topicalizations

Political tweets and posts are persuasive texts that induce the audience to perceive the world in the way the author of the text intends, often by imposing biased interpretations and questionable content (Lombardi Vallauri 2019). Persuasion through implicitness is particularly effective because meaning is co-constructed by the original speaker/writer and their audience. The original speaker/writer produces a possible set of implicit meanings, while their audience chooses the necessary inferences (Saul 2002; Sbisà 2021). As derogatory speech is mainly implicit in political settings, implicatures, presuppositions, and topicalizations are central elements of this kind of hate speech.

For the purposes of this study, I employ the traditional Gricean view on the cooperativeness of language (Grice [1975] 1989) and a textual-pragmatic distinction between implicatures and presuppositions (Stalnaker 1973, 1999; Sbisà 2007).

An implicature is the act of implying certain things by saying something else (Davis 2019): it is derived by inference from the speaker's speech and consists in an addition or adjustment to what the utterance says explicitly (Sbisà 2007). Following Grice's distinction ([1975] 1989), implicatures are traditionally divided into two main types: conventional and conversational. The former are generated by the stable *conventional* association with certain lexical items, while the latter can only be communicated and successfully understood in a specific conversational context. Implicating a meaning makes it less questionable and more acceptable for the addressees as it escapes full critical attention (Lombardi Vallauri 2019). This study focuses on the pragmatic features of connectives and adverbials and is thus primarily interested in the use of conventional implicatures within the political debate. For the analysis, I follow Sbisà (2001, 2021), who posits that implicatures are based on the intentions that can be attributed to the writer who produces a certain text in order for the text to be understood as intended: in other words, it is accepted as cooperative and appropriate (see also Chapter 9 in this volume). I adopt this view because political propaganda must be cooperative if it wants to succeed; this approach also represents a more manageable tool for discourse analysis (Sbisà 2021: 178). It must be noted that the goal of cooperation in the comment threads is less clear, but social media content often represents a way to signal belonging and affiliation to certain groups or ideologies (see, e.g., Crosset, Tanner, and Campana 2018).

Presuppositions are, on a very general level, truths in the text that must be taken for granted by the interlocutors in order to consider an utterance as appropriate (Bianchi 2003; Sbisà 2007). The presupposed propositions are established and accepted as the *common ground* among the participants to a conversation, their shared common belief (Kadmon 2001; Stalnaker 2002). I will focus on structural presuppositions (Yule 1996), because of the nature of the linguistic elements under investigation here. Structural presuppositions are linked to the use of specific words and linguistic structures that have been proven to be regular presup-

positional triggers, such as certain connectives or iterative expressions (Yule 1996; Sbisà 2007).

Akin to presuppositions, topicalizations are implicit strategies that allude to the fact that the content is already in the mind of the receiver and that the sender is merely reminding them of it (Lombardi Vallauri 2019: 164). 'Topic' is the name assigned to the linguistic codification of old information, while new information is the 'Focus'. A speaker usually begins from old information and only subsequently adds the Focus, which in pragmatic terms we can describe as the realisation of the informative purpose of a message (Lombardi Vallauri 2019).[1] In the context of hate speech, this informative purpose often aims to amplify pre-existing bias and prejudices rather than transmitting new information.

Focality is an essential requirement to draw attention to some content, and it demands a higher degree of cognitive processing. This means, for example, that an utterance with two Foci requires too much effort and two new pieces of information are still generally presented as Topic + Focus instead of Focus + Focus (Lombardi Vallauri and Masia 2015). Introducing information as topical 'gives the impression' that its content is already active in the linguistic context and therefore available in the short-term memory of the receiver, who processes it as knowledge that was already present in the discourse (Chafe 1994; Lombardi Vallauri 2019). The receivers can thus accept questionable topicalized content with low epistemic vigilance, aiding the formation of biased mental representations (Van Dijk 2006; Lombardi Vallauri 2021).

Differently from presuppositions, topicalizations do not appeal to the general knowledge of the interlocutor, but rather to the small set of things the interlocutor is thinking about at that moment (Lombardi Vallauri 2019: 167).

1 Marked intonations are a known exception to this general rule, but they will not be discussed in this chapter.

7.3.2 Discourse analysis and discourse strategies

The concept of discourse is multifaceted and open to various interpretations, and it has sometimes been criticised for being too vague (see Widdowson 1995). In this chapter, I employ tools of the DHA to Critical Discourse Analysis as developed by Ruth Wodak (2001), which studies linguistic productions in conjunction with extra-linguistic factors, such as the speaker's intention, the historical and political context, and socio-psychological factors (Titscher et al. 2000: 154–163). Wodak (2001: 66) characterises discourse as 'a complex bundle of simultaneous and sequential interrelated linguistic acts, which manifest themselves as thematically interrelated semiotic, oral or written tokens'. The most salient feature of a discourse is having macro-topics, for instance *immigration*, which are open and hybrid (Wodak 2001). During the early months of the COVID-19 pandemic, political discourse on social media saw a surge in derogatory comments against Chinese people and African migrants. Traumatic events, such as a pandemic, can indeed trigger socially defensive choices and tend to increase the production of hate speech (King and Sutton 2013; Caiani, Carlotti, and Padoan 2021; Della Porta 2022). The discourse analysis has identified several macro-topics: economic relationships between China and Italy, health issues, national security, immigration, and integration policies.

The production of tokens, or *texts*, is influenced by and related to the historical context, the genre, and the field of action. Besides the chronological events, the historical context also includes the geographical position, the institutions involved, and the socio-political situation. A genre is the socially accepted manner of using language in connection with a certain social activity (Fairclough 1995: 14). Lastly, a field of action is the portion of societal reality that creates and shapes the 'frame' of discourse and functions as a starting point for the different topics (Titscher et al. 2000; Wodak 2001). It must be noted that discourses cross over different fields of action and genres. Table 7.1 helps contextualise this study through the lens of DHA.

Table 7.1: Analytical schema of the study material according to DHA.

Historical context	Genres	Fields of action
Early stages of the COVID-19 pandemic (Feb–Jul 2020)	Online slogans	Political propaganda (top-down)
Online political debate in Italy	Policy announcements	Formation of public opinion (top-down and horizontal)
Government/ Opposition	Online comments	Self-representation (of politicians and other social media users)

Pragmatic and discursive implicitness clearly overlap, but they must be understood differently. Discursive and argumentative strategies can be linguistically explicit, but often work 'in an unconscious, irrational and emotional way' (Titscher et al. 2000: 156). The discourse-linguistic analysis can unearth these 'hidden' strategies, showing their role within a certain historical and political context.

I argue that the construction of the fundamental opposition between a positive and safe *us* and a negative and dangerous *them* is based on four main strategies: categorisation, perspectivation, intensification/mitigation, and argumentation. Firstly, the analysis focuses on how different groups of people are named and categorised, thus creating several out-groups with characteristics that are perceived as incompatible with the in-group's worldview (Russo and Tempesta 2017: 26). While categorisation is a normal cognitive process that allows us to better grasp reality (Cohen and Lefebvre 2005), its application to social groups often result in prejudice and stereotyping (Mazzara 1997; Russo and Tempesta 2017). The linguistic devices related to this strategy are mainly generalising membership categorisation, dehumanising metaphors, and synecdoche.

The other three strategies aim to justify the discrimination and the clear opposition between in- and out-groups. Perspectivation

Table 7.2: Discursive strategies.

Strategy	Aims	Devices
Categorisation	Construction of in- and out-groups	Generalising categorisation (membership categorisation devices, deictics, etc.)
		Dehumanising metaphors
		Implicit constructions
		Synecdoche
Perspectivation	Speaker's point of view in relation to others	Reporting/description of discriminatory events
		Quotations
Intensification/ Mitigation	Modifying the epistemic status of a proposition	Vagueness
		Hyperbolic expressions
Argumentation	Justification of the argument	Use of argumentative *topoi*[*]
		Implicit constructions

* I am aware that, within argumentation strategies, many include the analysis of fallacies as devices that justify the construction of certain arguments (see, e.g., Wodak 2001; Faloppa 2020). However, for reasons of space, in this chapter I decided to focus only on the analysis of *topoi* as effective means for the circulation of hateful argumentations online.

refers to how speakers represent perspectives, both theirs and others', and position themselves in relation to other subjects (Graumann and Kallmeyer 2002: 4). Common examples of these verbal practices are reported speech or modality (Zifonun 2002). Intensifying and mitigating devices, such as modals, tag questions, hyperboles, and litotes, modify the illocutionary force of an utterance, by either overtly expressing a concept or making it vague and opaque (Wodak 2001). I show how indeterminate expressions such as indefinite pronouns can also intensify derogatory content (see Section 7.6.2). Finally, discrimination is justified by means of *topoi*, explicit or implicit argumentative statements that must be

widely accepted by the participants in order to consider a certain conversation as cooperative and help justify the progress from the arguments to a conclusion (Kienpointner 1992: 194; Wodak 2001).[2] In Table 7.2, I summarise the aims of these strategies and the related linguistic devices.

7.4 Methodology

The qualitative analysis conducted in this study relied on manual investigation due to the inherent nature of political hate speech, which often operates implicitly, with seemingly neutral words carrying hateful messages. The study focuses on material published on Twitter (now X) and Facebook, the two social media platforms that have been an established part of political campaigns and communication since the early 2010s (Enli and Skogerbø 2013; Pietrandrea 2021). Twitter has become the most suitable arena for the strengthening and spread of propagandist content: its structure is open, the content is visible for all, and the militants can easily retweet the leader's comments (Pietrandrea and Battaglia 2022).[3] Facebook, on the other hand, is a more closed system, which allows users to address their followers directly: there is less backlash and negative comments, aiding the creation of militant in-groups.

The corpus of my study is composed of original tweets and Facebook posts by four Italian politicians (Matteo Salvini, Giorgia Meloni, Luca Zaia, and Vincenzo De Luca) published between February and July 2020, and the reaction of other users as seen in their comments. These politicians were chosen according to their representativity, activity on social media, and their main narratives. Salvini and Meloni are the party leaders of the main right-wing parties in Italy, while Zaia and De Luca are two regional administrators who earned great popularity during the pandemic.

2 For criticism of this definition of *topos*, see Žagar (2010).

3 These characteristics have changed since 2023, after Elon Musk's introduction of new rules on data access and visibility on X.

All four of these politicians are very active on social media and their communication style can be described as direct and, at times, aggressive.

The data were collected between August and October 2021. I used the web scraping application Octoparse 8 to automatically collect a total of 6364 tweets written by the selected politicians.[4] Manual annotation identified 128 posts with implicitly derogatory speech directed at either Chinese people or African migrants. Since Italian politicians tend to post the same content on both Twitter and Facebook, I then retrieved the same derogatory posts from Facebook (if possible). Subsequently, I collected 50 comments for each annotated token, using in the case of Facebook the 'most relevant' comment function. This allowed me to have a balanced and manageable corpus of 11,700 comments; of these, 542 were annotated as containing some degree of hate speech. In the case of Luca Zaia, I also included in the corpus an interview extract posted on a newspaper's Twitter page, because it was one of the first comments on the pandemic made by an important Italian politician (see (27) and (39) in sections 7.6.1 and 7.6.2).

The political affiliations of the politicians included in the study, together with some basic information about their popularity on the relevant social media platforms, are given in Table 7.3. In brackets is the number of tokens containing implicit hate speech collected for each politician.

The temporal frame of the study takes into consideration two important events. February 2020 marked the beginning of the pandemic in Italy, with the first local outbreak recorded in Northern Italy around 21 and 22 February. Meanwhile, in July 2020 Italy registered a rise in migrant landings on the coasts of Southern Italy. Cases of COVID-19 were very low, after a tight lockdown, and there was a widespread fear that African migrants might be responsible for a resurgence of the virus. To the best of

4 Octoparse (https://www.octoparse.com/) is a Windows-based web scraping tool that converts unstructured website data into a structured dataset without requiring the use of code.

Table 7.3: The politicians and the political context of the corpus

Name	**Party and political role**	**Followers**
Matteo Salvini (114)	Lega Nord (Northern League) – Senator	1.4 million on Twitter 5+ million on Facebook
Giorgia Meloni (12)	Fratelli d'Italia (Brothers of Italy) – Member of Parliament	1.2 million on Twitter 2.3+ million on Facebook
Luca Zaia (2)	Lega Nord – President of Veneto	130K on Twitter 1.1 million on Facebook
Vincenzo De Luca (0)	Partito Democratico (Democratic Party) – President of Campania	125K on Twitter 1.5+ million on Facebook

my knowledge this speculation was exaggerated, and no relevant outbreaks have been traced back to migrants arriving from North Africa.

Data are transcribed here in their original language and form, followed by an English translation. All translations are mine and are meant for non-native speakers of Italian to understand the examples; the analysis, however, is always based on the original texts. Regarding content published by public figures, the author, the account, the date of publication, and the link to each example is given as it was at the time of the analysis. Conversely, the links and accounts of other users will not be provided, in order to ensure a high degree of anonymity. An alphanumeric code is assigned to each post, according to the social media platform (FB = Facebook, TW = Twitter) and in order of appearance in the chapter.

In the next section, the analysis will focus on linguistic triggers of implicit hate messages such as connectives and adverbials. These seemingly neutral elements can be a vehicle for the expression of propositions that reinforce stereotypes or negatively categorise minorities and other vulnerable groups.

7.5 Pragmatic analysis: linguistic triggers

A fundamental part of any political debate is the establishment and propagation of certain shared ideologies between the elected officials, or electoral candidates, and their actual and future following. Ideologies are 'socially shared mental representations of social groups' (Van Dijk 1998) and the debates around them tend to be organised around polarisations and juxtapositions—a positively characterised in-group versus a negatively characterised out-group, a good *us* versus a bad *them* (Van Dijk 1998; Wodak 2001). In this section, the focus on the pragmatic levels aims to show how the use of certain connectives (Section 7.5.1) or adverbials (Section 7.5.2) aids the linguistic construction of out-groups and the circulation of stereotypes. This inevitably overlaps with some discursive strategies, but it is presented here both for clarity and to emphasise the significance of implicit linguistic strategies in the circulation of hate speech.

7.5.1 Connectives

Connectives are, in their most common meaning, invariable linguistic elements that logically connect textual units (Ferrari 2010, 2014). These elements might belong to different morphological classes such as conjunctions, conjunctive adverbs, or compound prepositions. Following Ferrari (2014), I will not consider as connectives those grammatical elements that establish a linguistic but not a logical connection. This section is concerned with a pragmatic aspect of the semantic properties of certain connectives: their ability to present a piece of information as given (presupposition) or to convey an implicit meaning to be inferred (implicature) by means of their conventional meaning (for an outline of the semantic and pragmatic use of connectives, see, e.g., Van Dijk 1979).

The following tweets represent an appropriate starting point for our analysis, since on a superficial level they do not seem to contain any derogatory content:

(1) Matteo Salvini [@matteosalvinimi] (28 April 2020).

In diretta da #Lampedusa: italiani chiusi in casa, negozi chiusi, **ma** porti sempre aperti!!!

'Live from #Lampedusa: Italians closed up at home, stores [are] closed, but ports [are] always open!!!'

(2) Matteo Salvini [@matteosalvinimi] (25 July 2020).

I modenesi e tutti gli italiani hanno subìto mesi di sofferenze e di lockdown **ma** ora il governo vanifica tutto attirando clandestini positivi.

'The people of Modena and all Italians have endured months of suffering and lockdowns but now the government thwarts it all by attracting positive illegal immigrants.'

The adversative conjunction *ma* (but) has the primary function of linking two contrasting clauses or propositions. Utterances introduced by 'but' have the pragmatic role of objections, persuading the addressee of the message to accept that the objection has met the necessary conditions to be considered appropriate (Sbisà 2007: 132).

I argue that the adversative conjunction is used to convey a similar implication in all of the examples reported above, which can be formulated as follows: (1) → that the government allows migrants into the country and grants them special liberties, while Italians cannot even work; (2) → that the government welcomes migrants that are potentially infected and this will result in Italians suffering again.[5]

The contrast is further strengthened by various dichotomies: closed stores/open ports (1) and suffering Italians/illegal and infected immigrants (2). Ports have specifically become the

5 In the analysis of implicit strategies, I will employ the symbol → for 'implicates'.

embodiment of the arrival and welcoming of migrants on to Italian soil and, in the right-wing rhetoric, they must be closed rather than open.

Another widely used adversative conjunction is *invece di* (instead of), which regularly appears in sentence-initial position. The content introduced by *invece di* is topicalized and presented as already given. As a result, the readers find themselves in a situation where the statement is not perceived as Salvini's, but as something they were themselves already thinking about:

(3) Matteo Salvini [salviniofficial] (12 May 2020).

Invece di assicurare un lavoro ai milioni di italiani disoccupati e ai tanti immigrati regolari e perbene presenti in Italia, il governo pensa a una MAXI-SANATORIA per migliaia di clandestini.

'Rather than guaranteeing a job to the millions of unemployed Italians and to the many legal and respectable immigrants in Italy, the government is considering a MAXI-AMNESTY for thousands of illegal immigrants.'

(4) Matteo Salvini [@matteosalvini] (12 July 2020).

Pazzesco. **Invece di** rinchiudere gli italiani, il governo pensi a chiudere i porti.

'Crazy. Rather than locking up Italians, the government should think about closing the ports.'

In these examples the connective *invece di* is used to criticise the actions of the government, either for what the government should do and does not (3) or for what the government does and should not do (4).

Another connective often employed to convey implicit meanings is the coordinating conjunction *e* (and). This conjunction is not only used to connect two sentences but it also generates parallelisms that implicate a pragmatic relation between the two or

more conjuncts of the parallelism (Lombardi Vallauri 2019; Pietrandrea and Battaglia 2022).

(5) Matteo Salvini [@matteosalvinimi] (14 March 2020).

Gli Italiani non possono uscire di casa, ma accogliamo immigrati **e** mettiamo in pericolo soccorritori e Forze dell'Ordine.

'Italians cannot leave their home, but we welcome immigrants and endanger rescuers and law enforcement.'

(6) Matteo Salvini [@matteosalvinimi] (05 May 2020).

Il governo annuncia una sanatoria per clandestini, e gli sbarchi aumentano (+350%). Stanotte a Lampedusa altri 136 arrivi. Italia campo profughi? NO, grazie.

'The government announces an amnesty for illegal immigrants, and landings increase (+350%). Tonight in Lampedusa another 136 arrivals. Italy refugee camp? NO, thanks.'

(7) Giorgia Meloni [@GiorgiaMeloni] (29 July 2020).

Non consentiremo al Governo di continuare con la sua furia immigrazionista e rendere vani tutti i sacrifici degli italiani.

'We will not allow the government to continue with its immigrationist fury and to render pointless all the sacrifices of Italians.'

These various parallelisms convey implicatures: (5) → that welcoming the migrants into the ports puts law enforcement workers at risk; (6) → that the government's decision to grant an amnesty to those undocumented migrants who already reside and work in Italy will cause a rise in illegal immigration; (7) → that the

government's lax policy on immigration will render pointless the sacrifices that Italians made during the pandemic.

So far, the focus has been on implicit meanings conveyed by single word connectives. Our focus will now shift to more complex connectives that express logical successions. Correlative connectives, such as 'either … or', 'both … and', 'not only … but also', bind sentences closely to one another. In the following examples the connective *non solo … ma anche* (not only … but also) is used to take the information introduced by the first element for granted:

(8) Matteo Salvini [Matteo Salvini official] (23 June 2020).

Non solo il virus che ha infettato il mondo, [**ma**] adesso questo nuovo massacro.

'Not only the virus that infected the world, now this new massacre.'

(9) [FB1] (23 June 2020).

La Cina sta distruggendo **non solo** questi poveri animali indifesi… **Ma** il mondo intero!!!!

'China is destroying not only these poor, defenceless animals… but the whole world!!!'

Salvini's Facebook post in (8) refers to a dog meat eating festival that takes place every June in the city of Yulin in south-eastern China. His short post comes with a link to an article in the Italian newspaper *Corriere della Sera* on the same issue and a picture of a seemingly stray dog in a cage.[6]

The content introduced by *non solo* is topicalized and therefore presented as already established information.

6 The post proved to be extremely popular: as of 25 April 2022, it has 9047 comments and has been shared 6576 times.

Another example of topicalized content is that triggered by the connectives *prima* (first) and *poi* (then) when used in conjunction:

(10) Matteo Salvini [@matteosalvinimi] (26 March 2020).

Prima infettano il mondo, **poi** rischiamo che lo ricomprino.

'First they infect the world, then we risk that they will buy it.'

(11) Matteo Salvini [@matteosalvinimi] (04 May 2020).

Prima lasciano che decine di mafiosi e assassini escano dal carcere, **poi** provano una sanatoria di centinaia di migliaia di clandestini. Abbiamo, tutti insieme, il dovere morale di fermarli.

'First they let tens of mafiosos and murderers leave jail, then they propose an amnesty for thousands of illegal immigrants. We have, all together, the moral obligation to stop them.'

Prima and *poi* are temporal connectives that in (10) and (11) carry a meaning of logical succession. In these examples they not only show a temporal evolution but also establish a consequential relation between the two utterances (for the various functions of *poi* in Italian, see Cruschina and Cognola 2021). Even in this case, the first connective, *prima*, introduces the Topic as the obvious starting point that does not need to be discussed. An explicit assertion such as 'The Chinese infected the world' would have most likely been subject to a larger backlash.

The second connective, *poi*, introduces the logical consequence of the previous proposition, activating implicatures: (10) → that the Chinese intentionally caused the pandemic in order to benefit from it for economic gain; (11) → that illegal immigrants are as

dangerous to Italian society as members of the mafia or murderers.

In the next section, I will conduct a similar pragmatic analysis on another syntactic class, adverbials.

7.5.2 Adverbials

The term 'adverbial' refers to a precise syntactic function that, in Maienborn and Schäfer's (2011: 1391) terms, specifies further the circumstances of the verbal or sentential referent according to limited semantic usage such as time, manner, and place.

Adverbials can be adverbs or adverb phrases ('emotionally', 'properly'), prepositional phrases ('at the restaurant', 'with great care'), or noun/determiner phrases ('the entire year'). This study finds that adverbs and prepositional phrases can be vehicles for implicit language:

(12) [TW1] (14 April 2020).

Non a caso hanno diffuso il virus. Rischio calcolato.

'It was not by chance they spread the virus. Calculated risk.'

Non a caso is an evaluative adverbial phrase with the meaning of intentionally or deliberately. If we analyse this comment (a reply to a tweet by Salvini) as a single sentence, this is how the information structure looks:

[Non a caso hanno diffuso il virus]$_T$. [Rischio calcolato]$_F$

As in previous examples, the Topic is accepted as having been part of the conversation prior to the comment; the Focus introduces new information that → that the Chinese took a calculated risk in infecting the world with the COVID-19 virus to gain some advantage.

Iterative adverbials, such as ‘again’ or ‘once again’, are well attested in the corpus and they are mainly used to add a sense of urgency to the matter at hand. Salvini in particular uses them to both amplify the dangerousness of a perceived migrant emergency and to attack political adversaries:

(13) Matteo Salvini [@matteosalvinimi] (21 June 2020).

Anche la nave Ong tedesca con 211 clandestini a bordo è arrivata in Italia, Sicilia **di nuovo** trasformata nel campo profughi d’Europa.

‘The German NGO ship with 211 illegal immigrants on board also arrived in Italy, Sicily is once again transformed into the refugee camp of Europe.’

(14) Matteo Salvini [@matteosalvinimi] (04 July 2020).

VERGOGNOSO. Sicilia trasformata **di nuovo** in un campo profughi, con delinquenti e violenti che vengono in Italia a metterci i piedi in testa.

‘Shameful. Sicily is transformed once again into a refugee camp, with thugs and violent people coming to Italy to push us around.’

The tweets in (13) and (14) trigger the presupposition that Sicily had already been Europe’s refugee camp in the past, with all the issues and social tensions that this entails. These tweets can be read in two ways. Firstly, they criticise Lega Nord’s Italian and European political adversaries, whose lax migration policies have let the situation evolve to a critical point. Secondly, Salvini’s tweets attack the migrants and their presence in Sicily. This attack is strengthened in (14) by a discursive characterisation of the migrants as violent criminals (see Section 7.6.1).

Lastly, a recurring prepositional phrase emerged from the corpus, *alle spese di* (at the expense of). This phrase is so popular and occurs so often that it could be considered an out-and-out motto of right-wing politicians:

(15) Matteo Salvini [@matteosalvinimi] (17 April 2020).

Ospiterà **a spese degli italiani** gli immigrati per la loro quarantena prima di essere sbarcati.

'It will accommodate, at the expense of Italians, migrants for their quarantine before they are disembarked.'

(16) Giorgia Meloni [@GiorgiaMeloni] (29 July 2020).

Governo pronto a varare una mega nave quarantena per immigrati che sbarcano in Italia: 4,8milioni di € per 92 giorni, per ospitare fino a 400 persone con vitto erogato 'in conformità ai dettami delle diverse religioni'. Un capolavoro politicamente corretto **a spese degli italiani.**

'The government [is] ready to launch a mega-quarantine ship for immigrants landing in Italy: €4.8 million for 92 days, to accommodate up to 400 people with food provided "in accordance with the requirements of the various religions". A masterpiece of political correctness at the expense of Italians.'

The phrase *alle spese di* is by nature ambiguous since it can be interpreted in more than one way. Both (15) and (16) hint at the fact that: a) the Italian government effectively spends its own money (and not, for example, European Union funds); and b) that in doing so, they are taking away resources from Italians. Ambiguity and vagueness (for the difference between these terms, see Machetti 2006) are essential parts of language use. Nevertheless, in textual contexts where the goal is to persuade, being opaque and ambiguous is more in the speaker's interest than being precise and clear (Channell 1994; Bazzanella 2011). I argue that Salvini and Meloni make a deliberate choice of leaving the readers with the 'burden' of interpreting their words. In fact, a more precise assertion might turn out to be either less credible, or less attractive (Lombardi Vallauri 2019: 99). In one of the replies to a post by

Salvini we can find a slightly modified version of the same prepositional phrase, where the ethnic adjective 'Italian' is substituted by the first-person plural possessive *nostre* (our):

(17) [TW2] (02 May 2020).

NOI IN QUARANTENA TAPPATI IN CASA, LE NOSTRE AZIENDE CHIUSE, MENTRE I NEGRI SBARCATI CLANDESTINAMENTE IN ITALIA VANNO IN ALBERGO, SERVITI E RIVERITI. **A NOSTRE SPESE.**

'We are quarantined and holed up in our homes, our businesses closed, while the negroes who illegally landed in Italy go to hotels, served and revered. At our expense.'

The tone of this comment is openly derogatory and insulting. The use of the possessive does not alter in any way the presupposition at play but rather strengthens it, and it is explained by the whole argumentative structure of the post. The author contrasts what is ours (Italians') and what is theirs (the migrants'), highlighting the waste of resources caused by the migrants' arrival and accommodation. This clear distinction between *us* and *them* will be the main topic of the following section, in which I focus on the discursive and argumentative strategies that categorise out-groups and justify their discrimination.

7.6 Discourse analysis: discursive and argumentative strategies

Social media platforms are, as the name itself indicates, interactive. The politicians (or a social media manager on their behalf) write their statements and other users can either share these posts or comment on them. The discourse analysis focuses on how the politicians and other social media users construe in- and out-groups, interactively and interdiscursively, and discuss issues of cultural belonging, national security, and health crisis. Cyberhate is amplified by the internet and social media, whose communica-

tion is global, participatory, immediate, and partially anonymous (see, e.g., Castaño-Pulgarín et al. 2021). This has introduced new aspects to the spread of online hate speech by increasing its replicability, visibility, searchability, persistence, and instantaneousness, making it more difficult to counter cyberhate compared to offline hate speech (Hrdina 2016; Assimakopoulos, Baider, and Millar 2017; Brown 2018; Faloppa 2020).

In Section 7.6.1 the analysis will focus on discourse strategies that create out-groups and categorise them negatively, thus justifying their defamation and vilification. In Section 7.6.2 the focus will shift to other strategies that help to demean minorities through vague or hyperbolic language (intensification/mitigation), or through the reporting of events from the point of view of in-group members (perspectivation).

7.6.1 Categorisation and argumentation: creating and justifying out-groups

Prejudice, hateful remarks, or, in extreme cases, incitement to violence against any kind of minority can arise from the reduction of their identity to one specific trait or situation that is perceived as threatening (Russo and Tempesta 2017: 28–29). When we look at the comments in response to (10), we can observe how the connection between Chinese people and the virus becomes more explicit:

(18) [TW3] (26 March 2020).

I **nostri** morti li dobbiamo a **loro**, esclusivamente a **loro**.

‘Our dead we owe to them, exclusively to them.’

(19) [TW4] (26 March 2020).

Sante parole, ste [sic] **bastardi** cinesi.

‘True words, these Chinese bastards.’

The example in (19) is a very explicit case of hate speech, with the use of a common insult, which aims to demean and attack the Chinese. The first comment (18) is interesting because the opposition between *us* and *them* is made quite explicit through the possessive *nostri* (our) and the pronoun *loro* (them). Personal pronouns tend not to be explicitly expressed in Italian and their overt use indicates an *othering* mechanism that discursively groups the subjects into an *us* and a *them* (Faloppa 2020: 169). This opposition had already emerged in the verbal forms chosen by Salvini in (10), which portray the Chinese as aggressors ('they infected'/'they will buy') and the Italians as victims ('we risk'). This *topos* of economic danger posed by the Chinese is extremely common, both in the politicians' rhetoric and in the comment threads (see (10) in Section 7.5.1):

(20) Matteo Salvini [@matteosalvinimi] (14 April 2020).

Cina, che ha contagiato il mondo, rischia di essere il Paese che cresce di più di tutti Imprenditori italiani chiedono TUTELE.

'China, which infected the world, risks being the country that grows [economically] the most. Italian entrepreneurs request PROTECTIONS.'

(21) [TW5] (14 April 2020).

Chiamiamolo virus **cinese** e non Covid-19.

'Let's call it the Chinese virus and not COVID-19.'

(22) [TW6] (14 April 2020).

#VirusChines *[sic]* chiama @realDonaldTrump e bombardiamoli una volta per tutte!!

'#ChineseVirus Call @realDonaldTrump and let's bomb them once and for all!!'

The argumentative evolution from (20) to (22) shows how a mitigated accusation (20) becomes firstly explicit but still mild (21) and eventually develops into a violent outburst of hate (22). In his tweet, Salvini is taking for granted the active role played by China in the pandemic and presents it as a commonly shared and accepted fact. Meanwhile, (21) and (22) contain much more aggressive language, which relates interdiscursively to the (in)famous characterisation of COVID-19 as the 'Chinese virus' made by Donald Trump, then president of the US. This widespread anti-Asian sentiment translated into a rise in anti-Asian crimes (Center for the Study of Hate and Extremism 2021; Dipoppa, Grossman, and Zonszein 2023) proving once more how these discursive strategies have a real-world impact.

In the example in (8), we saw how Salvini characterised the Chinese as a people that often commit massacres, not only against other human beings but also against innocent animals such as dogs. The comment section on that Facebook post contains a large degree of hate speech, justified by several *topoi*:

(23) [FB2] (23 June 2020).

I cinesi sono persone **subdole, sporche, pericolose, sono da isolare** no come dicono certi politici che dicono 'i nostri amici cimesi [sic]'

'Chinese people are devious, dirty, dangerous, they have to be isolated, not like some politicians say "our Chinese friends".'

(24) [FB3] (23 June 2020).

Questi **se non imparano il minimo dell'igiene**, sarà sempre così.

'These, if they don't learn a minimum of hygiene, it will always be like this.'

(25) [FB4] (23 June 2020).

Non gli *[sic]* ho mai potuto vedere **avari e furbi** che **mangiano tutto quello che cammina che striscia che vola e che nuota** ci manca solo che si mangiano tra di loro un altro po'.

'I have never been able to bear them, they are stingy and cunning, eating everything that walks, crawls, flies, and swims, probably they could even eat each other.'

(26) [FB5] (23 June 2020).

Perché in Italia **nei loro ristoranti cosa credete di mangiare? Nutrie e gatti come minimo.**

'Why, in Italy in their restaurants, what do you think you eat? Nutrias and cats at the very least.'

In the comments listed above, the attacks on Chinese people are justified through a negative characterisation or through stereotypical actions that are used to label the whole population. The negative characterisation emerges through the use of qualitative adjectives, as in (23) and (25), or by explaining what they do wrong, as in (24). On the other hand, we observe in (25) and (26) the reiteration of the negative stereotype of Chinese people who eat everything, regardless of the appropriateness of certain ingredients for human consumption. This same stereotype was also employed by Luca Zaia, president of the Veneto Region:

(27) Luca Zaia [interview reposted by @Corriere on Twitter] (28 February 2020).

La Cina ha pagato un grande conto di questa epidemia che ha avuto **perché li abbiamo visti mangiare tutti topi vivi.**

'This pandemic they had has cost China a lot, because we have all seen them eating live rats.'

Stereotypes are not simply a phenomenon of commonly shared superficial ideas, but they are in constant evolution in various discourses (Ivanou 2017). Widespread stereotypes or argumentative *topoi* can hold a great deal of power over minorities or vulnerable communities. There are numerous examples of stereotypes that were initially perceived as harmless but gradually became the justification for discriminating against certain groups and later led to widespread hate crimes, as in the case of the Rwandan genocide (Ivanou 2017; Määttä 2020).

Even in the comment section analysed here, the *topoi* of threat, barbarity, and incivility develop to the point of utter dehumanisation:

(28) [FB6] (23 June 2020).

Cina ma che nazione è? Che gente È? **Chiamarla gente è un'offesa al genere umano.**

'China, what kind of nation is it? What kind of people? Calling them people is an insult to humankind.'

(29) [FB7] (23 June 2020).

Un orrore da vomitare. **Non sono umani.**

'A horror that makes you throw up. They are not human.'

If the categorisation of Chinese people as an out-group has emerged in close relation to the pandemic, migrants (and particularly African migrants) have been consistently discriminated against and perceived as a threat to Italian society. Many of the *topoi* used in the anti-Chinese discourse are replicated in the categorisation of migrants, albeit with different stereotypical generalisations. As seen in (11), Salvini equated undocumented migrants to members of the mafia and murderers in terms of dangerousness. This *topos* of danger and threat is easily traceable in the replies to that tweet:

(30) [TW7] (04 May 2020).

Non sappiamo se siano stupratori assassini, rapinatori … non sappiamo nulla di loro, però li regolarizziamo.

'We do not know whether they are rapists, murderers, robbers … we know nothing about them, but we regularise them.'

(31) [TW8] (04 May 2020).

Altri tempi non eravamo pieni di **gentaglia** cosi..poi se a voi piace ok va bene

'In other times we weren't full of such riffraff..then if you like it, it's fine.'

(32) [TW9] (04 May 2020).

Sti cazzo di **zulù spacciatori, papponi e nullafacenti** fuori dalle balle!

'These fucking drug-dealing, pimping, do-nothing Zulus should get out the hell out!'

In the comments listed above, migrants are lexically characterised as criminals, in (30) and (32), or generally as less valuable people who are not to be trusted, in (31). The word *gentaglia* is a pejorative of *gente* (people) and reflects a demeaning view of the attacked group (for the role of morphological derivation in hate speech, see Faloppa 2020). The opposition between in- and out-group is made clear in (30) and (31) by the verbal forms on one side ('we know', 'we regularise', 'we weren't'), and the personal or object pronouns on the other ('about them', 'regularise them'). The relatively low infection rate in the summer of 2020 and the perceived risk brought by newcomers strengthened the negative categorisation of the migrants. We witnessed a shift in the stereotypical role of 'virus spreaders' that at the beginning of the

pandemic was attributed to the Chinese. It is the African migrants coming from the sea that might now bring the virus back to Italy and infect the Italians:

(33) Matteo Salvini [@matteosalvinimi]. (22 July 2020).

#Salvini: I pescatori tunisini vengono a pescare nel nostro mare e in cambio fanno arrivare da noi **clandestini col virus…**

'#Salvini: Tunisian fishermen come to fish in our sea and in return they bring us illegal immigrants with the virus…'

(34) Matteo Salvini [@matteosalvinimi] (30 July 2020).

Immigrati mandati a Treviso, ben 129 trovati positivi al Virus! **Se tornerà l'epidemia, sappiamo chi ne sarà colpevole.**

'Immigrants sent to Treviso, as many as 129 found positive for the virus! If the epidemic returns, we know whose fault it will be.'

In the examples listed above, the migrants are described as illegal and infected. The language in (34) is purposely vague. Salvini is apparently criticising the government's decision to welcome multiple migrants in the north-eastern city of Treviso, but the proximity of the two sentences creates the effect of a parallelism in which the reader is prone to consider the migrants as the guilty party in a possible resurgence of the disease.

Another very common *topos* in anti-migrant discourses is that of burden—migrants are described as slackers who avoid being active members of society and will forever be a burden on Italy's finances. The adverbial 'at the expense of Italians' in (16) and (17), which underlines a perceived waste of resources on the immigrants, has already been discussed. In (32) the user employs the adjective *nullafacenti* (do-nothings) in his very derogatory comment about African migrants. This discriminatory depiction

of migrants as a threat and a burden is well attested by previous research (see, among others, Assimakopoulos, Baider, and Millar 2017; Strani and Szczepaniak-Kozak 2018; Määttä, Suomalainen, and Tuomarla 2021; Bonhomme and Alfaro 2022).

In the following examples we see a very common metaphor that reinforces the prejudice of migrants as economic burden, widely employed by Matteo Salvini and many other users. It consists in labelling migrants as 'tourists':

(35) Matteo Salvini [@matteosalvinimi] (23 July 2020).

Ennesimi clandestini in arrivo a spese degli Italiani, mentre **i turisti veri** cancellano le vacanze su questa splendida isola.

'Yet more illegal immigrants arrive at the expense of Italians, while real tourists cancel their holidays on this beautiful island.'

(36) Matteo Salvini [@matteosalvinimi] (29 July 2020).

Solo questa notte e solo a Lampedusa altri 314 **'turisti per sempre'** in fuga dalla famosa 'guerra di Tunisia'…!

'Tonight alone and in Lampedusa alone another 314 "eternal tourists" fleeing the famous "Tunisian war"…!'

(37) [TW10] (29 July 2020).

2020, gli italiani non hanno soldi per fare 1 vacanza, ma **i migranti clandestini li mandiamo in Crociera** nel Mediterraneo??

'2020, Italians have no money to take 1 holiday, but we send the illegal migrants on a Mediterranean cruise?'

In (35) migrants are implicitly described as fake tourists: unlike real tourists, Salvini sees them as a liability and not as a resource. Similarly, 'eternal tourists' in (36), a widespread right-wing motto,

ironically underlines a perceived unwillingness to integrate into Italian society (Retta 2023). It is interesting to note how Salvini uses inverted commas in *'guerra di Tunisia'* to hint at the lack of prerequisites for these migrants to obtain political asylum. The commenter in (37) goes one step further and, using a similar metaphor, attacks both the migrants and the government, which is accused of gifting the migrants a holiday while Italians can no longer afford one. In anti-migrant discourses, hate speech is often directed not only at out-groups but also at those members of the in-group who support more open migration policies, for instance, or who simply reject aggressive approaches:

(38) [TW11] (29 July 2020).

Facciamo qualcosa per **fermare questi coglioni che svendono il paese a negri, islamici e zingari. La feccia dell'umanità**.

'Let's do something to stop these assholes selling out the country to niggers, Islamists, and gypsies. The scum of humanity.'

Here, hate speech is addressed not so much to its victims, but rather to political adversaries and to like-minded people as a means of engagement and hate group forming. Online hate speech manages to connect people who would not have otherwise been in contact with each other, thereby reinforcing the creation of in-groups and cementing intra-group community (Brown 2018; Baider and Constantinou 2020; Caiani, Carlotti, and Padoan 2021).

In this section the analysis focused on how the out-groups are created and the arguments employed to justify their discrimination. In the following section the focus will move to two other strategies that help to demean minorities: perspectivation and intensification/mitigation.

7.6.2 Perspectivation and intensification/mitigation

Speakers/writers use perspectivation strategies to position themselves in relation to others and to give their own account or explanation of other people's behaviour, opinions, and events (Faloppa 2020: 171). Perspectivation creates a distance between the in-group of the speaker and the attacked out-group, setting boundaries and contrasting elements between the two groups.

The following example is taken from an interview with Luca Zaia that was reposted on Twitter by the *Corriere della Sera* at the very beginning of the COVID-19 crisis. A small excerpt of the interview was previously analysed in (27).

(39) Luca Zaia [interview reposted by @Corriere on Twitter] (28 February 2020).

La mentalità che ha il nostro popolo a livello di igiene è quella di farsi la doccia, di lavarsi spesso le mani. L'alimentazione, il frigorifero, le scadenze degli alimenti **sono un fatto culturale**. La Cina ha pagato un grande conto di questa epidemia che ha avuto perché **li abbiamo visti tutti mangiare i topi vivi**.

'The mentality that our people have in terms of hygiene is to shower, to wash their hands often. Eating, the refrigerator, food expiry dates are a cultural fact. This pandemic they had has cost China a lot, because we have all seen them eating live rats.'

On 28 February 2020 the pandemic had not yet gained a foothold in Italy, and Zaia hence seems to consider it as an issue that concerns only the Chinese. He positions himself as someone who is commenting from the standpoint of an Italian, judging and blaming the pandemic on Chinese cultural habits. Expressions such as *il nostro popolo* (our people) and the use of the first-person plural in the verbal form *li abbiamo visti* (we saw them) underline this standpoint and reinforce the distance between the positive *us* and the negative *them*. Italians are described as hygienic people who

regularly wash themselves: storing and conserving food properly is part of their culture, *un fatto culturale* (a cultural fact). The negative perspective towards the Chinese is reinforced by his final statement: speaking once more on the behalf of all Italians, he claims that *we* all (Italians) have seen *them* (Chinese people) eating live rats. There is no truth to this generalising statement in real life, and it plays on a crude stereotype. This claim can also be seen as a dehumanising image, since the consumption of rats is normally associated with wild beasts. Images of bestiality or incivility are employed in other examples of perspectivation strategies in some of the replies to (8) (see Section 7.5.1):

(40) [FB8] (23 June 2020).

Quando passeggio con la mia cucciolona **i cinesi la guardano con certi occhi…**e' cosi bella che se la mangiano con gli occhi! **Loro sono cosi**, mandano al macello i cani e anche ogni genere che striscia!

'When I walk with my puppy, the Chinese look at her in a certain way… she is so beautiful that they eat her with their eyes! They are like that, they send dogs to the slaughter and also anything that crawls!'

(41) [FB9] (23 June 2020).

Ho visto video di cani scuoiati vivi solo per un collo di pelliccia. Ho visto immagini che non dimenticherò mai. Questo è l'Oriente.

'I have seen videos of dogs skinned alive just for a fur collar. I have seen images I will never forget. This is the Orient.'

The comment in (40) reiterates the stereotype that the Chinese consume all sorts of meat. The author reports a personal anecdote about Chinese people coveting their dog. 'Eating somebody with one's own eyes' is a common figure of speech in Italian to express

lust and desire. Here the figure of speech works as a double entendre: the Chinese people look at the dog not only because it is beautiful but also because they want to eat it. Using an anecdote creates engagement in the readers who already believe in the incivility of the Chinese, so much so that the commenter does not need to justify their position. They simply claim that Chinese people 'are like that'. The comment in (41) is also linked to a personal experience highlighted by the repeated use of the verbal form *ho visto* (I saw). This commenter first vaguely reports something seen in the past as a way of proving the brutality and incivility of Chinese people and then reduces them to this single prejudicial trait by using the demonstrative *questo* (this). The overlap between China and the whole concept of the Orient should also be noted here.

Intensification and mitigation are strategies employed to exaggerate certain derogatory concepts or to mitigate the seriousness of particularly egregious expressions and accusations. Intensification strategies are often linked to the use of certain adverbs or adjectives:

(42) [FB10] (23 June 2020).

> **Ogni virus** viene da quel paese perché **mangiano di tutto senza nessuna regola** come è accaduto con la malattia dell'aviara.
>
> 'Every virus comes from that country because they eat everything without any rules, as happened with the bird flu.'

(43) [FB11] (23 June 2020).

> I cinesi oltre che crudeli fanno **solo** danni.
>
> 'The Chinese, besides being cruel, only do harm.'

The linguistic markers of intensification in the examples above are the indefinite adjectives and adverbs in (42) and the degree adverb *solo* (only) in (43), which emphasise once more a perceived cruelty and incivility that is attributed to the Chinese (for the role

played by indeterminacy in language and its pragmatic relevance, see, e.g., Cutting 2007; Bazzanella 2011; Lombardi Vallauri 2019).[7]

The use of certain metaphors can also be seen as an intensification strategy. One of the most common metaphors employed against both Chinese people and migrants arriving from Africa is that of invasion. This is linked to the so-called white replacement theory, a widespread conspiracy theory in white ethnonationalist networks, which claims that global elites are trying to replace ethnically white populations with people coming from the Global South (Cosentino 2020). Here are a few examples:

(44) Matteo Salvini [@matteosalvinimi] (29 July 2020).

Lampedusa. **L'invasione organizzata continua**, giorno e notte. Conte-Lamorgese, sveglia! Sveglia!

'Lampedusa. The organised invasion continues, day and night. Conte-Lamorgese, wake up! Wake up!'

(45) [TW12] (12 May 2020).

Questa sostituzione etnica spinta dal Vaticano e dai sinistri verranno sul groppone degli italiani vogliono essere mantenuti bighellonando e aspettando il pranzo nel tempo libero poi si spaccia droga ai nostri giovani!!!

'This ethnic substitution pushed by the Vatican and the left will come on the backs of Italians[;] they [the immigrants] want to be supported while they're loitering and waiting for lunch in their spare time then dealing drugs to our youth!!!'

7 This argumentation of incivility could also be explained through the concept of explicature—that is, what is explicitly communicated. The author of (42), for example, does not want to convey the idea that the Chinese literally eat everything, but rather directly communicates that they eat things that should not be eaten (for a reflection on the implicature/explicature distinction, see Carston and Hall 2012).

The metaphorical use of language is not, naturally, a relevant indicator in the context of hate speech detection. Certain metaphors (and other figures of speech) are nonetheless cognitive instruments that can activate a transfer that disconnects the production of discourse from reality, making us perceive an unsettling and alternative reality that generates distrust, hostility, and hatred (Faloppa 2020: 174–175). White replacement theory has been one of the motivations behind several racist terrorist attacks since the mid-2010s, such as the mass murder of Black churchgoers in Charleston (South Carolina) in 2015 or the massacre of Muslims in Christchurch (New Zealand) in 2019.

There is a third intensifying strategy that is extremely common in right-wing rhetoric: the improper use of quantifiers, numbers, and statistics. This 'rhetorical number game' (Van Dijk and Wodak 2000: 75) aims to catch readers' attention and persuade them that the author is reliable. In the following examples we see how large numbers are used either to discredit the government that is causing the suffering of countless Italians or to underline the threat of an immigrant invasion:

(6') Matteo Salvini [@matteosalvinimi] (05 May 2020).

Il governo annuncia una sanatoria per clandestini, e gli sbarchi aumentano (+**350%**). Stanotte a Lampedusa **altri 136** arrivi. Italia campo profughi? NO, grazie.

'The government announces an amnesty for illegal immigrants, and landings increase (+350%). Tonight in Lampedusa another 136 arrivals. Italy refugee camp? NO, thanks.'

(46) Giorgia Meloni [@giorgiameloni] (12 May 2020).

Ma vi sembra normale che mentre **milioni di italiani ancora attendono i soldi promessi** per arrivare alla fine del mese, il governo abbia come priorità **regolarizzare centinaia di migliaia di clandestini**?

> ‘Does it seem normal to you that while millions of Italians are still waiting for the promised money to make ends meet, the government’s priority is to regularise hundreds of thousands of illegal immigrants?’

All the numbers presented in the tweets above are difficult to prove but also to rebut. No source or reference is presented for these statistics: for example, in (6) the 350 per cent increase in migrant arrivals could potentially be credible, but we have no means of knowing where it comes from or to what it actually refers. The only precise number presented here is the number of migrants, 136, who landed in Lampedusa on 5 May 2020. The number is preceded by *altri* (another), underlining that this quantity must be considered as the ‘tip of the iceberg’—in other words, that it is still remarkable even if it seems like a small number. When introducing precise numbers, Salvini often uses a similar expression, the adverb *ben* (as many as), in order to express urgency and dangerousness:

(34’) Matteo Salvini [@matteosalvinimi] (30-07-2020).

> Immigrati mandati a Treviso, **ben 129** trovati positivi al Virus! Se tornerà l’epidemia, sappiamo chi ne sarà colpevole.
>
> ‘Immigrants sent to Treviso, as many as 129 found positive for the virus! If the epidemic returns, we know whose fault it will be.’

Mitigating strategies appear to be rarer in the corpus and mostly relate to the rhetorical device of preterition or apophasis, which consists in the author bringing up a subject by professing to omit it. This strategy is very common in derogatory comments in sentences such as ‘I am not racist, but…’ or ‘I am not homophobic, but…’. The following comments are both related to the Yulin dog meat eating festival:

(47) [FB12] (23 June 2020).

Non voglio essere cattivo, ma i cinesi sono il popolo che non sarebbero dovuti nascere su questa terra.

'I don't want to be mean, but the Chinese are people who should not have been born on this earth.'

(48) [FB13] (23 June 2020).

Sono la rovina del mondo… spero vivamente che tutto quello che faranno a quei poveri cani venga fatto anche a loro… **vorrei dire molto di peggio ma** non mi voglio abbassare a quel "popolo" se si può chiamare così.

'They are the ruin of the world… I sincerely hope that whatever they do to those poor dogs will be done to them too… I would like to say a lot worse but I don't want to lower myself to those "people" if you can call them that.'

Preterition is used to mitigate the very strong accusations presented by the authors, who deny their will to insult and say 'worse things', but nevertheless use derogatory language against the Chinese.

After the empirical description of discourse strategies in sections 7.6.1 and 7.6.2, we can move on to the concluding remarks, in which I summarise the main results of the study and its implications for further studies in linguistics and other disciplines.

7.7 Conclusions

The structural characteristics of social media shape and guide the way users produce content and mediate their interaction (for the concept of *affordances*, see Biri 2023 and references therein). As moderation tools become more sophisticated, implicitness turns into a key feature of online political discourse, and in certain contexts it can lead to forms of hate speech. In this study, the analysis was performed on two interconnected levels: pragmatic

and discursive. The pragmatic analysis focused on the activation of implicit meanings by specific connectives, such as *ma* (but), *invece di* (instead of), *e* (and), *non solo … ma anche* (not only … but also), *prima … poi* (first … then). These words that do not possess a literal meaning of hatred provide politicians with the means to incite discriminatory discourses among their followers while adhering to the rules of conduct on social media. Similarly, adverbials such as *non a caso* (not by chance) or *alle spese di* (at the expense of) can also convey implicit messages: the latter, in its longer form of *alle spese degli italiani* (at the expense of Italians) has become a particularly widespread right-wing motto.

A search for implicit messages is an important instrument for further inquiries into the regulation and moderation of online hate speech. Internet platforms are better equipped to counter cyberhate than governmental agencies and institutions (Brown 2018: 310). While automated moderation will improve, algorithms struggle to recognise irony, misspelt words, neologisms, or implicit constructions. A linguistic approach to the digital education of moderators and users can provide valuable tools to decipher and prevent hate speech, reducing the risk of becoming victims or perpetrators of propaganda and discrimination.

The discourse analysis recognised the main discursive strategies that create out-groups and justify their vilification. The two out-groups that emerged from the corpus—Chinese people and African migrants—are predominantly discriminated against by dehumanising metaphors and prevalent *topoi* in racist discourses, such as those of dirtiness, disease, burden, threat, and incivility. Although the arguments were similar, the stereotypical imagery used was different. On the one hand, the Chinese were perceived as the cause of the pandemic and as a threat to the physical and economic well-being of the world. This perceived incivility is mostly underlined by comments on their eating habits, which are seen as unfit for modern civilisations. African migrants, on the other hand, are mainly categorised as physical threats or as a burden. The discrimination against them is also justified by perceived laziness and unwillingness to integrate, exemplified by the

common metaphor of 'eternal tourists'. Intensifying or mitigating strategies are employed to exaggerate the threatening attitudes of the out-groups in a derogatory way, or to mitigate the seriousness of particularly egregious expressions and accusations. Discourse analysis shows how hate speech goes beyond personal offence, and derogatory discursive and argumentative strategies naturalise and normalise discriminating attitudes.

The study and its methodology are not without limitations. First, implicit forms of hate speech have by nature a lower degree of intensity than explicitly conveyed hate speech. This means that, within different definitions of this concept, some of the comments presented in the study might be considered merely opinions—albeit distasteful and hostile—rather than expressions of hatred. Secondly, the analysis of the social media contents of only four politicians makes it difficult to generalise the observations in terms of how widespread implicit hate speech is within political discourses. Nonetheless, this choice is justified in terms of feasibility of the study, representativity of the chosen politicians, and relevance of the results. The main contribution of this chapter is specifically to show how the concept of implicitness can be applied to the detection and countering of hate speech, complementing previous works in different contexts worldwide (Baider 2019, 2023; Baider and Constantinou 2020; Parvaresh 2023).

Hate speech relates primarily to language use, which always constructs reality. An adequate linguistic and discursive definition is an essential step towards a better definition of what we can and cannot consider hate speech. This research represents an additional step in finding educational and normative tools to fight hate speech and is part of a growing multidisciplinary approach to the effects of the COVID-19 pandemic on political propaganda, hate speech, and online abuse.

References

Anderson, Luvell, and Michael Barnes. 2022. 'Hate speech'. In *The Stanford Encyclopedia of Philosophy* (Spring 2022 edition), edited by Edward N. Zalta. Accessed 9 September 2023. https://plato.stanford.edu/archives/spr2022/entries/hate-speech.

Assimakopoulos, Stavros, Fabienne H. Baider, and Sharon Millar, eds. 2017. *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*. Berlin: Springer. https://doi.org/10.1007/978-3-319-72604-5.

Baider, Fabienne. 2023. 'Accountability issues, online covert hate speech, and the efficacy of counter-speech'. *Politics and Governance* 11 (2): 249–260. https://doi.org/10.17645/pag.v11i2.6465.

Baider, Fabienne. 2020. 'Pragmatics lost? Overview, synthesis and proposition in defining online hate speech'. *Pragmatics and Society* 11 (2): 196–218. https://doi.org/10.1075/ps.20004.bai.

Baider, Fabienne. 2019. 'Le Discours de haine dissimulée: le mépris pour humilier'. *Déviance et Société* 43 (3): 359–387. https://doi.org/10.3917/ds.433.0359.

Baider, Fabienne, and Maria Constantinou. 2020. 'Covert hate speech: A contrastive study of Greek and Greek Cypriot online discussions with an emphasis on irony'. *Journal of Language Aggression and Conflict* 8 (2): 262–287. https://doi.org/10.1075/jlac.00040.bai.

Bazzanella, Carla. 2011. 'Indeterminacy in dialogue'. *Language and Dialogue* 1 (1): 21–43. https://doi.org/10.1075/ld.1.1.04baz.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.

Bianchi, Claudia. 2003. *Pragmatica del linguaggio*. Bari and Rome: Laterza.

Biri, Ylva. 2023. 'Stancetaking in Interest-Based Online Communities: A Corpus Pragmatic Comparative Analysis.' PhD thesis, University of Helsinki. http://urn.fi/URN:ISBN:978-951-51-9453-4.

Bonhomme, Macarena, and Amaranta Alfaro. 2022. ' "The filthy people": racism in digital spaces during COVID-19 in the context of South–South migration'. *International Journal of Cultural Studies* 25 (3–4): 404–427. https://doi.org/10.1177/13678779221092462.

Brown, Alexander. 2018. 'What is so special about online (as compared to offline) hate speech?' *Ethnicities* 18 (3): 297–326. https://doi.org/10.1177/1468796817709846.

Brown, Alexander. 2017. 'What is hate speech? Part 1: The myth of hate'. *Law and Philosophy* 36 (4): 419–468. https://doi.org/10.1007/s10982-017-9297-1.

Caiani, Manuela, Benedetta Carlotti, and Enrico Padoan. 2021. 'Online hate speech and the radical right in times of pandemic: the Italian and English cases'. *Javnost—The Public* 28 (2): 202–218. https://doi.org/10.1080/13183222.2021.1922191.

Carston, Robyn, and Alison Hall. 2012. 'Implicature and explicature'. In *Cognitive Pragmatics*, Handbooks in Pragmatics 4, edited by Hans-Jörg Schmid, 7–84. Berlin: Mouton de Gruyter.

Castaño-Pulgarín, Sergio Andrés, Natalia Suárez-Betancur, Luz Magnolia Tilano Vega, and Harvey Mauricio Herrera López. 2021. 'Internet, social media and online hate speech. Systematic review'. *Aggression and Violent Behavior* 58 (2021): 101608. https://doi.org/10.1016/j.avb.2021.101608.

Center for the Study of Hate and Extremism. 2021. 'Report to the nation: anti-Asian prejudice and hate crime'. Accessed 9 September 2023. https://www.csusb.edu/hate-and-extremism-center.

Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago, IL: University of Chicago Press.

Channell, Joanna. 1994. *Vague Language*. Oxford: Oxford University Press.

Cohen, Henri, and Claire Lefebvre, eds. 2005. *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier.

Cosentino, Gabriele. 2020. 'From Pizzagate to the Great Replacement: the globalization of conspiracy theories'. In *Social Media and the Post-truth World Order*, 59–86. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-43005-4_3.

Crossett, Valentine, Samuel Tanner, and Aurélie Campana. 2018. 'Researching far right groups on Twitter: methodological challenges 2.0'. *New Media in Society* 21 (4): 939–961. https://doi.org/10.1177/1461444818817306.

Cruschina, Silvio, and Federica Cognola. 2021. 'From connective adverb to modal particle: a generative analysis of *poi*'. In '(Nuove) Prospettive di analisi dei connettivi', edited by Angela Ferrari and Filippo Pecorari, special issue, *Studi italiani di linguistica teorica e applicate* 50 (1): 52–68.

Cutting, Joan, ed. 2007. *Vague Language Explored*. London: Palgrave Macmillan.

Davis, Wayne. 2019. 'Implicature'. In Zalta, *The Stanford Encyclopedia of Philosophy*. Accessed 9 September 2023. https://plato.stanford.edu/entries/implicature/.

Della Porta, Donatella. 2022. 'Social movements in the emergence of a global pandemic'. In *Social Movements and Politics During COVID-19*, edited by Bringel Breno and Geoffrey Pleyers, 131–133. Bristol: Bristol University Press.

Dipoppa, Gemma, Guy Grossman, and Stephanie Zonszein. 2023. 'Locked down, lashing out: COVID-19 effects on Asian hate crimes in Italy'. *Journal of Politics* 85 (2): 389–404. https://doi.org/10.1086/722346.

Enli, Gunn Sara, and Eli Skogerbø. 2013. 'Personalized campaigns in party-centered politics: Twitter and Facebook as arenas for political communication'. *Communication and Society* 16 (5): 757–774. https://doi.org/10.1080/1369118X.2013.782330.

Fairclough, Norman. 1995. *Critical Discourse Analysis: The Critical Study of Language*. New York: Longman Publishing.

Fairclough, Norman. 1989. *Language and Power*. London: Longman.

Fairclough, Norman, and Ruth Wodak. 1997. 'Critical Discourse Analysis: an overview'. In *Discourse and Interaction*, edited by Teun A. Van Dijk, 67–97. London: Sage.

Faloppa, Federico. 2020. *#ODIO: manuale di resistenza alla violenza delle parole*. Turin: UTET.

Ferrari, Angela. 2014. *Linguistica del testo: principi, fenomeni, strutture*. Rome: Carocci editore.

Ferrari, Angela. 2010. 'Connettivi'. In *Enciclopedia dell'Italiano*. Rome: Istituto della Enciclopedia Italiana. Accessed 25 May 2022. https://www.treccani.it/enciclopedia/connettivi_%28Enciclopedia-dell%27Italiano%29/.

Ferrini, Caterina, and Orlando Paris. 2019. *I discorsi dell'odio*. Rome: Carocci editore.

Fish, Stanley. 1994. *There's No Such Thing as Free Speech and It's a Good Thing, Too*. New York: Oxford University Press.

Fumagalli, Corrado. 2019. 'Discorsi d'odio come pratiche ordinarie'. *Biblioteca Della Libertà* 224: 55–75. https://doi.org/10.23827/BDL_2019_1_3.

Graumann, Carl F., and Werner Kallmeyer, eds. 2002. *Perspective and Perspectivation in Discourse*. Amsterdam and Philadelphia, PA: John Benjamins.

Grice, H. Paul. 1989. 'Logic and conversation'. In *Studies in the Way of Words*, 22–40. Cambridge, MA: Harvard University Press. Originally published in 1975 in *Syntax and Semantics*, vol. 3: *Speech Acts*, edited by Peter Cole and Jerry L. Morgan, 41–58. New York: Academic Press.

Hrdina, Matouš. 2016. 'Identity, activism and hatred: hate speech against migrants on Facebook in the Czech Republic in 2015'. *Naše společnost* 14 (1): 38–47.

Ivanou, Uladzislau. 2017. 'Stereotyping vulnerable groups'. In Assimakopoulos, Baider, and Millar, *Online Hate Speech in the European Union*, 29–32. https://doi.org/10.1007/978-3-319-72604-5_3.

Kadmon, Nirit. 2001. *Formal Pragmatics: Semantics, Pragmatics, Preposition, and Focus*. Oxford: Blackwell.

Kienpointer, Manfred. 1992. *Alltagslogik: Struktur und Funktion von Argumentationsmustern*. Stuttgart: Frommann-Holzboog.

King, Ryan D., and Gretchen M. Sutton. 2013. 'High times for hate crimes: explaining the temporal clustering of hate-motivated offending'. *Criminology* 51 (4): 871–894. https://doi.org/10.1111/1745-9125.12022.

Lombardi Vallauri, Edoardo. 2021. 'Manipulative shallow processing induced by presuppositions and topics: theoretical perspectives and experimental evidence'. *Frontiers in Communication* 6 (7 April 2021). https://doi.org/10.3389/fcomm.2021.610807.

Lombardi Vallauri, Edoardo. 2019. *La lingua disonesta: contenuti impliciti e strategie di persuasione*. Bologna: Il Mulino.

Lombardi Vallauri, Edoardo, and Viviana Masia. 2015. 'Cognitive constraints on the emergence of Topic–Focus structure in human language'. In *Immagine e pensiero: bilanci nelle scienze cognitive attuali*, edited by Alessandra Chiera and Vittorio Ganfi, 180–204. Rome and Messina: Corisco Edizioni.

Määttä, Simo. 2020. 'Syrjivän kielenkäytön rajoilla'. *Niin&Näin* 3: 71–77.

Määttä, Simo, Karita Suomalainen, and Ulla Tuomarla. 2021. 'Everyday discourse as a space of citizenship: the linguistic construction of in-groups and out-groups in online discussion boards'. *Citizenship Studies* 25 (6): 773–790. https://doi.org/10.1080/13621025.2021.1968715.

Machetti, Sabrina. 2006. *Uscire dal vago: analisi linguistica della vaghezza del linguaggio*. Bari and Rome: Laterza.

Maienborn, Claudia, and Martin Schäfer. 2011. 'Adverbs and adverbials'. In *Semantics: An International Handbook of Natural Language Meaning*, edited by Claudia Maienborn, Klaus von Heusinger, and Paul Portner, 1390–1420. Berlin and Boston, MA: De Gruyter.

Mazzara, Bruno M. 1997. *Stereotipi e pregiudizi*. Bologna: Il Mulino.

Miller-Idriss, Cynthia. 2018. *The Extreme Gone Mainstream*. Princeton, NJ: Princeton University Press.

Parvaresh, Vahid. 2023. 'Covertly communicated hate speech: a corpus-assisted pragmatic study'. *Journal of Pragmatics* 205: 63–77. https://doi.org/10.1016/j.pragma.2022.12.009.

Perry, Barbara. 2001. *In the Name of Hate: Understanding Hate Crime*. London and New York: Routledge.

Petrilli, Raffaella. 2019. 'Il meccanismo dell'odio nel discorso politico'. In *Hate speech: l'odio nel discorso pubblico*, edited by Raffaella Petrilli, 41–58. Rome: Round Robin.

Pietrandrea, Paola. 2021. *Comunicazione, dibattito pubblico, social media: come orientarsi con la linguistica*. Rome: Carocci editore.

Pietrandrea, Paola, and Elena Battaglia. 2022. ' "Migrants and the EU": the diachronic construction of ad hoc categories in French far-right discourse'. *Journal of Pragmatics* 192: 139–157. https://doi.org/10.1016/j.pragma.2022.02.012.

Pöyhtäri, Reetta. 2015. 'Vihapuhe haasteena uutismedialle ja journalismille'. In *Vihapuhe Suomessa*, edited by Riku Neuvonen, 241–266. Helsinki: Edita.

Retta, Mattia. 2023. 'A pragmatic and discourse analysis of hate words on social media'. *Internet Pragmatics* 6 (2): 197–218. https://doi.org/10.1075/ip.00096.ret.

Russo, Ernesto, and Pablo Bernardino Tempesta. 2017. 'Categorisation and defence strategies'. In Assimakopoulos, Baider, and Millar, *Online Hate*

Speech in the European Union, 25–29. https://doi.org/10.1007/978-3-319-72604-5_3.

Saul, Jennifer M. 2002. 'Speaker meaning, what is said and what is implicated'. *Noûs* 36 (2): 228–248. https://doi.org/10.1111/1468-0068.00369.

Sbisà, Marina. 2021. 'Presupposition and implicature: varieties of implicit meaning in explicitation practices'. *Journal of Pragmatics* 182 (September): 176–188. https://doi.org/10.1016/j.pragma.2021.05.027.

Sbisà, Marina. 2007. *Detto non detto: le forme della comunicazione implicita*. Bari and Rome: Laterza.

Sbisà, Marina. 2001. 'Intentions from the other side'. In *Paul Grice's Heritage*, edited by Giovanna Cosenza, 185–206. Turnhout: Brepols.

Stalnaker, Robert. 2002. 'Common ground'. *Linguistics and Philosophy* 25: 701–721. https://doi.org/10.1023/A:1020867916902.

Stalnaker, Robert. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Stalnaker, Robert. 1973. 'Presuppositions'. *Journal of Philosophical Logic* 2 (4): 447–457. https://doi.org/10.1007/BF00262951.

Stanley, Jason. 2015. *How Propaganda Works*. Princeton, NJ: Princeton University Press.

Strani, Katerina, and Anna Szczepaniak-Kozak. 2018. 'Strategies of othering through discursive practices: examples from the UK and Poland'. *Lodz Papers in Pragmatics* 14 (1): 163–179. https://doi.org/10.1515/lpp-2018-0008.

Titscher, Stefan, Michael Meyer, Ruth Wodak, and Eva Vetter. 2000. *Methods of Text and Discourse Analysis*, translated by Bryan Jenner. London: Sage.

United Nations. 2019. 'United Nations strategy and plan of action on hate speech'. Accessed 12 August 2024. https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf.

Van Dijk, Teun A. 2006. 'Discourse and manipulation'. *Discourse and Society* 17 (3): 359–383. https://doi.org/10.1177/0957926506060250.

Van Dijk, Teun A. 1998. 'Opinions and ideologies in the press'. In *Approaches to Media Discourse*, edited by Allan Bell and Peter Garrett, 21–63. Oxford: Blackwell.

Van Dijk, Teun A. 1979. 'Pragmatic connectives'. *Journal of Pragmatics* 3 (1979): 447–456.

Van Dijk, Teun A., and Ruth Wodak, eds. 2000. *Racism at the Top*. Klagenfurt: Drava.

Widdowson, Henri G. 1995. 'Discourse analysis: a critical view'. *Language and Literature* 4 (3):157–172.

Wodak, Ruth. 2001. 'The discourse historical approach'. In *Methods of Critical Discourse Analysis*, edited by Ruth Wodak and Michael Meyer, 63–94. London: Sage.

Yule, George. 1996. *Pragmatics*. Oxford: Oxford University Press.
Žagar, Igor. 2010. '*Topoi* in Critical Discourse Analysis'. *Lodz Papers in Pragmatics* 6 (1): 3–27. https://doi.org/10.2478/v10016-010-0002-1.
Zifonun, Gisela. 2002. 'Grammaticalization of perspectivity'. In Graumann and Kallmeyer, *Perspective and Perspectivation in Discourse*, 89–109.

CHAPTER 8

A linguistic analysis of nationality-based hate speech on Facebook

The case of the Italian language

Alicja Paleta
Jagiellonian University

Anna Dyda
Jagiellonian University

Abstract

This chapter addresses some linguistic characteristics of hate speech in social media. The research was carried out on the basis of a corpus in Italian comprising posts and comments published on two Facebook groups: Italiani a Cracovia (Italians in Krakow) and Italiani in Polonia (Italians in Poland). These groups have about 26,500 members, mainly Italians who live or plan to come and live or travel in Poland and Poles who for various reasons are linked to Italian culture and/or language. This is an important factor as the heterogeneity of the group has an impact on the

How to cite this book chapter:
Paleta, Alicja, and Anna Dyda. 2024. 'A linguistic analysis of nationality-based hate speech on Facebook: The case of the Italian language'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 281–313. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-8.

language used within it, which is also varied. The study analysed utterances on different topics in order to understand whether and how the idea of belonging to a given nation (in this particular case, Italy or Poland) can form the basis of hate speech.

The analysis revealed significant variation in the manifestation of hatred that can be expressed through the use of specific words, such as slurs or vulgarisms, but also through grammatical choices, for instance pronominal contrasts. The research also confirmed that not only is hate speech transmitted lexically and grammatically but also through context-dependent irony and cynicism.

Keywords: hate speech, Italian language, social media, Facebook, nationality, in-group versus out-group

8.1 Introduction

This chapter aims to identify and describe some of the characteristics of hate speech (HS) in social media by means of a case study on the role of national identities. It focuses on examining posts and comments in Italian published in two Facebook groups mainly addressed at Italian expatriates in Poland,[1] in order to verify whether they contain elements of HS referring to the two involved nationalities, namely Italians and Poles, as well as to establish the linguistic elements through which this type of HS is manifested. The research question is whether and how the idea of belonging to a given nation (in this particular case, Italy or Poland) can form the basis from which HS develops in the online interactional dynamic of virtual communities. The analysis takes into account the linguistic mechanisms through which HS is expressed towards a group of people or a member of such a group. The category of nationality was chosen because it is an area in which latent

1 The features of the two groups will be specified in Section 8.4. For now, it is important to note that since the early 2010s Poland has become a popular destination for Italians in terms of long-term stays (mainly for work purposes), which is why some virtual communities (such as Facebook groups) have appeared, grouping primarily Italians but also Polish nationals among their members.

(linguistically covert, implicit) HS is particularly insidious. Since designations of nationality, such as nationality adjectives ('Italian', 'Pole'), are neutral *per se*, they do not fall within the markers that guide automatic detection of HS within online social media. Hence, a broader consideration of the context and of the specific linguistic mechanisms adopted is needed in order to single out harmful messages.

Section 8.2 presents the prerequisites of the analysis, by discussing some terminological and conceptual considerations. In Section 8.3 we talk about Facebook and the procedures that the platform implements in order to counter hatred. In Section 8.4 we outline the specific field of investigation, the corpus, and the objectives of the research, after which we move on to the results of the linguistic analysis of the manifestations of HS in Section 8.5. Section 8.6 presents the conclusions and possible extensions of the research, with notes on the relevance of this work for the fields involved in the study of HS.

8.2 Hate speech: general considerations

It is undeniably true that social media has become part of everyday life. It allows us to communicate, to stay informed about current events, to develop our interests, and to exchange opinions, but it also exposes us to HS, as witnesses or even victims. The link between social media and HS is indeed increasingly central to debates on communication that take place with regard to new social media (see, e.g., Daniels 2008; Foxman and Wolf 2013; Baider 2020;[2] Banaji and Bhat 2022). Because of its heterogeneity, the issue is studied from the perspective of a range of disciplines, from politics and social sciences (Van Blarcum 2005; Bleich 2011; Brown and Sinclair 2020), through jurisprudence (Casarosa 2020; Guillén-Nieto 2023), to linguistics and philosophy of language

2 This contribution is included in a special issue of *Pragmatics and Society* (Baider, Millar, and Assimakopoulos 2020) dedicated to HS that contains further relevant discussion on HS in online communication.

(Bianchi 2014, 2015a,b, 2021; Cepollaro 2015; Brindle 2016; Knoblock 2022; Guillén-Nieto 2023). Although the phenomenon has been the subject of extensive research, it has not yet been possible to arrive at a transdisciplinary definition of HS either at a global or European level. In an attempt to better characterise the subject of our research, we first turned to the legal field and, specifically, to the legislation in force within the European Union.

In 1997 the Council of Europe expressed its opinion on the matter in a Recommendation of the Committee of Ministers (see also Chapter 1):

> the term 'hate speech' shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin. (Council of Europe 1997)

This first definition, which is still very generic, emphasises two elements: namely, different forms of expression—though these are not further specified—and target groups who are defined on the basis of geographical and ethnic origin.

At the European level, the issue was taken up again in 2008, when Framework Decision 2008/913/JHA was published (Council of the European Union 2008). The document specifies the content of the aforementioned Recommendation, and defines hatred, encompassed by the term 'hate speech', as based on prejudices relating to race, colour, religion, ancestry, and national or ethnic origin, which are concepts protected by law. At the same time, all Member States are invited to take the necessary measures to prevent the incitement of hatred expressed towards a group of people or its members.

Subsequently, in 2016 the European Commission against Racism and Intolerance published General Policy Recommendation no. 15, which was approved in December 2015 and is devoted entirely to HS. It promotes a definition that encompasses the

assumptions expressed up to then by the community bodies, specifying and expanding them:

> Hate speech … entails the use of one or more particular forms of expression—namely, the advocacy, promotion or incitement of the denigration, hatred or vilification of a person or group of persons, as well any harassment, insult, negative stereotyping, stigmatization or threat of such person or persons and any justification of all these forms of expression—that is based on a non-exhaustive list of personal characteristics or status that includes 'race', colour, language, religion or belief, nationality or national or ethnic origin, as well as descent, age, disability, sex, gender, gender identity and sexual orientation. (Council of Europe 2016)

In fact, while the concepts discussed above are specified in the General Recommendation (forms of expression and target groups), a margin of freedom remains: indeed, the list of characteristics on the basis of which the target groups are defined is to be considered 'non-exhaustive'.

In 2016, based on the Framework Decision of 2008, the Code of Conduct on Countering Illegal Hate Speech Online was developed (European Commission 2016). This was initially an agreement between the European Commission and Facebook, Microsoft, Twitter, and YouTube, with other platforms joining later. Its purpose is 'to prevent and counter the spread of illegal hate speech online', while keeping alive the principle of freedom of expression. The internal regulations of social media platforms such as Facebook, YouTube, or Twitter (now X) can also constitute another source of input for a potential definition of HS. In an interesting attempt at theoretical synthesis, Fortuna and Nunes (2018) analyse the conditions and terms in force on the various platforms, and on the basis of these elements they propose a definition which contains an important reference to 'different linguistic styles, even in subtle forms or when humour is used' (Fortuna and Nunes 2018: 2), which opens the way to considering also irony and sarcasm, which are not susceptible to automatic detection in the same way that concrete words are, for example. In fact, compared to the con-

siderations previously cited, this definition allows the inclusion of different nuances that the language expresses which, as we will see, will be particularly important in the case of the corpus analysed in this chapter. Given the current fast development of social media, however, it seems that even the cited definition needs some further expansion to take into consideration linguistic phenomena concerning HS that can be observed in that environment.

Moving on to the field more directly linked to language, of particular importance is the work of Claudia Bianchi (2014, 2015a,b, 2021), who deals with HS from the perspective of the philosophy of language and emphasises the harmful effect of any HS element.[3] She takes into consideration both the descriptive and performative dimension of the language:

> If in a descriptive perspective language is a simple mirror of society and as such *reflects* phenomena, classifications, hierarchies and social conflicts, in a performative perspective linguistic practices, strictly connected to collective practices, contribute to *creating* and transforming social objects and therefore to building, reinforcing or revoking classifications, hierarchies and conflicts. (Bianchi 2021: 9)[4]

If the object of this research were to be limited to the theoretical considerations mentioned above—which, for obvious reasons, cannot be understood as exhaustive—it could be said that by HS we mean linguistic expressions of any kind that are addressed to groups or individual persons as members of such groups, which

---

3 Bianchi is by no means the only scholar to adopt this perspective. Her work constitutes the point of reference for this chapter as it primarily concerns the Italian language.

4 Unless otherwise specified, the translation of Italian sources into English is by the authors. 'Se in una prospettiva descrittiva il linguaggio è semplice specchio della società e come tale *riflette* fenomeni, classificazioni, gerarchie e conflitti sociali, in una prospettiva performativa le pratiche linguistiche, strettamente connesse a pratiche collettive, contribuiscono a *creare* e trasformare gli oggetti sociali e quindi a costruire, rinforzare o revocare classificazioni, gerarchie e conflitti.'

convey contempt, derision, and the like, based on a range of specific factors such as origin, ethnicity, religion, sexuality, and so on. To ensure that our study provides a complete picture, however, we must bear in mind that HS is closely linked to the intrinsic characteristics of language. In other words, it is necessary to consider the fact that the illocutionary and perlocutionary force of a speech act is very often not only dependent on the denoting meanings of a given expression, but is based on the context in which it is used.[5] The issue in question is addressed, among others, by Brambilla and Crestani (2021), who adopt the distinction between open and latent HS, where the former is based on statements in which hatred appears as explicit (e.g. *Gli immigrati sono ladri* [The immigrants are thieves][6]), while the latter refers to cases in which hatred is hidden behind irony or sarcasm, or behind a symbol or a graphic form (such as in the replacement of some letters with signs like the asterisk: *n*gro*, *c*zzo*[7]). The coded character of the latent form of HS makes it essential to give it due importance, as content characterised by 'hidden' hatred easily escapes the automatic detection systems on social media and can therefore be visible to users for a long time, hence doing greater damage.

8.3 Facebook and online HS

As already mentioned, one of the most fertile terrains for the development of HS is the online social network environment. For this reason, our research is oriented towards Facebook (henceforth FB). It is one of the signatory platforms of the Code of Conduct and therefore operates according to various rules designed to avoid the dissemination of hateful content. In fact, the community

5 For the notion of HS from the perspective of speech act theory see Bianchi (2015a, 2021).

6 In this text, examples of hateful expressions are cited in italics. For each expression its English translation is also provided, which is presented in square brackets. If the English equivalent is missing, an approximate translation is reported.

7 The coding refers to the word *negro* [nigger] and *cazzo* [dick].

standards section includes a complex definition that refers to all the documents previously discussed, and in which HS is defined as a direct attack against people on the basis of 'protected characteristics' that are analogous to what we have observed above (race, colour, religion, ancestry, national or ethnic origin, gender, sexual orientation). There is a direct reference to harmful stereotypes, exclusion, and negative comparisons as prohibited mechanisms. In addition to the previous definitions cited, here we also find age as one of the determining factors and, among the protected categories, migration is explicitly mentioned (Facebook 2021).

The document repeats the concept of the target groups defined on the basis of characteristics protected by law. The list is long because it contains not only the consolidated elements that appear frequently in theoretical discussions on HS but also concepts that were rarely mentioned previously, such as sexual orientation, sex, gender identity, and serious illness. There is also an attempt to specify the forms of expression that HS takes, such as stereotypes, various types of statements, comparisons, and more.

In 2019 FB published a report on the actions taken by the entities that signed the Code of Conduct in which it declared that it had a global network of about 15,000 employees responsible for the review of the content posted on the site. The document also shows that, on average, 89 per cent of the content reported as offensive is examined within 24 hours and that the removal rate is around 70 per cent (Council of the European Union 2019).[8] However, the question of the impossibility of drawing a clear limit between offensive content and freedom of expression was reiterated.

Even though automatic detection of HS has consistently improved, it still presents some issues: besides the difficulties concerning latent HS, an even broader obstacle is represented by the lack of an international agreement on the definition and classification of HS. We are also witnessing rapid social and linguistic development with the constant extension of which characteristics are considered to define minority groups and as such make them

8 The data date back to the reporting period in 2019.

worthy of protection. Automatic detection systems hence appear to operate on the basis of unstable rules to determine whether content is marked as HS. The work of the algorithms, based on artificial intelligence, and the work of human moderators can sometimes lead to confusing situations; for example, the phrase 'Irish teenagers are dump' would not automatically be classified as HS while 'Irish women are dump' would be reported as offensive. This is due to the fact that women constitute a protected social category and are thus recognised by the automatic detection system, while adolescents do not constitute such a category (Fortuna and Nunes 2018: 6).

In our opinion, both sentences cited above fall within the definition of HS that we adopt and on which our research is based (linguistic expressions that convey a spectrum of hatred, addressed to groups or individual persons as members of such groups), because they communicate contempt (i.e. a demeaning attitude) and hostility (i.e. an aggressive attitude) against a social group based on national identity. Since, in the examples above, national identity is referred to by means of a nationality adjective that is, in principle, neutral and objective, automatic detection may fail to recognise this form of HS, for which the surrounding context plays a crucial role. For this reason, although we acknowledge that FB has already initiated a number of procedures to counter HS, we have decided to examine a corpus based on statements taken from FB, taking as a starting point the hypothesis that automatic detection and user complaints fail to identify all forms of expression of hatred. Automatic work will never be able to replace human work, since, alongside words and expressions with an evident derogatory potential (e.g. *frocio* [faggot], *crucco* [Kraut]), there are words and expressions that in themselves do not necessarily carry hatred, but become offensive in certain contexts which Bianchi (2021) calls 'denigratory uses'. In effect, the utterance *Si salvano solo i pierogi poi il resto e tutto munnezza* [Only the pierogi are worth saving; the rest is all rubbish] (see (19) below), written by a FB user of Italian descent who lives in Poland, shall be considered HS as the offensive charge increases due to the context of use.

The decision to base the research specifically on the language used on FB was motivated by the nature of the platform itself: FB is a social means of communication that has undergone the fastest development in the social media era and can, at the time of writing, be defined as a multifunctional platform that no longer solely facilitates digital social exchanges between friends. Over time, it has acquired new functions, including doing business, making donations and fundraising, and advertising and promotion. A further possible and very popular function is the creation of virtual communities of people with common interests and/or needs, known as 'groups', which can be public (membership through acceptance by the administrator) or private (membership upon invitation of an administrator or a member). FB users (with personal profiles) must be real individuals who provide some personal data upon registration; once a profile is created, however, it can work with a false name and surname or even a nickname, therefore making it impossible for a general user to establish the true identity of those who choose not to use their real name. This certainly allows and, perhaps, also invites greater freedom of expression compared to face-to-face communication, in which the interlocutors are obviously identifiable. Thus, the specific affordances of the virtual environment (besides potential anonymity, also the facility of reaching a target and the opportunity to find a sympathetic audience in a 'group', leading to attitude reinforcement) may amplify tendencies towards aggression and abuse that exist in the offline world, negatively influencing the production, consumption, and distribution of contents on the social platform and resulting in a polarisation of discourses in online communication (see, e.g., the review by Walther 2022).

8.4 Outline of the field of investigation and the corpus

In our data collection, we assumed that HS always involves reference to a group, understood in the broad sense of the term, and that it is not limited to explicit manifestations, but can assume

latent forms, conditioned by the context. On this understanding, we analyse the phenomenon of HS directed against two nationalities, namely Italians and Poles in the content published on two public FB groups which take nationality as a fundamental defining aspect: Italiani in Polonia [Italians in Poland] and Italiani a Cracovia [Italians in Krakow], which between them, at the beginning of November 2021, had about 26,500 members (17,660 in Italiani in Polonia and 8,836 in Italiani a Cracovia). Both groups had been created with the aim of connecting Italians who live or intend to live in Poland and Krakow, respectively, and to exchange views and help each other in case of doubts or difficulties. Over time, however, many Italians who travel to Poland for short periods for tourism purposes have also joined the groups, as well as those who are only considering relocation, and, finally, numerous Poles who know Italian or for various reasons are linked to Italy.

Given the nature of the two groups, the published content is in Italian, although there is also a strong Polish language influence. In fact, is it linguistically a highly heterogeneous group, particularly with regard to diatopic and diastratic factors, as in both groups there are people who come from different parts of Italy and Poland, of different ages, sexes, and levels of education, and in most cases it is not possible to trace these data since the profile of the person does not provide them (sometimes it is even difficult to tell whether a profile belongs to an Italian or a Pole). It should also be borne in mind that the Poles present in the groups demonstrate a knowledge of the Italian language that varies from the fluency of a native speaker to only fairly basic with deficiencies in written language use. This results in frequent errors at all levels (orthographic, morphosyntactic, lexical, etc.) and in the presence of examples of code mixing (whether deliberate or not) between Italian and Polish.[9] There are of course also errors in the posts and comments published by the Italians, which is related to

9 The examples shown in this work have not been subject to any linguistic correction by the authors and are presented in their original form. They have been analysed in the original language in which they were

the authors' variable literacy competence, as well as to the digital nature of the published texts, to the speed with which posts are written, to the lack of editing. Posts and comments are rarely re-read before being published, differently from other types of digital communication, such as emails. There are in fact a number of characteristics that are typical of online texts, of which we list a few. With regard to Italian, the structural characteristics include a generally reduced use of punctuation: the most frequently used punctuation marks are full stops, question marks, exclamation signs, commas, and ellipses, while the colon and the semicolon are rarer. On the morphosyntactic side, the use of personal pronouns is reduced while that of connectives and discursive signals, employed to create textual cohesion within and across posts and comments, seems to be increased. Differences are also observed on the lexical level, enriched by a vast spectrum of neologisms, including calques, anglicisms, acronyms, and adaptations.[10]

In order to examine and verify the presence of elements of nationality-driven HS referring to Italians and Poles, as well as to establish by which linguistic elements this type of HS is manifested, we identified relevant material dating from a specific period (November 2021 to February 2022). The examination was carried out manually by reading each post, in the original language in which it was composed, with the respective comments, in search of statements that contained negative references to the two nationalities in question. We then moved on to the analysis of HS manifestations based on the concept of nationality.

The results presented in the next section must be understood in purely qualitative terms, as it was not possible to establish a precise number of utterances in the corpus due to the fact that it would have been too challenging to transfer the linguistic mate-

published. The English translations are only working translations meant to make the content accessible to a broader audience.

10 Studies on the language used on the web are already numerous. To deepen the issue, by way of example, see Fiorentino (2007), Prada (2003), Rossi (2010), Miłkowska-Samul (2019).

rial into a file that would allow us to proceed with quantitative analysis. The process of reading the posts and comments lasted several days during which the corpus was repeatedly modified by the users of the two groups, who could always delete and modify all posts and comments (the content was read directly from FB). In fact, some comments were deleted or suppressed, and others were added. In addition, users who publish comments often 'divide' them into several parts, posting them separately even if they constitute a single message. We do not therefore have the necessary information to carry out a quantitative analysis. This limitation, however, does not impact significantly on our results, given the way in which we frame our research question: our aim is to demonstrate that, despite the work of the automatic detection systems implemented by digital platforms, the utterances that convey HS can still appear on social media as HS is not limited to a set of words and/or expressions that can be captured by algorithmic formulae. Nationality-driven HS makes this particularly clear. We have thus focused on the qualitative side, to discover what escapes the automatic tools and is not reported by users or is not classified as offensive by FB staff.

The idea of nation, namely 'a large group of people of the same race who share the same language, traditions, and history, but who might not all live in one area',[11] is strongly present in the corpus. This is inevitable given that the discussions under analysis are between the representatives of two different peoples who are describing, comparing, and evaluating life in two different states (which implies contact between two cultures).[12] In an environment in which two peoples clash, situations that underline the contrast between the two nations often arise, creating a fertile ground for HS. When the utterances in which we noticed the presence of

11 Cambridge Dictionary, s.v. 'nation (*n*)', accessed 20 May 2022, https://dictionary.cambridge.org/dictionary/english/nation.

12 The contrast between two cultures represented by two nations points inevitably to the idea of the stereotype. The perspective adopted in this contribution is linguistic, but it is worth mentioning studies such as Tajfel (1982), Yzerbyt and Demoulin (2010), Fiske (2017).

hatred based on the idea of belonging to a nation were identified, we proceeded with the analysis in order to categorise the expressions of HS with respect to the linguistic features in which we are interested.

The linguistic material collected for the purposes of this research includes utterances selected on the basis of the presence of linguistic elements that can denote HS with reference to Italian and Polish nationality. Within the meaning of HS we shall include: i) a reference to the group, and ii) some offensive content against the group itself.[13] In the first (manual) phase of analysis, we began by identifying expressions of nationality, attempting to reveal in this way the discourses that genuinely referred to the nationalities in question. We consequently identified: i) nationality adjectives and state names used as metonymies (Alfonzetti 2019: 75), in which the concept of nation/nationality is transferred to the names of states (Italians → Italy; Poles → Poland); ii) personal pronouns and place adverbs referring to the two nations and states; and iii) cultural traits characteristic of a nation and used to describe the entire concept of nation/nationality. Having identified the ways in which reference is made to the Italian and Polish nations respectively, we moved on to identifying those statements that include expressions of hatred towards the nationalities in question. We found examples of various linguistic mechanisms used for this purpose, such as negative evaluation, direct contrast between two or more nationalities,[14] pronominal contrast *noi* [we] versus *voi* [you], but also *qui* [here] versus *là* [there], offensive criticism against cultural traits, use of hate words such as vulgar and derogatory terms, and so on.[15] In what follows, we present and comment on the examples of HS found in the corpus, divided by mechanism of reference to nationality.

13 For the components of slurs, see Cepollaro (2015: 155).

14 For a more in-depth account of the in-group and out-group dimension, see Russo and Tempesta (2017).

15 HS can manifest itself through very different linguistic mechanisms that can be classified as various speech acts, of which the most productive seems to be that of the insult. See in this regard Bazzanella (2020).

8.5 Analysis

8.5.1 'Gli italiani sono...': nationality adjectives and state names

We begin with constructions that use adjectives of nationality, in our case *italiani* [Italians] and *polacchi* [Poles], also expressed through the use of metonymies such as *Italia* and *Polonia*. In this section we want to show that the situation within the groups is not one-sided or one-directional—in other words, it is not only Italians who offend Poles or vice versa. The examples selected for this section show that Italy or Poland, Italians or Poles may and do become the object of offence by any member of the group. The nationality of the interlocutors is considered relevant here, and is expressed in square brackets, using the abbreviations IT for Italian and PL for Polish. The cases in which we do not know the speaker's origin are accompanied by a question mark. We also note whether the sentences refer to people (in general) from one of the two nations [of] or are addressed to a person who belongs to one of the two nations [to].

We shall consider the following examples:

(1) **italiani** sempre sono pronti a criticare tutto che non é come sono abituati loro. [PL of IT]

'**Italians** are always ready to criticise everything that differs from what they are used to.'

(2) **Italia** e 30 anni in dietro con tutto!!! Topi giganti X le strade, sporcizia dapertutto, criminalità. governo mafioso…….. però loro si vantano tanto poi 80/ della popolazione non conosce bene loro lingua e sapete cosa dichiarato ué settimana scorsa???? Che in Europa più ignorante popolo e **ITALIANI**. [PL of IT]

'**Italy** is 30 years behind with everything!!! Giant rats on the streets, dirt everywhere, crime, mafia government …….. but they brag a lot, then 80% of the population

does not know their language well and you know what the EU declared last week? That in Europe the most ignorant people are **ITALIANS**.'

(3) a) ma i polacchi non sono cretini e falsi come **gli italiani**. Si sono venduti per 4 denari. [IT of PL and IT]

'but **the Poles** are not stupid and false like **the Italians**. They sold themselves for pennies.'

b) chi si è venduto??? il popolo non si è venduto……la classe politica è allo sbando più completo, forse a loro ti riferisci……lascia perdere **i polacchi**……si vede che **la Polonia** la conosci poco……. [IT of PL and IT]

'who sold themselves? the people did not sell themselves……the political class is in complete disarray, maybe you are referring to them……forget **the Poles**…… it is evident that you know little about **Poland**…….'

(4) **i Polacchi** sempre a difendere tutto pure il 'lekktor' per voi e stupendo. [PL of PL]

'**the Poles** always defend everything, even the "lekktor" for you is wonderful.'

(5) **i polacchi** sono culturalmente più indietro, mangiano e bevono da schifo, altro che noi che sappiamo mangiare. [IT of PL]

'**the Poles** are culturally behind, they eat and drink disgusting stuff, unlike us who know how to eat.'

(6) solita cazzuta feminista è intollerante come tutte **le Polacche**. [IT of PL]

'she is the usual stupid feminist and is intolerant like all **Polish women**.'

(7) Ti assicuro che in Italia è pieno di **donne polacche** che si prostituiscono, mi spieghi come mai **le nostre donne** non vengono in Polonia a prostituirsi? Vedete di fare meno i fenomeni. Mi spieghi te come mai **le vostre donne** si mettono insieme a italiani se **i vostri signori polachi** sono tutti ricchi e vivono in un paese meraviglioso? […] Dimenticavo **le nostre** hanno delle difficoltà a mettersi con un uomo più vecchio di 20 anni. **Le donne polacche** di mettono volentieri con un uomo più vecchio basta che ho il portafoglio più largo. [IT of PL]

'I assure you that Italy is full of **Polish women** who prostitute themselves, can you explain to me why **our women** don't come to Poland to prostitute themselves? Try to reduce the phenomena. Can you explain to me why **your women** get together with Italians if **your Polish gentlemen** are all rich and live in a wonderful country? […] I forgot that **our women** have a hard time dating a man 20 years older. **Polish women** happily get together with an older man as long as he has a big wallet.'

(8) poi non e che **qua in Polonia** la gente sia specialmente intelligente… [IT of PL]

'then it's not that **here in Poland** people are especially intelligent…'

(9) Aspettative, sogni, ambizioni… cosa ha da offrire a un quindicenne **l'anonima Polonia** rispetto alle mille possibilita' di una grande citta' statunitense come per esempio New York? E poi vuoi metter l'inglese con il polacco? daaaai Disegnino? [IT of PL]

'Expectations, dreams, ambitions… what does the **anonymous Poland** have to offer compared to the thousand possibilities of a big American city like New York for example to a fifteen-year-old? And then you want to put English with Polish? Come onnnn Disegnino?'

In (1) and (3a)–(7), nationality adjectives have been used directly, while in (8) we see a periphrasis, *la gente in Polonia* [people in Poland]; in (2), (3b), and (9), instead, there are metonymies. The examples considered have different syntactic structures (Alfonzetti 2019: 76), but the most frequent is presented as '*x* is/are (not) O', in which *x* is the subject to which HS is addressed (*italiani/Italia* [Italians/Italy]; *polacchi/Polonia* [Poles/Poland]) and O is the offensive expression in question. For this phenomenon, consider examples (1) and (2) as well as (3a), (5), (7), and (8). In the remaining sentences, either the verb *essere* [to be] is omitted, as in (4) and (6), or some different linguistic elements are used, such as other verbs as in (7) or descriptive adjectives as in (9).

The offensive expressions against Italians relate to culture and temperament (*cretini* [cretins], *falsi* [fake people]), while the basis for contempt towards the Poles is spread across cultural traits, as in (4) and (5), stereotypes about Polish women, as in (6) and (7), reference to intelligence, as in (8), and the characteristics of the country itself, as in (9).[16]

8.5.2 Adverbs of place and personal pronouns

Moving on to the presence of adverbs of place and pronouns, we must start from the concept of these parts of speech as they are used in Italian. Treccani reports that 'Adverbs of place are used to specify the place of an action, the location of a person or object in space and the distance of a person or object from the speaker or listener',[17] whereas for pronouns—and in our case, subject pronouns in particular—it should be noted that, 'Unlike other languages (such as English and French), in Italian the expression of the personal pronoun subject is almost always optional

---

16 For a more specific lexical analysis, see Dyda and Paleta (2023).

17 Treccani (2012), s.v. 'luogo, avverbi di' [adverbs of place], https://www.treccani.it/enciclopedia/avverbi-di-luogo_%28La-grammatica-italiana%29/: 'Gli avverbi di luogo servono a specificare il luogo di un'azione, la collocazione di una persona o di un oggetto nello spazio e la distanza di una persona o di un oggetto rispetto a chi parla o ascolta.'

and not mandatory. It is indispensable, however, when it serves to avoid ambiguities, or in emphatic expressions'.[18] Two essential facts emerge from these definitions: first, adverbs can denote the distance between the speaker and the interlocutor or object to which they refer, and second, the presence of subject pronouns is optional and has an emphatic function, particularly with regard to the idea of the contrast between me/us and you/them.[19] Consider the following examples:

(10) **qui** i divorzi partono dai 18 anni in su, fanno i figli quando sono ubriachi e il giorno dopo non si ricordano neanche come si chiamava quello che le ha farcite qualche ora prima e il gioco è fatto, si lasciano e tutti vissero felici e contenti. [IT of PL]

'**here** they start getting divorced at 18, they have children when they are drunk and the next day they don't even remember the name of the one who "filled" them a few hours before and that's it, they break up and everyone lived happily ever after.'

(11) a) Fanno la fila per delle schifezze. Poi **qua** tutti fanno le stesse cose. Uno fa una cosa e tutti a fare lo stesso. Sembrano dei robot che eseguono gli stessi comandi. [IT of PL]

18 Treccani (2012), s.v. 'personali, pronomi' [personal pronouns], https://www.treccani.it/enciclopedia/pronomi-personali_%28La-grammatica-italiana%29/: 'A differenza di altre lingue (come l'inglese e il francese), in italiano l'espressione del pronome personale soggetto è quasi sempre facoltativa e non obbligatoria. È indispensabile, però, quando serve a evitare le ambiguità, oppure in espressioni enfatiche.'

19 The pronominal contrast is one of the techniques exploited by HS to differentiate between two groups—a belonging group and the other reference group. To receive the same contrast effect, adverbs of place are used to differentiate two places and to create a distinction between two realities, thus referring to the peoples or people who live there. On the subject of pronouns used as a reinforcement of stereotypes and prejudices, see Cruschina (2020).

‘They line up for some crap. Then everyone **here** does the same things. One does one thing and everyone does the same. They look like robots executing the same commands.’

b) esatto, come le scimmie delle multinazionali. Ma dopotutto, dopo Trieste in poi tipo a Cracovia c’è un alto livello di disperati… La Polonia come anche gli altri Paesi dell’est stanno ripulendo l’Italia per fortuna. [IT of PL and IT]

‘that’s right, they’re like corporate monkeys. But after all, from Trieste onwards, like in Krakow there is a high level of desperate people… Poland as well as the other Eastern countries are cleaning up Italy thankfully.’

(12) le **vostre** fanno le corne con un vicino non hanno nemmeno vergogna. [IT of PL to PL]

‘**your** [women] cheat on you with a neighbour, they are not even ashamed.’

(13) e già.. la ‘vodka’ **li** avrà bruciato il cervello. [PL of PL]

‘and yes.. the “vodka” will have burnt **their** brains.’

(14) scusa so che se la volpe non arriva all’uva dice che è acerba(se non sai cosa vuol dire acerba poi te lo spiego) ma **voi** ancora usate il filo rigido per fare gli impianti elettrici, facendo murandolo al muro, sperò che tu sappia che da **noi** sono 40 anni che non esiste più. [IT of PL to PL]

‘sorry, I know that the fox said that the grapes were sour when he could not reach them (if you don’t know what sour means then I’ll explain it to you) but **you** still use the rigid wire to make electrical systems, chasing the wire

into the wall. I hope you know that this system has not existed **here** for 40 years.'

(15) Godete**vi** la Polonia, le sue belle città e soprattutto la cortesia e l'educazione dei suoi cittadini, a casa **nostra** merce sempre più rara. In alternativa tornate**ve**ne al paesello a mangiare la parmigiana di mammà e lasciate spazio a chi vive all'estero, con curiosità intellettuale e senza cercare di rinchiudersi nel proprio ghetto mentale. [IT to IT]

'Enjoy Poland, its beautiful cities and above all the courtesy and education of its citizens, an increasingly rare commodity in **our** home. Alternatively, go home to eat mama's parmigiana and leave space for those who live abroad, with intellectual curiosity and without trying to lock themselves up in their own mental ghetto.'

(16) **Io** da italiano dico che **gli italiani** sono una brutta razza specialmente chi sta fuori casa sono i peggiori non fanno altro che criticare si credono che stanno in Italia che si siedono a tavolino e parlano male di uno e di un altro. [IT of IT]

'As an Italian, **I** say that **Italians** are a bad breed, especially those who are away from home. They are the worst, they only criticise, they believe they are in Italy where they sit at a table and speak badly of one and another.'

In (10) and (11a), sentences produced by Italians who already live in Poland, reference is made to Poland through the use of the adverbs of place *qui* and *qua* [here]. In (11b), as a response to the statement (11a), the implied subject refers directly to Poles. In all cases, discussing both Italy and Poland, the third person is used, which means that the HS employed refers to a third party, and not directly to the interlocutor. For the use of adverbs of place, we shall consider example (8), which is strengthened by the direct reference to the state.

Moving on to the use of pronouns in the examples we have identified, it can be assumed that their function is primarily to highlight and give particular emphasis to the subject within the sentence. In (14) and (15) the pronouns are used to identify the subject, which would remain unidentifiable without the context. In (12) *le vostre* [your] refers to Polish women, while in (13) the pronoun refers to all Poles. Pronouns, in addition, play a very significant role in the particular type of HS that is manifested under the guise of pronominal contrasts. In (14) we see the opposition between *voi* [you] and *noi* [we]—where 'you' refers to Poles and is opposed to 'us', Italians. In (15) we see a particular situation in which the contrast between *loro* [they] and *noi* [we] is emphasised, but direct contempt is pronounced towards the second-person plural, namely *voi* [you]. Pronominal contrasts also can serve to express a distancing from a group (*io* vs *loro* [I vs they]), as in (16).

Moreover, it is not uncommon to see mixed situations in which both nationality adjectives and pronouns are used—this is the case of the strengthening function of the pronoun. A similar case is observed in (2), where next to Italy, reference is made to *loro* [they], meaning Italians.

It should consequently be noted that the pronominal contrast not only constitutes a method in itself but is also used together with other HS elements in order to emphasise the message, further attracting the reader's attention. It can reinforce, for example, the use of adjectives of nationality and other references to subjects, as in (7), where we see the contrast between *le nostre donne* [our women] and *le vostre donne* [your women]. Example (5), using both the adjective of nationality (*polacchi* [Poles]) and the pronoun (*noi* [we], meaning Italians) strongly emphasises the contrast that the speaker assigns to *loro polacchi* [they, the Poles] and *noi italiani* [we Italians]. In this case, the contrast could also be motivated by the idea behind the FB group itself, in that it was created by Italians for Italians (us) and all the others are considered as third parties (them).

Example (4) is of great relevance and interest. It is produced by a Pole against other Poles, where the sender firstly speaks of Poles as *loro* [they] (i.e. *i Polacchi*, 'the Poles' in the text), and then refers to his compatriots with the pronoun *voi* [you], thus creating a strong distancing effect between *io* [I] and *loro/voi* [they/you]. This distancing, as well as the comparison with others, constitutes the basis for HS as they presuppose the existence of something that is better and something that is worse, as in the example in (14), where life in Poland and in Italy is expressly compared, giving great prevalence to the latter.

8.5.3 Cultural traits

Cultural traits that are characteristic of a certain nation and that are used to describe it as a whole constitute another means of referring to the group that has been identified in the corpus.[20] These references, through offensive and critical words, form part of the HS phenomenon. The following examples may be considered:

(17) A riguardo mangiare e naturale che se vuoi paragonarlo al cibo italiano è una Schifezza. [IT to PL]

'As far as eating is concerned, it is natural that if you want to compare it to Italian food it is rubbish.'

(18) sono le foto della mensa della Caritas? [IT of PL]

'Are these photos from the Caritas canteen?'

(19) Si salvano solo i pierogi poi il resto e tutto munnezza. [IT of PL]

'Only the pierogi are worth saving; the rest is all rubbish.'

20 By cultural traits we mean what is referred to in translation studies as culture-specific items (Snell-Hornby 1988: 1–2; Aixelà 1997: 56–57) or cultural words (Newmark 1988: 4).

(20) Continuate a votare PIS per un paese sempre più violento, razzista ed omofobo. Alla fine PIS è anche una rappresentazione abbastanza indicativa di come sono la maggioranza dei polacchi. Sennò questo partito non sarebbe al governo. [IT of PL]

'Continue to vote PiS for an increasingly violent, racist and homophobic country. In the end, PiS is also a fairly indicative representation of what the majority of Poles are like. Otherwise this party would not be in government.'

(21) Popolo di paraculi. [IT? of PL]

'A population of opportunists.'

(22) in italia sti fenomeni votano lega e fratelli d'Italia, statisticamente abbiamo meno imbecilli. [IT to PL]

'in Italy these phenomena vote Lega and Fratelli d'Italia, statistically we have fewer idiots.'

(23) Ma quanto fanno schifo quegli italiani che, pur di difendere PiS che nessuno ha tirato in ballo, arrivano quasi ad esaltare il pestaggio di loro connazionali? Geni, per PiS voi siete stranieri quanto un musulmano o un africano… [IT of IT]

'But how disgusting are those Italians who, in order to defend PiS that no one has mentioned, almost come to exalt the beating of their compatriots? You are geniuses, for PiS you are as foreign as a Muslim or an African…'

In (17)–(19) HS is manifested through the criticism of Polish cuisine, which is compared to its Italian counterpart (17), and again demonstrates the concept of the contrast between two nations as a carrier of hatred; in other examples, such as (18) and (19) it is criticised without any reference to other cultures. Example (11a) is

also linked to cuisine, as the term *schifezze* [rubbish] refers to the sausages that are sold in Krakow from a travelling van. Sentences (20)–(23) refer, in turn, to the political sphere: (20)–(22) refer to Polish politics, while (23) refers to Italian politics. An interesting example can be found in (22) as it first criticises some Italians, namely, those who vote for Lega and Fratelli d'Italia, and on the basis of this negative assessment it then goes further, criticising Polish people. It should also be noted that there is a strong contextual bond between (18) and (21) in which it would be quite difficult to identify the object of the utterances. Example (18) refers to a photograph that shows some typical Polish products, such as *pierogi* (a kind of traditional dumpling) and sausages, while in example (21) the speaker talks about the behaviour of Poles with regard to the war in Ukraine.

The variety of uses analysed confirmed that HS can manifest in very different ways. Not only can a nation be insulted by direct reference to the state or to the nation itself, as in (1)–(9) or (16), but the contempt can be also conveyed through various less direct elements, such as pronominal contrasts or cultural traits. HS can therefore manifest itself through the lexicon and syntactic structures, thanks to which the utterances acquire an offensive power, expressing judgement or derision, but it can also be hidden in the context, appearing through irony or cynicism, as for instance in (18).

8.5.4 The idea of belonging to a group

Taking up the concept of group belonging, which in the present study refers to a social group defined by nationality, an important consideration is how the speaker identifies themselves. The fact of belonging to a social group does not determine a real mental identification with this group. It is possible for a speaker to identify themselves with the nation to which they belong, while at the same time maintaining a distance from it by placing themselves in

the position of an external observer/commenter.[21] Given the binational character of the two FB groups (Italian/Polish), the phenomenon was observed from two perspectives, Italian and Polish. We have seen situations in which Italians speak of other Italians and Poles, and there are similarly Poles who make comments about other Poles and Italians. Starting from the Italian perspective—i.e., from the utterances posted by Italians—we observed that both Italians and Poles are described by other Italians either in the third-person plural (*loro* [they], as in (16) for Italians and (3a) for Poles and (3b) for both nations), or in the second-person plural, (*voi* [you], which can be seen in (15) referring to Italians while in (14) and (20) it appears to refer to Poles). It is not surprising that an Italian should use the second- or third-person plural towards Poles, since it is a people to which the speaker does not belong, so they put themselves in the position of a third party. It is interesting, however, when the distancing is created by an Italian who speaks of other Italians in the third-person plural as in (16) and (22).

Turning to the Polish perspective, we observed that when speaking of Italians, the reference is often made using generalisations (Italian = Italy) and *loro*, as in (2). An interesting case is again those Poles who, speaking of their compatriots, create a distancing effect by referring to them through the third-person plural, as in (4) and (13). In those cases in which the third-person plural is used towards the speaker's compatriots, it would be interesting to know whether the speakers are only distancing themselves from their group of origin, or if they are identifying themselves more closely with the other nation (perhaps due to the fact that the speakers already live abroad).

21 At this point we could open a parenthesis about belonging to the group and social identity. For reasons of space this issue will not be developed here but a more in-depth treatment of the topic can be found in Speltini and Palmonari (2007) and Bertani and Manetti (2007).

8.5.5 Themes and topics

Over the course of the analysis, a question arose regarding whether there are any issues that can spark HS in a particular way. Hot topics such as migration, gender issues, and ethnic minorities undoubtedly come into play. However, these issues are not often discussed in the groups that form the basis of this corpus; there are, however, some other subjects that are particularly fertile in terms of HS, leading to some intense exchanges of views, such as when Italian and Polish cuisine and politics were compared in our examples above. The analysis also highlighted that HS can arise from any type of conversation, regardless of its theme. Moreover, it often manifests itself in the chain situation in which an offensive comment engenders further comments of this kind, heightening the emotional charge.

Where the lexical field is concerned, it is worth noting that there are some words that underline the offensive nature of the utterances and these play a very important role in the identification of HS. Slurs such as *negro* [nigger], *frocio* [faggot], and *puttana* [bitch] are very significant in this field (Bianchi 2015b: 285). However, in the FB content that was analysed, these epithets were not used, and had we limited ourselves only to the presence of slurs, the results would have been negative. Nevertheless, in the examples presented there are some derogatory words and expressions that—in the contexts in which they appeared—were used towards the entire group (nation): *cretini* [idiot], *imbecilli* [stupid], *essere indietro* [to be backward], *essere poco intelligente* [to be unintelligent]. As can be seen from the examples, these are all terms that refer to limited mental capacity. Another lexical field comprises words that express disgust or that describe and object as rubbish, such as *schifo* [disgust] (also in the expression *far schifo* [to be disgusting]) and its derivatives *schifezze* [disgusting things] or *monnezza* [rubbish] (written as *munnezza* in the comment). There were also some vulgarisms such as *cazzuto* [badass] and references to ethnicity: *brutta razza* [bad breed].

Most of the offensive content present in the corpus was based, however, on irony and cynicism, for example: *sono dei robot* [they're robots]; *scimmie delle multinazionali* [corporate monkeys]—expressions used to describe Poles. In situations where it is more the context and the sensitivity of the interlocutors that carries the offence (in other words, in situations where terms of pure hatred, such as derogatory epithets, are not present), it is very difficult to establish the boundary between neutral content (freedom of expression) and HS content, pointing inevitably to the limitations of the automatic detection system.

8.6 Conclusions

As we have seen, the penetration of hostility connected to nationality is visible in various aspects of the content published on FB. This hostility can refer to an individual person or to a single person by virtue of their belonging to a certain social group. It is often difficult to establish whether a statement should be considered HS, due to the lack of a uniform definition of HS, and the contextual nature of HS itself, which has been highlighted in the analysis of the corpus presented in Section 8.5. In our study, we have attempted to identify the manifestations of HS that refer to a group of people on the basis of their nationality—Italian or Polish—as the corpus includes posts and comments that come from two public FB groups with both Italian and Polish members. Taking into consideration the content posted, we focused on the conversations that refer to the idea of the nation in a broad sense of the concept and that carry some offensiveness towards it. We have identified i) nationality adjectives and state names used as metonymies; ii) adverbs of place and personal pronouns referring to both nationalities in question; and iii) cultural traits characteristic of a nation that may reflect it as a whole. When accompanied by offensive content and, in some cases, contrasted one with another (e.g. pronominal contrasts), all these elements acquire offensive power by embodying contempt, judgement, or derision, thereby constituting HS. It should be borne in mind that such HS reduces

individuals to a single dimension, characterising them only on the basis of belonging to a group—in our case, the social identity derived from belonging to a nation. At the same time, however, it should be emphasised that the offensive content based on belonging to the Italian and/or Polish nation in the analysed corpus does not originate with the clear intention of calling on Italians to leave Poland, for example, which would suggest that they (Italians) are a hated native group. Instead, the topic of nation, as one of the topics triggering HS and recurring in its definitions, seems to fuel discussions about where a person lives best and why someone decided to emigrate. It still holds true that this topic easily leads to the escalation of differences of opinion within conversations.

It is clear, then, that discussing insults and hatred expressed online is essential, since such speech not only reflects our reality but also contributes to creating it in a certain way, inciting a greater aggression and offensiveness of language. It could be said that we are dealing with a 'chain reaction', in which an insulting post or comment constitutes a trigger for other even more offensive utterances. Furthermore, it is often the case that the initial post is absolutely neutral in terms of offensiveness but nonetheless immediately provokes insulting reactions towards the author or other commenters.

The study presented here must therefore be considered as just one of the voices in a discussion that must be further developed from the perspectives of disciplines such as linguistics, philosophy of language, psychology, and jurisprudence. The identification of existing manifestations of HS is only the first step along the arduous and complex road to the ultimate goal, namely the minimisation of HS content in social media. By minimisation we mean increasing user awareness about offensive content and the damage caused by this content, inviting mutual moderation in conversations that take place on the platforms; however, minimisation also requires engagement from the managers of the platforms themselves, given the weaknesses that have been revealed in the anti-hate systems currently in place.

However, it should also be noted that this type of study is not free of limitations. In addition to the impossibility of conducting an accurate quantitative analysis, it must be considered that the two authors of the present chapter are native speakers of Polish, of Polish nationality, and, therefore, the subjectivity of the point of view cannot be excluded. Furthermore, due to the nature of data collection and the impossibility of archiving material from FB, it is also possible that during the process of reading the linguistic material not all utterances that included elements of HS were detected.

The study also found some possible extensions of the research. For example, it would be interesting to analyse which linguistic mechanisms incite other expressions of insult/hatred, leading to the intensification of the emotional charge. It would similarly be interesting to verify whether, within the content identified for the purposes of this and similar studies, it is possible to observe interventions made by the users themselves aimed at mitigating the negative content that has been already published.

References

Aixelà, Javier Franco. 1997. 'Culture-specific items in translation'. In *Translation, Power, Subversion*, edited by Román Álvarez and Carmen-África Vidal, 52–78. Bristol: Multilingual Matters.

Alfonzetti, Giovanna. 2019. 'Gli insulti: alcuni criteri di categorizzazione'. In *Studi linguistici in memoria di Giovanni Tropea*, edited by Salvatore Carmelo Trovato, 67–78. Alessandria: Edizioni dell'Orso.

Baider, Fabienne. 2020. 'Pragmatics lost? Overview, synthesis and proposition in defining online hate speech'. *Pragmatics and Society* 11 (2): 196–218. https://doi.org/10.1075/ps.20004.bai.

Baider, Fabienne, Sharon Millar, and Stavros Assimakopoulos, eds. 2020. 'Hate speech: definitions, interpretations and practices'. Special issue, *Pragmatics and Society* 11 (2). https://doi.org/10.1075/ps.11.2.

Banaji, Shakuntala, and Ramnath Bhat. 2022. *Social Media and Hate*. London: Routledge.

Bazzanella, Carla. 2020. 'Insulti e pragmatica: complessità, contesto, intensità'. *Quaderns d'Italià* 25: 11–26.

Bertani, Barbara, and Mara Manetti, eds. 2007. *Psicologia dei gruppi: teoria, contesti e metodologie d'intervento.* Milan: Franco Angeli.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio.* Bari and Rome: Laterza.

Bianchi, Claudia. 2015a. 'Parole come pietre: atti linguistici e subordinazione'. *Esercizi Filosofici* 10 (2): 115–135.

Bianchi, Claudia. 2015b. 'Il lato oscuro delle parole: epiteti denigratori e riappropriazione'. *Sistemi intelligenti* 2: 285–302.

Bianchi, Claudia. 2014. 'The speech acts account of derogatory epithets: some critical notes'. In *Liber Amicorum Pascal Engel*, edited by Julien Dutant, Davide Fassio, and Anne Meylan, 465–480. Geneva: Université de Genève.

Bleich, Erik. 2011. *The Freedom to be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism*. Oxford: Oxford University Press.

Brambilla, Marina, and Valentina Crestani. 2021. '"Der kriminelle …", "I soliti …": parole dell'odio in tedesco e in italiano'. In *Il linguaggio dell'odio: fra memoria e attualità*, edited by Marilisa D'Amico, Marilisa Brambilla, Valentina Crestani, and Nannerel Fiano, 85–105. Milan: Franco Angeli.

Brindle, Andrew. 2016. *The Language of Hate: A Corpus Linguistic Analysis of White Supremacist Language*. New York and London: Routledge.

Brown, Alexander, and Adriana Sinclair. 2020. *The Politics of Hate Speech Laws*. New York and London: Routledge.

Casarosa, Federica. 2020. 'L'approccio normativo europeo verso il discorso dell'odio online: l'equilibrio fra un sistema di "enforcement" efficiente ed efficace e la tutela della libertà di espressione'. *Questione Giustizia*, 8 July 2020. Accessed 20 May 2022. https://www.questionegiustizia.it/articolo/l-approccio-normativo-europeo-verso-il-discorso-dell-odio-online-l-equilibrio-fra-un-sistema-di-enforcement-efficiente-ed-efficace-e-la-tutela-della-liberta-di-espressione.

Cepollaro, Bianca. 2015. 'Gli epiteti denigratori: presupposizioni infami'. *Esercizi Filosofici* 10 (2): 154–168.

Council of Europe. 2016. 'ECRI General Policy Recommendation no. 15: on combating hate speech'. Adopted on 8 December 2015. Strasbourg: Council of Europe, European Commission against Racism and Intolerance (ECRI). https://rm.coe.int/ecri–general–policy–recommendation–no–15–on–combating–hate–speech/16808b5b01.

Council of Europe. 1997. 'Recommendation No. R (97) 20 of the Committee of Ministers to Member States on "Hate Speech"', 30 October 1997. https://rm.coe.int/1680505d5b.

Council of the European Union. 2019. 'Assessment of the Code of Conduct on Hate Speech on line State of Play', information note sent to the Permanent Representatives Committee/Council (12522/19), 27 September 2019. https://commission.europa.eu/document/download/a5c92394-8e76-434a-9f3a-3a4977d399bb_en?filename=assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play_.pdf.

Council of the European Union. 2008. 'Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law'. Official Journal of the European Union L 328/55 (6 December 2008). https://eur-lex.europa.eu/eli/dec_framw/2008/913/oj.

Cruschina, Silvio. 2020. 'Il linguaggio d'odio'. *Linguisticamente*, 10 July 2020. https://www.linguisticamente.org/il-linguaggio-dodio/.

Daniels, Jessie. 2008. 'Race, civil rights, and hate speech in the digital era'. In *Learning Race and Ethnicity: Youth and Digital Media*, edited by Anna Everett, 129–154. Cambridge, MA: MIT Press.

Dyda, Anna, and Alicja Paleta. 2023. 'Vocabulary of hatred in social media: an example from Italian'. *Neuphilologische Mitteilungen* 124 (1): 103–127. https://doi.org/10.51814/nm.130138.

European Commission. 2016. 'Code of conduct on countering illegal hate speech online'. 30 June 2016. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

Facebook. 2021. 'Hate speech'. Facebook Community Standards. Version as updated on 24 November 2021. Accessed 20 May 2022. https://transparency.fb.com/policies/community-standards/hate-speech/.

Fiorentino, Giuliana. 2007. 'Nuova scrittura e media: le metamorfosi della scrittura'. In *Scrittura e società: storia, cultura, professioni*, edited by Giuliana Fiorentino, 175–207. Rome: Aracne.

Fiske, Susan T. 2017. 'Prejudices in cultural contexts: shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion)'. *Perspectives on Psychological Science* 12 (5): 791–799. https://doi.org/10.1177/1745691617708204.

Fortuna, Paula, and Sérgio Nunes. 2018. 'A survey on automatic detection of hate speech in text'. *ACM Computing Surveys* 51 (4): 1–28. https://doi.org/10.1145/3232676.

Foxman, Abraham H., and Christopher Wolf. 2013. *Viral Hate: Containing Its Spread on the Internet*. New York: St. Martin's Press.

Guillén-Nieto, Victoria. 2023. *Hate Speech: Linguistic Perspectives*. Berlin: De Gruyter.

Knoblock, Natalia, ed. 2022. *The Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive, and Dehumanizing Discourse*. Cambridge and New York: Cambridge University Press.

Miłkowska-Samul, Kamila. 2019. *(S)cortesia e social network: opportunità e rischi del dibattito pubblico su Facebook*. Warsaw: Wydawnictwo Uniwersytet SWPS.

Newmark, Peter. 1988. *A Textbook of Translation*. New York: Prentice Hall.

Prada, Massimo. 2003. 'Lingua e web'. In *La lingua italiana e i mass media*, edited by Ilaria Bonomi, Andrea Masini, and Silvia Morgana, 249–289. Rome: Carocci.

Rossi, Fabio. 2010. 'Internet, lingua di'. *Enciclopedia dell'Italiano Treccani*. https://www.treccani.it/enciclopedia/lingua-di-internet_(Enciclopedia-dell'Italiano).

Russo, Ernesto, and Pablo Bernardino Tempesta. 2017. 'Categorisation and defence strategies'. In *Online Hate Speech in the European Union*, edited by Stavros Assimakopoulos, Fabienne H. Baider, and Sharon Millar, 25–29. Berlin: Springer. https://doi.org/10.1007/978-3-319-72604-5_3.

Snell-Hornby, Mary. 1988. *Translation Studies: An Integrated Approach*. Amsterdam and Philadelphia, PA: John Benjamins.

Speltini, Giuseppina, and Augusto Palmonari. 2007. *I gruppi sociali*. Bologna: Il Mulino.

Tajfel, Henri. 1982. *Social Identity and Intergroup Relations*. New York: Cambridge University Press.

Van Blarcum, Christopher D. 2005. 'Internet hate speech: the European framework and the emerging American haven'. *Washington and Lee Law Review* 62 (2): 781–830.

Walther, Joseph B. 2022. 'Social media and online hate'. *Current Opinion in Psychology* 45 (June 2022): 101298. https://doi.org/10.1016/j.copsyc.2021.12.010.

Yzerbyt, Vincent, and Stéphanie Demoulin. 2010. 'Intergroup relations'. In *Handbook of Social Psychology*, vol. 2, edited by Susan Fiske, Daniel Gilbert, and Gardner Lindzey, 5th edn. New York: Wiley.

CHAPTER 9

Hate speech in non-cooperative contexts

Question types as a measure of cooperativity

Valentina Bianchi
University of Siena

Silvio Cruschina
University of Helsinki

Abstract

Research on hate speech has identified various aspects of social media that affect the speaker's attitude in this specific type of communication. In this chapter we discuss some structural aspects of the context of utterance as analysed in dynamic pragmatics, and we show that with respect to these, certain online contexts qualify as inherently non-cooperative; we hypothesise that non-cooperativity favours the emergence of excessive language and, in particular, of hate speech. To test our hypothesis, we analyse three small corpora of discussion threads from two different social platforms. We propose that different types of canonical and non-canonical

How to cite this book chapter:
Bianchi, Valentina, and Silvio Cruschina. 2024. 'Hate speech in non-cooperative contexts: Question types as a measure of cooperativity'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 315–346. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-9.

questions are indices of (non-)cooperativity, and we analyse their distribution in each discussion thread.

Keywords: online context, cooperativity, dynamic pragmatics, question types, canonical questions, non-canonical questions

9.1 Introduction

The proliferation of hate speech and excessive language on social media has become a central issue from various disciplinary perspectives. Research since the late 2010s converges on the view that alongside social, psychological, and ideological factors, the format and affordances of the medium itself plays a crucial role (for general discussion, see Brown 2018; Baider 2020; Biri 2023; Esposito and KhosraviNik 2024). The reciprocal invisibility between the conversational participants, due to the lack of visual or auditory contact, reduces empathy on the part of the author, as well as their sense of accountability and their moral engagement; the latter may be further reduced by the author's perceived anonymity in the online environment. The speed and instantaneousness of the medium may lead to a less reflective attitude, leaving considerable space for unfiltered emotive content. Moreover, in addition to the immediate addressee, the author is also aware of a potential remote 'audience', since the medium makes the written exchange available to any reader for an undefined period: thus, hate speech is often used to mark the author's affiliation to a generic community of like-minded users.

Besides these general factors, however, something specific to the discourse context must be at play, since not all discussion threads are conducive to hate speech. In this chapter we propose that one crucial aspect is the degree of cooperativity among participants.

Following Grice (1975), the conversation is a cooperative activity involving rational agents who jointly pursue a common discourse goal. This conception has been adopted and implemented in the framework of dynamic pragmatics that evolved from the seminal work of Stalnaker ([1978] 1999), which defines the *context of the conversation* by characterising different components

and the way they are involved in cooperative speech acts (see Section 9.2).[1] We will argue that the analysis of conversation that follows from this approach presupposes the stability of the conversational group for cooperative speech acts to be successfully performed (Section 9.3). We then extend this approach to online written exchanges on social platforms, assuming that a conversation context corresponds to *a single continuous discussion thread*. We characterise a certain type of online context as lacking to a significant extent the required stability, which makes it inherently non-cooperative (Section 9.4). We then hypothesise that the lack of cooperativity favours the emergence of hate speech and that, conversely, cooperative online conversation contexts show a comparatively low incidence of hate speech (Section 9.5).

To operationalise our hypothesis, we adopt the characterisation of speech acts in formal pragmatics, and we propose that certain types of questions—information-seeking questions and deliberative questions—are evidence for context cooperativity, as opposed to rhetorical questions. On these grounds, we expect that discussion threads with a higher proportion of questions of these types show a lower incidence of hate speech as compared to discussion threads where such questions are scarce (Section 9.6). As a proof of concept, we analyse three small corpora composed of discussion threads from two social networks, Facebook and Reddit (sections 9.7 and 9.8). We compare one corpus from Facebook and one from Reddit where the discussion thread concerns gender identity, and we observe that the two corpora differ significantly, in line with our expectation; this shows that it is not the topic *per se*, however potentially divisive, that triggers hate speech. We then compare the first Facebook corpus to another Facebook corpus whose discussion threads concern potentially less divisive topics (the life of Italian immigrants in Finland): again, we observe a stark difference, which leads us to conclude that the social media platform *per se* is not crucial either: what is relevant is the specific conditions under which the online conversation develops.

1 On the connection with Grice's approach, see Stalnaker (2002: 702–705).

9.2 The dynamic pragmatic approach

In the approach starting from Stalnaker ([1978] 1999) and Lewis (1979), the conversational context is represented in the terms of possible world semantics. Each conversation is based on a set of propositions that constitute the participants' common ground: the presupposed propositions that each participant accepts as true for the purposes of the conversation, and assumes to be accepted by all participants.[2] In a non-defective context, all the participants implicitly agree on which propositions are in the common ground (Stalnaker [1978] 1999: 84–85).[3] The common ground circumscribes a region of logical space: the subset of possible worlds in which all the propositions are true, dubbed the 'context set'. The multiplicity of worlds in the context set represents the fact that the common ground information is partial and leaves various possibilities undecided.

In a typical conversation, the essential goal is to increase the information jointly accepted by all the participants. A proposition is *informative* relative to a common ground if and only if it is not true in all the worlds of the context set, but it is true in some of them (i.e. it is neither entailed by the context set, nor inconsistent with it).[4] When an informative proposition is asserted by one participant and is jointly accepted by the conversational group, it is added to the common ground and it eliminates from the context set those worlds in which it is not true. Thus, the increase in shared information reduces the region of logical space to be taken into account (Stalnaker [1978] 1999: 86).

---

2 The propositional attitude of acceptance, or 'common belief', is a public attitude, whereby each participant is committed to act as if the common ground propositions were true in the actual world. For an in-depth discussion of the formal properties of the relevant attitude relation, see Stalnaker (2002: 706–708).

3 'The information state will include two different kinds of information: first, information about the participants in the conversation—about what they know about each other and their common environment; second, information about the subject matter of their discourse' (Stalnaker 2018: 384).

4 See Stalnaker's ([1978] 1999: 88–89) principle I.

The increase in shared information does not proceed randomly, but is guided by discourse goals, modelled as *questions under discussion* (QUDs; Ginzburg 1996; Roberts 1996, 2018). When a question is asked, its effect is to *partition* the context set into disjoint subsets of worlds, each corresponding to a possible answer. The discourse goal, then, is to select one of these subsets.

By way of illustration, consider a toy context with just two restaurants A and B. The question in (1) partitions the context set into the four cells schematically represented in Figure 9.1.[5]

(1) Which of the two restaurants has a vegan menu?

Each cell corresponds to a complete (exhaustive) answer. It is possible, however, to provide a partial answer, by breaking down the QUD into two sub-questions:

(2) Does restaurant A have a vegan menu?

(3) Does restaurant B have a vegan menu?

A yes-answer to (2) retains only the upper half of the context set, while a no-answer retains only the lower half. Symmetrically, a yes-answer to (3) retains only the left-hand half of the context set, and a no-answer retains the right-hand one. Each of these answers is relevant to the super-question (1)—since it discards some cells from the context set—but it is partial in that it does not single out a unique cell (Roberts 2012: 11–12).[6] Notably, different QUDs induce different partitions on the context set, which predetermine some specific possible evolutions of the common ground. Thus, for the conversation to proceed, the addressee must proffer an answer that is as informative as possible, or else indicate that no informative answer can be provided.

5 We adopt here the partition semantics for questions (Groenendijk and Stokhof 1984) because it allows for a neat and concise definition of a discourse goal.

6 The addressee can assert a reply from which an answer can be obtained via inference. We leave aside this phenomenon, limiting ourselves to an overview of the general approach.

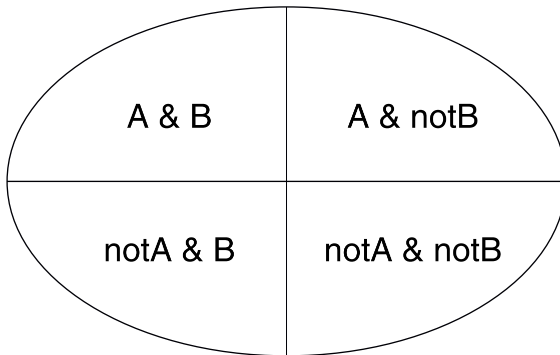


Figure 9.1: Partition of the context set.

9.3 Context stability

The view of the cooperative conversation outlined above rests on an implicit assumption—namely, that the context is (relatively) stable. Firstly, the group of participants is stable, in that when a participant joins or leaves the group, this is explicitly acknowledged by all the others. For instance, if a newcomer is assumed to lack some common ground information, the other participants will inform them, so that they will be able to contribute to the conversation in a relevant and informative way.[7]

Crucially, group stability guarantees that the common ground can be *monotonically updated* by new information—that is, an assertion can be straightforwardly assessed as compatible with the current common ground and informative with respect to it, and if it is, it can update the common ground without requiring any revision. In addition, the common ground of the conversation can be *non-generic*—that is, it may contain information that is only shared by the specific group in that specific moment. In turn, a stable common ground allows the group to pursue a common discourse goal by asking and answering QUDs.

7 See note 3 above.

Group stability and mutual acknowledgement also guarantee that every discourse move by a speaker is taken into account by all the others (even when the move is directed to a specific addressee in the group). In a situation of disagreement, where not all the participants accept as true an asserted proposition, the latter is not added to the common ground, but all the participants are aware of the incompatible commitments that have arisen; on the other hand, if no participant explicitly rejects an assertion, this typically counts as tacit acceptance by the conversational group, since acceptance can be considered the default reaction to an assertion (Farkas and Bruce 2010: 86, 99). When disagreement arises, it is often possible to open a conversational negotiation through which a shared commitment is eventually reached. If this is (assumed to be) impossible, the participants will agree to disagree. However, such a context of settled disagreement can still be cooperative: for instance, the participants may decide to pursue another relevant QUD.

It is important to stress that a given conversation context will display these properties to different degrees, and these properties may change in the course of the conversation. In a context where the participants' group is not fully stable, there may be a stable subgroup. Moreover, there are different proportions of common ground information specific to the conversational group; this in turn affects the range of possible QUDs. But crucially, there must be a stable core, however minimal, for cooperativity to be possible at all.

9.4 Unstable contexts

We defined above an online conversation context in a very narrow way, as a *single continuous discussion thread* starting from a post on a specific topic. Of course, the environment in which a post is published will already define a theme or orientation for the communication, so as to condition the participants' interests and beliefs: for instance, a post on a politician's Facebook page (see Corpus A in Section 9.7.3) will mostly attract comments from

users with an interest in politics, raising a number of expectations about the development of the discussion. The online environment thus constitutes a 'context' in a very broad sense, and some general information about it will be taken for granted by any participant. From the present perspective, however, what is relevant is the conversation dynamics, and for this reason we focus on the narrower notion of conversation context.

The type of conversation context that we are interested in is the discussion thread that follows a post on an open page, such that participation may be occasional. Here, the conversational group is unstable in a specific way: any participant can join in at any moment, participate for an undefined stretch of time, and leave at any moment, without prior notice and without the other participants being aware of it; note that this holds independently of how long and how often a participant intervenes. Because of this fundamental instability, replies or reactions to a discourse move are not guaranteed: even the smallest or most minimal responses may not be forthcoming. However, lack of reaction on the part of a given participant cannot be interpreted as default tacit acceptance, contrary to what may happen in face-to-face conversations. Whenever a reaction is manifested, this typically gives rise to interaction within a subgroup of participants – which is also unstable – and it is not possible to perceive its effect on any other participants in the discussion thread who do not directly intervene in it. The conversation context thus lacks an essential ingredient, namely mutual acknowledgement: since the discourse moves are not mutually acknowledged by a stable group of participants, it is virtually impossible—beyond the occasional interactions just mentioned—to maintain a common representation of how the conversation is evolving, and how the common ground is being updated.

Moreover, the very instability of the conversational group also implies that the participants can only presuppose a minimal and *generic* common ground; consequently, when a question is asked, it creates a partition on an exceedingly wide context set, and the pursuit of a specific discourse goal would preliminarily require an

unusually long chain of ordered QUDs. Under these conditions, it is extremely hard for the participants to engage in an exchange aiming at a common discourse goal. This implies that they will not be consistently committed to expanding the common ground by sharing new information among themselves.[8] For these reasons, we characterise this type of unstable context as *structurally non-cooperative*.

9.5 Our hypothesis

We have argued that in non-cooperative online contexts, speech acts do not—actually, *cannot*—aim to pursue an information-oriented discourse goal. The question, then, is what such speech acts aim at. We propose that they have an essentially *expressive* function: the author expresses their evaluation and their sentiment regarding the topic of the discussion thread (or some subtopic), and expects other participants (not *the* other participants, since the group is unstable) to express their own similar or opposite sentiment. Indeed, any (pseudo-)factual information is reported to justify their expressive stance.[9]

These speech acts have two alternative essential goals. One is that of maintaining the author's inclusion in an undefined 'in-group' of like-minded users, by manifesting and encouraging agreement, so as to yield the so-called 'chorus effect'.[10] The

8 Again, we may find fragments of cooperative exchanges among subgroups of participants, but crucially, the discussion thread as a whole does not pursue a general discourse goal.

9 In an information-oriented conversation, expressive content is present but is not at issue. See Potts (2007) for an approach to expressive content in possible world semantics.

10 This can be viewed as the common goal of a group of individuals. Note, however, that this type of cooperativity does not correspond to our definition: such a common goal is not a *discourse* goal—that is, it is not aimed at sharing information that restricts the context set. It is also important to emphasise that the actual communication expectations may depend on the affordances of the specific virtual environment—that is, the technical features of a social media platform that enable and

in-group does not coincide with the conversational group, which is unstable, nor with a recognisable collective entity; it is an imagined entity, whose membership condition is declaring oneself a member of it.[11] The alternative goal is to achieve intrusion into a discussion thread characterised by some 'out-group' stance that is perceived as opposite to one's own.

Note that when participants in such a discussion thread are expected to be largely like-minded, the expressive and emotive content conveyed by the author may well go unnoticed within the general chorus effect. We hypothesise that this is what leads the participants to intensify the content that they express on a relevant emotive and/or evaluative scale. This mechanism gives rise to excessive language, which exacerbates the polarisation between the perspectives of the in-group and the target out-group(s). Excessive language involves emotive, offensive, and aggressive communication that is not necessarily directed at specific groups or single individuals, as is the case with hate speech. However, the lack of information exchange easily shifts the target of excessive language from the topic itself to public persons who are perceived as representative of a target out-group, and to the participants who are perceived as opponents. Thus, excessive language easily degenerates into hate speech.

We thus propose that the purely expressive function of speech acts is linked to the structural non-cooperativity of the context. This leads us to expect that the less cooperative a context is, the more likely it is for excessive language to be found. In order to avoid a circular argument, it is necessary to identify some independent linguistic features that characterise (non-)cooperativity.[12]

constrain the actions and interactions of the network community (see Biri 2023 and references therein).

11 On the linguistic underpinnings of generic in-groups and out-groups, see the corpora analysis in Olmastroni, Bianchi, and Duguid (2021: 203–207).

12 As discussed above, disagreement does not *per se* imply lack of cooperativity: this is why we do not consider the presence of overt expressions of disagreement as a reliable indicator.

To this aim, in the next section we return to the notion of 'question under discussion'.

9.6 A typology of questions

In the framework described in Section 9.2, questions are normally asked in order to enhance the informative content of the participants' common ground, by partitioning the context set into disjoint alternatives (see the discussion around (1) above). Farkas (2020) proposes that at the pragmatic level, canonical questions are associated with the following default assumptions about the participants' epistemic states:

(i) *Open issue:* the speaker assumes that all the alternatives introduced by the question are neither positively nor negatively decided with respect to the current context set; in other terms, the context set neither entails not excludes any of the alternatives.

(ii) *Speaker ignorance*: the speaker presents themselves as having an epistemic state that does not support their commitment to any of the alternatives.

(iii) *Addressee competence*: the speaker presents themselves as assuming that the addressee's epistemic state supports the commitment to the 'true' alternative—that is, the cell of the partition that contains the actual world.

(iv) *Addressee compliance:* the speaker presents themselves as assuming that the addressee will resolve the issue by publicly committing to the true alternative. (Farkas 2020: 21)

Canonical information-seeking questions then require a cooperative addressee who fulfils (iv) by providing the most informative answer (complete or partial) that is supported by their epistemic state. Therefore, we can consider them as marking a cooperative dynamics. According to Farkas, non-canonical questions deviate from one or more of these assumptions.[13]

13 On non-canonical questions, see also Obenauer (2004), Garzonio (2004), Cruschina (2012), Giorgi (2016), Giorgi and Dal Farra (2019),

Deliberative self-addressed questions are non-canonical questions that introduce an open issue but do not require the addressee to provide the true answer as per (iii) and (iv):

(4) Should we go there by car or by train?

The goal of a deliberative question is thus to start the negotiation of a possible choice: the question is asked with the aim of partitioning the context set into disjoint cells that correspond to different action choices (Cariani, Kaufmann, and Kaufmann 2013), each of which is then cooperatively evaluated. Since there is no informational asymmetry between the participants, the default assumptions of speaker ignorance and addressee competence are suspended. However, deliberation-oriented questions introduce an open issue (i) and elicit a cooperative conversational dynamics; for this reason, we assume that they too characterise cooperative contexts.

Another type of non-canonical questions is *exclusively self-addressed questions* like (5):

(5) Teacher: Why do you have to show your work? Because I want to know how you reached the solution. (Farkas 2020: 24)

These questions introduce an open issue, as per (i), and are intended to increase the information publicly available to all participants; however, the speaker asks the question to point out an issue on which the addressee is assumed to be ignorant, and they immediately provide the answer. Here assumptions (ii)–(iv) are suspended; however, the question highlights an issue that is open in the current context set (i). Although commonly dubbed 'rhetorical', this question type introduces a sequence of speech acts that are cooperative, in that they are aimed at enhancing the participants' common ground.

Bianchi and Cruschina (2022), Cruschina and Bianchi (2022a,b), among others.

Real rhetorical questions, instead, do not comply with the open issue assumption:

(6) Should we leave them to drown in the sea?

The speaker does not assume that the alternatives in the question denotation are undecided, as per (i): on the contrary, they ask the question to emphasise that the resolution of the issue is obvious to everyone (this implies that addressee competence is assumed). Thus, the question does not ask for an answer on the part of the addressee, nor does it elicit a cooperative discourse dynamics (see Biezma and Rawlins 2017). Indeed, rhetorical questions have the effect of marking one of the alternatives as *non-negotiable* (Farkas 2020). We wish to stress that rhetorical questions are not *per se* an indication of context non-cooperativity: they can be used by an author to emphasise a point, or to mark expected similarity with the other participants' stances. It is rather the scarcity of the other question types (on the total number of questions asked) that characterises a non-cooperative conversation context.

To summarise, within this approach, canonical, deliberative, and self-addressed questions signal an information-oriented attitude on the part of the speaker, who assumes that they are participating in a cooperative exchange. For the purposes of our analysis, these three types are grouped together under the label 'cooperative questions'.[14] On the other hand, rhetorical questions are not assumed to introduce an open issue and do not require the addressee to answer; therefore, they are compatible with a non-cooperative context.

Based on this typology, our hypothesis can be operationalised in the following way: the more cooperative questions are found in a discussion thread (on the total number of questions asked), the less hate speech is expected to emerge. Conversely, in a context with a high incidence of hate speech, we expect a low proportion of cooperative questions. In the next section we present a proof-of-concept study conducted through corpus analysis.

14 Short for 'questions marking a cooperative conversation context'.

9.7 Corpora analysis

9.7.1 Independent factors

In designing our proof-of-concept study, two factors independent of our hypothesis were taken into account that might influence the incidence of hate speech in a discussion thread. The first one is the nature of the topic under discussion. It is likely that a highly controversial issue would arouse heightened emotions and would engender among the participants a less reflective attitude that could be conducive to hate speech. In this case, the very nature of the topic could lead the conversational community to a radical polarisation independently of the structural aspects of the context that we discussed above.

We believe that there is indeed a topic effect. Note, however, that if this were the main factor, it should condition the incidence of hate speech independently of the type of context in which it is discussed. In contrast, from our perspective, a polarised conversation context can still remain cooperative in the sense that we define above; in this case, a controversial issue can be discussed without giving rise to excessive language and hate speech.

The second independent factor is the social platform on which a discussion thread develops. It is probably the case that different social platforms differ with respect to their prevalent use in the online community, as well as in the ethical guidelines stated in the terms and policies with which users must agree when signing up. We claim, however, that the social platform should not be taken as the relevant notion of 'context': it is at the level of a single discussion thread that a conversational community arises and its dynamics is deployed. We will indeed show that one and the same social network can host quite different types of contexts.

In order to check for these independent factors we carried out two pairwise comparisons. In the first comparison, we built two corpora of discussion threads concerning two issues related to the LGBTQIA+ community, which were *equally controversial* in the Italian public debate at the time of observation. Corpus A is a corpus of Facebook comments about equality between homosexual

and heterosexual couples. Corpus B is a long discussion thread on Reddit about the nature of transgender identities, on which different participants express quite different views. The rationale is that if the controversy of the topic is the main conditioning factor, then we should expect a similar incidence of hate speech in the two corpora.[15]

In the second pairwise comparison, we compared Corpus A to another Facebook corpus, Corpus C, containing five discussion threads about the life of Italian immigrants in Finland, where the issues raised are unrelated to those of corpora A and B and are not particularly controversial. In this case, if the social platform is the main conditioning factor, we should expect a similar incidence of hate speech in A and C.

9.7.2 Criteria

In our analysis, we manually counted the number of posts/comments containing hate speech and the occurrence of different types of questions in each corpus, following the typology outlined in Section 9.6. For the purposes of this study, we adopted a fairly broad definition of hate speech, which subsumes the more recent definitions with an emphasis on the aggressive side of hate speech (see, e.g., Bianchi 2021), but which also includes excessive language not directly aimed at a specific individual or group (e.g. swear words). We thus classified as hate speech the following expressions:[16]

- insults, swear words, irony with explicit reference to sex;
- slurs, including innovative slurs referring to a political stance, e.g. *sinistroidi* 'leftoids', *pidioti* 'PD idiots';[17]

15 On the differences between the two social media platforms in terms of technological affordances, see Biri (2023) and references therein.

16 Note that this classification is functional to the specific goals and data of this study and may therefore differ from other definitions of hate speech (see Chapter 1).

17 PD (*Partito Democratico*, the Democratic Party) is a social-democratic political party in Italy.

- evaluative epithets such as *questa incapace* 'this incompetent [woman]'. Crucially, in this use the negative evaluation is not part of the at-issue content, but it is presupposed and hence not subject to explicit negotiation (Potts 2005; Murray 2014);
- verbs in an irrealis mood denoting violent actions, e.g. *frustarla sulle chiappe e poi mandarla ai lavori forzati* '[we should] whip her butt and then give her hard labour'.

All of these expressions were classified in the general category of hate speech. We did not classify as hate speech other forms of irony, GIFs, and emoji, except for one case: when an emoji was used to substitute a word used as a derogatory term—for example, the emoticon for 'shit' referring to an individual.

We manually counted the occurrences of questions classified as follows:

- canonical information-seeking questions, which manifest the speaker's intention to gain information from the interlocutor(s) by introducing a QUD;
- canonical questions directly answered by another participant, which manifest the occurrence of a cooperative exchange;
- deliberative questions, which address an open issue and invite the participants to express their evaluation of alternative action choices;
- rhetorical questions, which are characterised by the lack of an open issue and thus do not elicit an answer;
- self-addressed questions, through which the speaker introduces an open issue and provides the answer themselves. As discussed above, these differ from rhetorical questions proper in that they are intended to structure an argumentation providing information to the interlocutor, and in this respect can be considered cooperative speech acts.

We collected the results from each corpus in tables, and then compared the tables (see sections 9.7.2.–9.7.4 and Section 9.8).

9.7.3 Corpus A

Corpus A consists of two discussion threads from the official Facebook page in Italian of Simone Pillon (@SenatorePillon), a Member of Parliament who famously opposed the 'DDL Zan', a bill that concerned, most prominently (at least in public opinion), the legitimacy of equal rights between homosexual and heterosexual couples.[18] The first discussion thread is dated 27 October 2021—the end of the first parliamentary debate on the DDL Zan; the second is dated 13 January 2022—the date of a second discussion round for a modified version of the DDL. The two discussion threads consist of 246 and 166 comments, respectively, for a total of 412 comments, amounting to *c.*10,000 words. The initiating posts were not included in the corpus, since they were not introduced by a participant in the discussion but rather by the page owner or manager.

Most hate speech in the discussion threads was produced by users who expressed hate or disgust toward the Member of Parliament himself and his stance on the relevant issue. Table 9.1 shows the number of comments containing hate speech.

Table 9.1: Hate speech in Corpus A.

	Comments with hate speech	Total comments	%
Post 1	60	246	24.4%
Post 2	27	166	16.2%
Corpus A total	87	412	21.2%

18 The *disegno di legge* (DDL) Zan is a bill proposed by Alessandro Zan, a Partito Democratico MP. In Italy, a DDL is the initial phase of the process in which proposed new legislation is introduced by one or more members of parliament. The bill contains a set of articles that need to be discussed and (eventually) approved, one at a time, by the different branches of the parliament before becoming an effective law.

In Corpus A, we find a significant proportion of comments with hate speech. As we can see, the number of posts with hate speech is different in the two discussion threads, amounting to 60 out of 246 in the first thread and 27 out of 166 in the second thread. Overall, 21.2 per cent of comments in Corpus A contain hate speech.

In (8)–(10) we reproduce some examples of comments that were classified as containing hate speech. (For privacy reasons, we do not provide the users' names.)

(8) Devi morire! Ammazzato!!!!!

'You must die! Killed!!!!!'

(9) Figa se mi fai vomitare

'Fuck, you make me vomit'

(10) Sei una merda senza fine

'You're an endless [piece of] shit'

The manually counted totals of the different question types are summarised in Table 9.2.

Table 9.2: Question types in Corpus A.

	Canonical Qs not answered	**Canonical Qs answered**	**Deliberative Qs**	**Exclusively self-addressed Qs**	**Rhetorical Qs**	**Total**
Post 1	12	3	2	1	25	43
Post 2	13	1	1	0	73	88
Corpus A total	25	4	3	1	98	131
%	19.1%	3.1%	2.3%	0.8%	74.8%	100%

We observe a very low proportion of cooperative questions, which are significantly outnumbered by rhetorical questions (see examples (11)–(13)), the latter amounting to 78 per cent of the total. Among the former, the low number of answered questions shows the low incidence of cooperative exchanges addressing a QUD.

(11) E tu quando ti togli dal cazzo?

‘When are you getting the fuck out of here?’

(12) Perché tua mamma non ti ha ingoiato?

‘Why didn’t your mum swallow you?’

(13) Che squallore certi contenuti fb dovrebbe bloccarli, cosa si prova nel sentirsi così inutile?

‘What squalor! Certain fb contents should be blocked; what does it feel like to feel so useless?’

Finally, the virtual lack of self-addressed questions in this corpus is compatible with the absence of structured argumentation.

9.7.4 Corpus B

Corpus B consists of a single discussion thread published on Reddit on 3 May 2021.[19] The author who initiated the discussion thread published a long and detailed post entitled ‘Persone transgender e identità di genere’ (Transgender persons and gender identity), which was followed by a discussion between a limited number of participants. The discussion thread contains 216 posts for a total of *c.*20,000 words. The ratio shows that the posts in this thread were, on average, significantly longer than those in Corpus A.

19 CYP4502D6, ‘Persone transgender e identità di genere’, Reddit (r/italy), 3 May 2021, https://www.reddit.com/r/italy/comments/n3rftm/persone_transgender_e_identit%C3%A0_di_genere/.

The same classification criteria were adopted as for Corpus A. Crucially, the topic of transgender identities was controversial in the discussion, where different positions were expressed. Despite this, the ratio of hate speech is negligible, as shown in Table 9.3: virtually all instances of hate speech consisted in the expression of a heightened tone through swear words.

Table 9.3: Hate speech in Corpus B

	Comments with hate speech	Total comments	%
Corpus B	7	216	3.2%

On the other hand, the manual count of different question types yielded very different results from Corpus A, as shown in Table 9.4.

Table 9.4: Question types in Corpus B.

	Canonical Qs not answered	Canonical Qs answered	Deliberative Qs	Exclusively self-addressed Qs	Rhetorical Qs	Total
Corpus B	16	32	2	20	38	108
%	14.8%	29.6%	1.8%	18.5%	35.2%	100%

We can observe that 18.5 per cent of the posts in Corpus B contain self-addressed questions, which suggests the presence of structured argumentation in the discussion thread. The percentage of canonical questions (14.8 per cent) and especially of answered canonical questions (29.6 per cent) shows that there was a significant incidence of cooperative exchanges driven by QUDs or attempts at initiating them. The combined percentage of canonical questions, answered or otherwise, amounts to 44.4 per cent

and exceeds that of rhetorical questions (35 per cent). In (14) an example is reported of an answered canonical question, while in (15) we can see examples of canonical questions consecutively produced by the same speaker:[20]

(14) A: Il problema di fatto è che le definizioni di genere, identità di genere, uomo e donna sono o vuote anche nell'astratto o con conseguenze pratiche non ideali. Tu che definizioni daresti a quelle 4 parole?

'The actual problem is that the definitions of gender, gender identity, man and woman are either empty also at an abstract level, or they have undesirable practical consequences. How would you define those 4 words?'

B: Io darei le definizioni che ho utilizzato nel post.

'I would give the definitions that I used in the post.'

C: Provo a rispondere io, anche se, tieni presente, al momento ciascuno usa i vari termini un po' come preferisce perché non c'è un consenso netto

'I'll try to answer on my part, even though, you should realise, as of now everyone uses the terms as they prefer, because there is no clear consensus.'

(15) Hai fonti da linkare? Ci sono affermazioni e conclusioni da parte di studiosi? O più semplicemente potresti motivare la tua affermazione?

'Do you have sources to link? Are there statements and conclusions by scholars? Or could you simply justify your statement?'

20 In the following, capital letters are used to list examples from different speakers and to list their conversational turns (see (14), (18), and (19)), while the use of lower-case letters, as in (16), indicates that the examples were consecutively produced by the same speaker.

The examples in (16) illustrate two consecutive self-addressed questions, answered by the speaker themselves, while in the comment in (17) we find a rhetorical question:

(16) a. Allora perché non siamo tutti transgender? La differenza sta nella cosiddetta DISFORIA DI GENERE.

'So why aren't we all transgender? The difference lies in so-called gender dysphoria.'

b. Cos'è la disforia? È un sentimento spiacevole che si prova nei confronti di se stessi e di come si viene percepiti dagli altri in merito al proprio GENERE.

'What is dysphoria? It is an unpleasant feeling that one experiences about oneself and about how one is perceived relative to one's gender.'

(17) Per le persone transgender non capisco questo accanimento sulla questione, dopo un accurato controllo medico e psicologico, se i medici lo ritengono necessario che problema c'è?

'For transgender people I do not understand such doggedness on the issue, after an accurate medical and psychological examination, if the physicians think it [transition] necessary what's the problem?'

The results summarised in Table 9.4 show that the overall discussion thread qualifies as a (mostly) cooperative context in the sense defined above.

9.7.5 Corpus C

Corpus C was taken from the Facebook private group *Gruppo degli italiani in Finlandia* (Italians in Finland), at a time when the members numbered around 4000. The corpus consists of eight discussion threads that appeared between 14 and 29 September

2021. The total number of posts and comments is 290, for a total of *c.*10,000 words. In this case, as for Corpus B, the initiating posts were counted together with the comments because, unlike in Corpus A, they were introduced by an actual participant.

The discussion threads were about various aspects of life in a foreign country, with an emphasis on bureaucratic procedures and the issue of learning the official language; the posts mostly focused on sharing useful information. Two threads were initiated by a participant who was planning to move to Finland, and another participant who was planning a holiday there. In both cases, the long-term Italian immigrants responded by providing information and at the same time commenting on living conditions in Finland.

The manual count followed the same classification criteria as in corpora A and B. Table 9.5 shows that the percentage of comments with hate speech is 0.

Table 9.5: Hate speech in Corpus C.

	Total comments	**Comments with hate speech**	**%**
Corpus B	290	0	0%

The occurrences of question types are reported in Table 9.6.

Table 9.6: Question types in Corpus C.

	Canonical Qs not answered	**Canonical Qs answered**	**Deliberative Qs**	**Exclusively self-addressed Qs**	**Rhetorical Qs**	**Total**
Corpus C	20	35	0	1	10	66
%	30.3%	53.0%	0%	1.5%	15.1%	100%

The distribution shows a very high percentage of canonical questions, most of which (53 per cent) were answered. There are virtually no self-addressed questions, which is compatible with the lack of structured argumentation in the posts. The percentage of rhetorical questions is 15.1 per cent; all in all, the ratio of rhetorical to non-rhetorical questions is almost symmetrical to that of Corpus A. In (18) and (19) we present examples of cooperative exchanges in Corpus C:

(18) A: Buongiorno, adesso vivo in Finlandia posso rimanere solo 3 mesi. Ma la mia ragazza è finlandese posso rimanere qui oltre questo tempo visto che siamo una coppia? Cosa potrei fare? Grazie

'Good morning, I currently live in Finland, I can only stay for 3 months. But my girlfriend is Finnish, can I stay here longer since we are a couple? What could I do? Thanks'

B: Buonasera. Sapresti lavorare come cameriere?

'Good evening. Would you be able to work as a waiter?'

C: scusa ma che lavoro stai cercando? Qui la situazione non è delle più rosee.

'Sorry, but what kind of job are you looking for? The situation is not so good here.'

(19) A: Ciao a tutti, mi consigliate qualcosa di bello da fare per trascorrere il mio compleanno in famiglia? Abbiamo 2 bimbi di 6 e 8 anni. Non conosco granché essendo qui da poco più di un anno, magari qualcuno di voi ha qualche bella idea da propormi.

'Hello, everyone, could you suggest something nice to do for my birthday with my family? We have 2 children aged 6 and 8. I don't know much around here

since I have been here for little over a year, perhaps someone has a nice idea to suggest.'

B: Escape room, ce ne sono di semplici da fare coi bambini. Parco acquatico Flamingo. Se il tempo è bello, ci sono i parchi di arrampicata sugli alberi. A costo zero, passeggiata a Nuuksio con salsiccia grigliata. Cinema e pizza.

'Escape room, there are easy ones that you can do with children. Aqua Park Flamingo. If the weather is good, there are parks where you can climb trees. For free, a walk in Nuuksio with grilled sausages. Cinema and pizza.'

A: Abbiamo scelto Flamingo, grazie mille! Tu ci sei stata? Hai qualche consiglio da darmi? Basta portarsi solo costume e accappatoio o serve altro?

'We chose Flamingo, thanks a lot! Have you been there? Any advice? Do we need to take just swimming costumes and bathrobe or anything else?'

Most of the rhetorical questions in Corpus C, examples of which are presented in (20) and (21), came from a single discussion thread—that is, a post on integration and language policies in Helsinki:

(20) Abbassare le tasse? Ahah ma se sono più basse che in Italia

'Lowering the taxes? Haha, but if they are lower than in Italy'

(21) perché mai uno dovrebbe sbattersi ad imparare il finlandese, quando con le stesse qualifiche può andare altrove, con meno problemi, meno tasse, ed uno stile di vita anche migliore?

'why on earth should one make the effort to learn Finnish, when with the same qualifications one can go elsewhere, with less problems, less taxes, and an even better quality of life?'

9.7.6 Summary

The results from the three corpora are summarised in Table 9.7.

Table 9.7: Summary of results for all three corpora.

	Canonical Qs not answered	Canonical Qs answered	Deliberative Qs	Exclusively self-addressed Qs	Rhetorical Qs	Comments with hate speech
Corpus A	25	4	3	1	98	87/412
131 Qs	19.1%	3.1%	2.3%	0.8%	74.8%	21.1%
Corpus B	16	32	2	20	38	7/216
108 Qs	14.8%	29.6%	1.8%	18.5%	35.2%	3.2%
Corpus C	20	35	0	1	10	0/290
66 Qs	30.3%	53.0%	0%	1.5%	15.1%	0%

In Table 9.8, we group together all cooperative question types.

Table 9.8: Comparison of the three corpora: cooperative vs rhetorical questions.

	Cooperative questions	Rhetorical questions	Comments with hate speech
Corpus A	25.2%	74.8%	21.1%
Corpus B	64.7%	35.2%	3.2%
Corpus C	84.8%	15.1%	0%

9.8 Discussion

The trends that we expected are confirmed. There is, in particular, a dramatic split between Corpora B and C, where cooperative questions are significantly above 50 per cent, and Corpus A, where cooperative questions make up around 25 per cent of all questions; the incidence of hate speech in B and C is at least seven times less than in Corpus A.

On the other hand, the incidence of rhetorical questions is more than double in Corpus A with respect to Corpus B, and it is almost five times higher in Corpus A than in Corpus C. There still is a non-negligible proportion of rhetorical questions in B and C; however, as noted above, rhetorical questions *per se* are not incompatible with cooperativity. What is relevant, instead, is the incidence of cooperative questions that introduce, at least potentially, a QUD.

The asymmetries between the three corpora are even sharper if we compare the number of *answered* canonical questions, as shown in Table 9.9.

Table 9.9: Comparison of the three corpora: answered canonical questions.

	Canonical Qs answered	**Comments with hate speech**
Corpus A	3.1%	21.1%
Corpus B	29.6%	3.2%
Corpus C	53.0%	0%

We conclude that typology of questions seems to be a reliable indicator for distinguishing different types of conversation contexts. Following this criterion, the discussion threads in Corpus B and those in Corpus C qualify as significantly more cooperative than the two discussion threads in Corpus A. The comparison of A and C shows that there is indeed a topic effect, such that a highly controversial topic, unsurprisingly, leads to polarisation.

But crucially, the topic *per se* is not a *sufficient* condition for the emergence of hate speech. In other terms, it is not only the topic that counts but also the conversation context in which it is discussed.

As a side note, we observe a difference between Corpora B and C when it comes to self-addressed questions. We take this question type to mark the presence of a structured argumentation in the discussion thread. The asymmetry corresponds to the different nature and goals of the contexts of the two corpora: while in B the discussion revolves around a general topic at an abstract ideological level, in Corpus C it mostly revolves around practical questions.

9.9 Summary and conclusions

In this chapter we have proposed an approach to hate speech that focuses on the intrinsic properties of the online conversation contexts in which it might emerge. Through the tools of formal pragmatics, we identified some properties that distinguish cooperative contexts from less (or non-)cooperative ones. In less cooperative contexts, the conversational group is unstable and, for this reason, they do not share a specific common ground; reactions to any speech act are not guaranteed, and it is virtually impossible to establish a common discourse goal to be pursued through QUDs. We argued that under these conditions, the participants' speech acts have a merely expressive function, conveying the author's sentiment, and are aimed at an ideologically homogeneous group, for the author to be recognised as a member of that group (or to intrude into it). To produce an expressive speech act that will have a significant impact, the author is led to express extreme evaluations or sentiments, which may lead to hate speech against a specific target: a person or an out-group. This led us to hypothesise that unstable, non-cooperative conversation contexts are more likely to host hate speech.

To operationalise our hypothesis, we adopted a typology of questions that distinguishes those that potentially introduce a

QUD from those that do not; we assumed that the former are an indicator of context cooperativity. We conducted a proof-of-concept study of three corpora, which suggests that the rate of cooperative questions is a reliable indicator of a significant property of online conversation contexts.

We are aware that our results are far from conclusive: the number of corpora analysed and their dimensions are very limited. The procedure of manual counting is time-consuming, and to the best of our knowledge, there is no way to automatically classify questions according to the typology that we adopted, because there are no systematic lexical or syntactic cues that distinguish the various types. However, we have developed a methodology that we hope has been shown to be promising. Note that the index that we have proposed here—question types—should in principle be valid across languages: thus, we envisage a possible cross-linguistic comparison of online exchanges in different linguistic communities. In future work we plan to investigate other possible indicators of context cooperativity.

We conclude with a bold and hopeful suggestion. At the beginning of the era of social media, one could hope that public debate would be enhanced and supported by easy access to information; but things turned out differently, with a proliferation of 'fake news' and of hate speech. We believe that by shedding light on the relevant *structural* properties of online contexts, an approach might emerge that goes beyond mere censorship of pathological online behaviour, and instead aims to raise the users' awareness of the conversational dynamics that characterise online contexts.

References

Baider, Fabienne. 2020. 'Pragmatics lost? Overview, synthesis and proposition in defining online hate speech'. *Pragmatics and Society* 11 (2): 196–218. https://doi.org/10.1075/ps.20004.bai.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.

Bianchi, Valentina, and Silvio Cruschina. 2022. 'Ignorance and competence implicatures in central Sicilian polar questions'. In 'Romance Languages and Linguistic Theory 17', edited by Ora Matushansky, Laurent Roussarie, Michela Russo, Elena Soare, and Sophie Wauquier, special issue, *Isogloss: Open Journal of Romance Linguistics* 8 (2): 1–20. https://doi.org/10.5565/rev/isogloss.117.

Biezma, María, and Kyle Rawlins. 2017. 'Rhetorical questions: severing asking from questioning'. *Proceedings of Semantics and Linguistic Theory* (SALT) 27: 302–322. https://doi.org/10.3765/salt.v27i0.4155.

Biri, Ylva. 2023. 'Stancetaking in Interest-Based Online Communities: A Corpus Pragmatic Comparative Analysis'. PhD thesis, University of Helsinki. https://helda.helsinki.fi/items/49596ae2-12c9-490f-b978-aef57bc36177.

Brown, Alexander. 2018. 'What is so special about online (as compared to offline) hate speech?' *Ethnicities* 18 (3): 297–326. https://doi.org/10.1177/1468796817709846.

Cariani, Fabrizio, Magdalena Kaufmann, and Stefan Kaufmann. 2013. 'Deliberative modality under epistemic uncertainty'. *Linguistics and Philosophy* 36: 225–259. https://doi.org/10.1007/s10988-013-9134-4.

Cruschina, Silvio. 2012. *Discourse-Related Features and Functional Projections*. Oxford and New York: Oxford University Press.

Cruschina, Silvio, and Valentina Bianchi. 2022a. 'The syntactic encoding of conventional implicatures in Sicilian polar questions'. *Probus* 34 (1): 77–109. https://doi.org/10.1515/probus-2022-0006.

Cruschina, Silvio, and Valentina Bianchi. 2022b. 'Variation at the syntax–pragmatics interface: discourse particles in questions'. *Languages* 7 (3): 183. https://doi.org/10.3390/languages7030183.

Esposito, Eleonora, and Majid KhosraviNik, eds. 2024. *Discourse in the Digital Age: Social Media, Power, and Society*. London: Routledge.

Farkas, Donka. 2020. 'Canonical and non-canonical questions'. Unpublished manuscript, University of California at Santa Cruz. Accessed 22 August 2024. https://semanticsarchive.net/Archive/WU2ZjIwM/questions.pdf.

Farkas, Donka, and Kim B. Bruce. 2010. 'On reacting to assertions and polar questions'. *Journal of Semantics* 27 (1): 81–118. https://doi.org/10.1093/jos/ffp010.

Fogal, Daniel, Daniel Harris, and Matt Moss, eds. 2018. *New Work on Speech Acts*. Oxford: Oxford University Press.
Garzonio, Jacopo. 2004. ‘Interrogative types and left periphery: some data from the Fiorentino dialect’. In *Quaderni di Lavoro ASIS* 4, edited by Barbara Patruno and Chiara Polo, 1–19. Padua: Unipress.
Ginzburg, Jonathan. 1996. ‘Dynamics and the semantics of dialogue’. In *Logic, Language, and Computation*, vol. 1, edited by Jerry Seligman and Dag Westerstahl, 221–237. Stanford, CA: CSLI Publications.
Giorgi, Alessandra, and Chiara Dal Farra. 2019. ‘On the syntax/pragmatics interface: expressing surprise and disapproval’. *Intercultural Pragmatics* 16 (3): 335–361. https://doi.org/10.1515/ip-2019-0017.
Giorgi, Alessandra. 2016. ‘On the temporal interpretation of certain surprise questions’. *SpringerPlus* 5 (1): 1390. https://doi.org/10.1186/s40064-016-2951-5.
Grice, H. Paul. 1975. ‘Logic and conversation’. In *Syntax and Semantics*, vol. 3: *Speech Acts*, edited by Peter Cole and Jerry L. Morgan: 41–58. New York: Academic Press.
Groenendijk, Jeroen, and Martin Stokhof. 1984. ‘Studies on the Semantics of Questions and the Pragmatics of Answers’. Unpublished PhD thesis, University of Amsterdam.
Lewis, David. 1979. ‘Scorekeeping in a language game’. *Journal of Philosophical Logic* 8: 339–359. https://doi.org/10.1007/BF00258436.
Murray, Sarah E. 2014. ‘Varieties of update’. *Semantics and Pragmatics* 7 (2): 1–53. https://doi.org/10.3765/sp.7.2.
Obenauer, Hans-Georg. 2004. ‘Nonstandard *wh*-questions and alternative checkers in Pagotto’. In *Syntax and Semantics of the Left Periphery: Interface Explorations 9*, edited by Horst Lohnstein and Susanne Trissler, 343–384. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110912111.343.
Olmastroni, Francesco, Valentina Bianchi, and Alison Duguid. 2021. ‘A deliberative bridge over the mass-elite rift: Effects of online deliberation on support for European integration’. In *The EU through multiple crises*, edited by Maurizio Cotta and Pierangelo Isernia, 186–220. New York: Routledge.
Potts, Christopher. 2007. ‘The expressive dimension’. *Theoretical Linguistics* 33 (2): 165–198. http://dx.doi.org/10.1515/tl.2007.011.
Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.
Roberts, Craige. 2018. ‘Speech acts in discourse context’. In Fogal, Harris, and Moss, *New Work on Speech Acts*, 317–359. https://doi.org/10.1093/oso/9780198738831.003.0012.
Roberts, Craige. 2012. ‘Information structure: towards an integrated formal theory of pragmatics’. *Semantics and Pragmatics* 5 (6): 1–69. https://doi.org/10.3765/sp.5.6. Originally published in 1996 in ‘Papers in

Semantics', edited by Jae-Hak Yoon and Andreas Kathol, *Working Papers in Linguistics* 49: 91-136. Columbus, OH: Ohio State University, Department of Linguistics.

Stalnaker, Robert. 2018. 'Dynamic pragmatics, static semantics'. In Fogal, Harris, and Moss, *New Work on Speech Acts*, 394–399. https://doi.org/10.1093/oso/9780198738831.003.0014.

Stalnaker, Robert. 2002. 'Common ground'. *Linguistics and Philosophy* 25: 701–721. https://doi.org/10.1023/A:1020867916902.

Stalnaker, Robert. 1999. 'Assertion'. In *Context and Content: Essays on Intentionality in Speech and Thought*, Chapter 4, 78–95. Oxford: Oxford University Press. Originally published in 1978 in *Syntax and Semantics 9: Pragmatics*, edited by Peter Cole, 315–332. New York: Academic Press.

Index

Page numbers with n refer to footnotes.

A

accountability principle 166
activated meanings 123
ad hominem arguments 58
addressee competence, assumed 325–327
addressees, effect of insults on 67, 73–78
  see also targets of hate speech
adjectives
  annotation of corpora 172
  apparently neutral terms 283, 289
  implicit meanings 254–255
  indefinite adjectives/adverbs 267–268
  insults 83
  intensification/mitigation 267–268
  irony 95
  nationality adjectives 283, 289, 294, 295–298, 302
  out-group construction 14, 262
  stereotyping 259
adverbials 97–98, 238, 252–255, 262–263, 272
adverbs 14, 252, 267–268, 270, 294, 298–303
adversatives 247, 248
affectivity 137, 179–181
  see also emotions
affiliations, group 37, 238, 305–306, 316
  see also belonging; in-groups/out-groups; insider/outsider identities
affordances 2, 271, 290, 316, 323n10, 329n15
aggressive language
  affordances of virtual environment 290
  chorus effects 324
  co-occurrence with offensiveness 113
  COVID-19 anti-Chinese hate 258
  definition of hate speech not limited to 6
  distinguished from hate speech 7
  in modern culture of communication 3
  tagset design 118
  TWEER corpus 120, 121
AI 289
  see also algorithms
Alfonzetti, Giovanna 69
algorithms 235, 272, 289, 293
ambiguity 254–255
ambiguity account of appropriation 43n14
anaphor 87

Anderson, Luvell 42n10, 43n14, 44n16, 46n19
anecdotes 267
animal metaphors 82–83, 90–91, 131
annotation of corpora 113–114, 116–119, 122, 145, 164n2, 171, 172–178, 188, 244
annotation of stylistic aspects of language 161
anonymity 60, 64, 290, 316
AntConc 124
anthropological homophobia 131
anti-Chinese discourse 233–279
anti-immigrant discourse 233–279
anti-politeness 66n8
Antonelli, Giuseppe 58, 59, 60, 63
API (Application Programming Interface) on Twitter/X 161n1, 165, 168, 176
apophasis 270
appropriation of slurs
 ambiguity account of appropriation 43n14
 bilingualism's impact on 193–230
 connotation 206
 context-dependency of insults 17
 contextual clues of offensiveness in homophobic hate speech 206
 derogative speech 205–206
 discrimination 205
 echoic account of appropriation 43n14, 205
 expressivity 43n14, 205
 to fight discrimination 44
 in-groups 205
 identity and hate speech 29, 42–45, 149
 indexical account of appropriation 43n14
 intimacy 43
 legitimating environments 9
 non-derogatory uses of slurs 42–45, 47–48
 political appropriation 44
 pragmatics 195, 205–206
 solidarity 43
 symbols 45–46
 uses of slurs between friends 43, 44
archaic terms 168
argumentation
 homotransphobia in Twitter corpus 128, 145
 limited on Twitter/X 61
 as lost art 58
 and metaphor 90
 non-cooperative contexts 333, 334, 342
 pathos precedes 59
 political discourse 58–61, 235–236, 241, 243, 256–271
Aristotle 59
'asides' 14
assault-like speech acts 36, 38
assimilation strategies 128–129, 142
asterisks 287
attention-grabbing words/arousal 214, 216, 222
attitudinal data 170
attributive structures 83
auctoritas 70
audiences 37, 316
 see also bystanders; targets of hate speech
Austin, John L. 17–18, 29, 33, 34, 35, 36n3
author profiling 112
authoritative subordinating speech acts 36
authority, construction of 13–14
automatic detection systems
 see detection systems
automatic filters 64, 89, 103, 134
automatic user detection 112
axiology 70, 80, 82, 103

B

Baider, Fabienne 15
banter 44

behaviour, attempts to police 39
belonging 66, 238, 299n19, 305–306
see also in-groups/out-groups
Berlinguer, Bianca 127
Berlusconi, Silvio 58, 91n19
Bianchi, Claudia 6, 32n1, 43n14, 44n16, 68, 102, 205, 206, 286, 289
biased mental representations 239
bilingualism 193–230
Bilingualism and Emotion Questionnaire (BEQ) 197, 198
biphobia 150
Biri, Ylva 271
'bitch' 46n19
'Black' 43
blasphemy 64, 74
Boeckmann, Robert J. 204, 208, 224
Boldrini, Laura 38n5
Borrillo, Daniel Angel 138
Bossi, Umberto 58
boundary terms 142–143
Brambilla, Marina 287
Brown, Alexander 31, 202, 203
'but' 12–13
bystanders 33, 38, 68–69, 203

C

Calvert, Clay 203–204
Camp, Liz 47
Canobbio, Sabina 70, 102
Carey, James W. 203
Carnaghi, Andrea 33
Cepollaro, Bianca 32n1, 42
chains 307, 309
chorus effects 323–324
Cirinnà, Monica 127
classifications/categorisations of people 30–33, 35, 40, 241, 256–264
see also hierarchical rankings
clinical homophobia paradigm 132–133
co-construction of meaning 235, 237
Code of Conduct on Countering Illegal Hate Speech Online (European Commission 2016) 285, 287–288
code-mixing 291
collocations 123, 147, 148
comedy 16
comment threads 256–257, 260
common ground/shared beliefs 238, 258, 260, 318–323, 326–327
common knowledge/references 89, 102, 104
comparatives 14
compensatory strategies 187
complicity, feelings of 47
computational linguistics 112, 114, 115–122, 165
computer-mediated communication 59, 114, 145
conceptual metaphors 90
conditional inference trees 177–188
connectives 12–13, 238, 246–252, 272, 292
connotation
apparently neutral terms 167
degree of offensiveness 186
emotions 179–181
in-groups 17
homophobic hate speech 126–131, 169
L1 versus L2 199
political discourse 64, 70, 80, 82–87, 103–104
slur appropriation 206
stylistics 187
TWEER corpus 141
conspiracy theories 268
constative language 17
Conte, Guiseppe 95, 96, 100, 101
contempt 6, 31, 43, 45n17, 67, 69, 289, 298, 302, 305
'context sets' 318, 319
context-dependency

boundary between neutral and hate content 308
context stability 320–321
contextual clues of offensiveness in homophobic hate speech 159–191, 206
cultural traits in nationality-based hate speech 305
defining hate speech 8–9
denigration 289
detection systems 289
discussion threads 328
illocutionary and perlocutionary force 287
implicit versus explicit hate speech 14–15, 163
insults and derogatory terms 17, 69, 70
natural language processing (NLP) 163
unstable contexts 321–323
continuous discussion threads 317, 321–323
continuum/spectrum, hate speech as 6, 115, 237, 289
controversial topics 328–329
conventional implicatures 238
conversational implicatures 238
co-occurrences 123, 147, 148, 162
see also collocations
cooperative interactions 71, 237, 238, 242–243, 316–341
coordinating conjunctions 248–249
corpus notable absences 138
Corpus of Contemporary American English (COCA) 123
correlative connectives 250
Costa, Albert 200, 222
Council of Europe 5, 284, 285
covert prestige 68
COVID-19 233–279
credibility 67, 94, 254–255, 270
Credtani, Valentina 287
Criado Perez, Caroline 38n5
crimes
crimes against humanity 40
hate crimes 144, 202, 235, 258
inciting crime 40
insults as 66
legislative definitions of hate speech 5, 8, 284
Critical Discourse Analysis (CDA) 113–114, 234, 236, 240
Croom, Adam 33n2, 45n18, 47
crosslinguistic transfer 222
crosslinguistic word pair tasks 208–209
cruelty, rationalisation of 40
culture
cultural referents in political discourse 100–102, 104
cultural traits in nationality-based hate speech 303–305
culture-specific insults 79, 82, 90, 100–104
culture-specific speech acts 199
and nationality 293–294
offensive expressions 298
political discourse in COVID-19 265–266
curses 69
cynicism 305, 308

D

DDL (*disegno di legge*) bill 331n18
De Luca, Vincenzo 243–244, 245
deep learning tools 112
defamation 7, 69n10, 73, 74, 77, 78
defining hate speech 4–9, 31, 162–163, 202–203, 236–237, 273, 283–287, 329
degrading speech acts 36
degree adverbs 267–268
dehumanisation
effect on targets 204
homotransphobia 117–118, 128, 131–132, 151

political discourse in COVID-19 241, 260, 265–266, 272–273
deixis 128, 176, 179, 181, 186–187
delegitimisation of opponents 63, 67, 117–118
deliberative questions 317, 326, 330–342
democracy 235
demographic information 176
demonstratives 13, 176, 267
denigration 65, 289
derivational morphology 176, 261, 307
derogatory speech
adverbials 252–255
apophasis 270
automatic detection systems 289
changing derogatory force over time 32–33
connectives 246–252
context-dependency 169, 181
defining 237
demonstratives 13
derogatory epithets/slurs and identity construction 29–49
emojis 330
epithets as insulting strategies 79
gender of Twitter/X users 187
implicit meanings 237
as indicator of hate speech 294
inherently derogative words 185
intensification/mitigation 267
lexical fields 307
slur appropriation 205–206
descriptive versus performative functions of language 286
detection systems
annotation of corpora 145
apparently neutral terms 283, 308
context-dependency 289
deep learning tools 112
derogatory speech 289
false positives 147
implict hate speech 272, 273, 287
irony and sarcasm 113, 272, 285
lack of international agreement on definition of hate speech 288–289
lexical fields 307
metaphors 269
nationality-based hate speech 293
natural language processing (NLP) 162, 187, 203
offensiveness 113
pragmatics 147
and rapid language change 288–289
target detection 112
user reports as detection system 289
Dewaele, Jean-Marc 198–199
Di Maio, Luigi 95, 96
Di Maio, Marco 73, 81, 88
dialectal words 175, 181, 183
dichotomies 97, 128, 143, 247–248, 273
dictionaries 140, 166
disabled people 82
disagreement 324–325
disambiguation 146, 172–173
discourse analysis 234, 236, 240–243, 256–271, 272–273
discourse frames 240
discourse strategies 240–243, 256–271, 272–273, 292
discourse-historical approach (DHA) 236, 240, 241
discrediting opponents 60, 61, 62, 67, 80
discrimination
in definitions of hate speech 7
incitement to 118
insults 66, 68
propaganda-like speech acts 37
racism 33–34, 39–40, 79, 129, 147, 233–279, 281–313
sexism 34, 38–39, 127
slur appropriation 205

slurs as constitutive of 34, 41
TWEER corpus 120
using appropriation to fight 44
victim/perpetrator reversals 128
discursive structures 89–92, 235–236
discussion threads 317, 321–323, 324, 327, 331–342
disgust words 67, 98, 202, 307, 331
dislocations 175–176
'distance' conceptual space 142
distancing from the target group 33, 299, 302–303, 306
dog-whistles 17
Domaneschi, Filippo 65, 66, 67, 68, 70, 74, 80, 82, 91n19, 102
double entendres 267
dynamic pragmatics 316, 318–320

E

echo chambers 235
echoic account of appropriation 43n14, 205
Eisenstein, Jacob 165
election campaign periods 61
ellipsis 98
embodied insults 224
emojis 95, 330
emotions
affordances of virtual environment 316
bilingualism 195, 197–202, 206–207, 224
chorus effects 324
defining hate speech 202–205, 236
discourse strategies 241
dislocations 176
insults 66–67, 92, 94
and lexical variation 179–181
listeners' 203
in political discourse 59–60, 62, 63
speech acts 199
syntactic variation indicating 163–164
emphatic expression 175, 299, 327
epistemic states 242, 325
epithets 78–89, 307, 330
see also insults; slurs
errors 291–292
essentialism condition 40
ethical values 135
ethnophaulisms 200
Eugenides, Jeffrey 28
euphemism 134
European Commission against Racism and Intolerance 284, 285
European Court of Human Rights 117
European Union 284
evaluative adverbs/adverbials 14, 252
evidentials 14
excessive language 324, 328
exercitives 36n3
explicature 268n7
expressivity 43n14, 69, 171, 179, 205, 323

F

Facebook
Code of Conduct on Countering Illegal Hate Speech Online (European Commission 2016) 285
community standards 287–288
constructing a corpus from 290–294
discussion threads and non-cooperative contexts 317, 328–329, 331–333
immigration 336–340
nationality-based hate speech 281–313
political discourse 62, 243, 244, 250–251, 258
virtual communities 290
face-threatening acts 67, 71
fallacies 242

Faloppa, Federico 162–163, 265, 269
false claims 94
false identities 64
false positives 147
familiarity as variable in bilingual study 214–218, 221–226
fantasy fiction 42–43
fare + noun 147
Farkas, Donka 325, 327
Ferrari, Angela 246
Feuerbach, Ludwig 100
fiction, slurs in 42
fields of action 240
figures of speech 267, 269
see also metaphors
filler words 121
filters 64, 89, 103, 116, 134
first language attrition 201, 206–207, 222
first-person constructions 74–76, 255, 265–266, 267
flaming 60, 64n6
focality 239
focus/topic structures 239, 252
Fontana, Lorenzo 126
formal register 58
Fortuna, Paula 285, 289
Forza Italia 71, 72
foul language 141–142
see also swearing; vulgar words
Framework Decision 2008/913/JHA (EU) 284, 285
Fratelli d'Italia 62, 72, 73, 92, 93, 94
freedom of expression 3–4, 66, 236, 308
frequency wordlists 124
friends, uses of slurs between 44
Fumagalli, Corrado 13
functional variation features 40

G

Galinsky, Adam D. 45
Gawinkowska, Marta 200, 201, 222
Gelber, Katharine 204, 224
gender identity 111–157, 317, 333–336
gender non-conformity 45
gender norms 38–39
gender of Twitter/X users 176, 181, 187
generalisations 4, 128, 183, 241–242, 260, 266, 306
generic pronouns 13
genocide 40, 133, 260
genres 240
GIFs 330
given information 248, 250–251
see also common ground/shared beliefs
grammar
see morphosyntax; syntax
graphic forms 287
Greenberg, Jeff 33
Grice, Paul 16, 237, 238, 316
Grillo, Beppe 58, 79n17
Grondelaers, Stefan 165, 187, 188
Grosjean, François 201
group affiliation, indicating 37, 238, 305–306, 316
see also belonging; in-groups/out-groups
Gulpease index 59
Gümüşay, Kübra 140–141

H

hard and soft hate speech 15–16, 164
harm
bystanders 33, 68–69
in definitions of hate speech 286
emotional harm 203–204, 207, 223–224, 226
to identity 204, 207
intangible harms 203–204
to out-group 151
psychological damage 33, 69, 204, 207, 223–224, 226

psychological distance 66, 200
speech acts 40
Harris, Kamala 38n5, 91
hate campaigns 38n5
hate crimes 144, 202, 235, 258
HateBERT 112
Herbert, Cassie 46n19, 48, 48nn20–21
heteronormative paradigm 135, 148, 151
hierarchical rankings 11, 35–36, 40, 46–47, 128, 138, 148, 286
Hom, Chris 32, 43, 44, 205
homonormativity 148–149
homonymy 172
homophobia 33, 34, 79, 111–157, 159–191
homosexuality 41n9, 111–157, 159–191, 328–329, 331–333
homotransphobia 111–157
Hornsby, Jennifer 45n17, 205
humanity, essential 40n8
humour 10n3, 285
hyperboles 92–100, 242
hypotext identification 100

I

identification (versus definition) of hate speech 8–9
see also detection systems
identity
anonymity 60, 64, 290, 316
belonging 66, 238, 299n19, 305–306
bilingualism 196, 198
construction of 29–49
'deviant' 141
false identities 64
harm from hate speech to 207
hate speech and identity 27–54
migrants 129
performative functions of language 18
reduction to single trait 256
signalling identity on Twitter 165
silencing effects of hate speech on 204
see also in-groups/out-groups
ideologies 246
idiomatic language 146
illocutionary dimension of speech acts 10n3, 34, 35, 36, 66, 242, 287
imageability 214, 216, 218, 222
implicatures 10n3, 11, 235, 237–239, 249–250, 251–252
implicit meanings
adverbials 252–255
connectives 246–252
connotation 246–252
detection systems 272, 273, 287
homophobia 120, 163
homotransphobia 148
implicit discrimination 117
implicit versus explicit hate speech 9–16, 164, 287
political discourse 233–279
inciting crime 40
inclusive pronouns 14
in-concept versus out-of-concept tokens, selecting 172–173
indefinite adjectives/adverbs 267–268
indexical account of appropriation 43n14
indexicality of tweets 165
indirect strategies 42–45, 186–187
see also implicit meanings
inference trees 177–188
inferences 146, 238, 319n6
inflected and derived forms 123–124, 149, 176, 261, 307
information state 318
information-oriented conversations 323–324
information-seeking questions 317, 325, 327, 330–342

informative propositions 318
in-groups/out-groups
  argumentative strategies 128
  belonging 66, 238, 299n19, 305–306
  derogatory meanings 16
  dichotomies 143
  discursive and argumentative structures 256–271
  extension of slur appropriation 205–206
  Facebook 294
  in-group boundaries 37, 38–41, 44, 46–47, 68
  non-cooperative contexts 323
  out-group construction 39, 128, 151, 206, 256–271, 273, 324
  perspectivation 241–242, 265–271
  polarisations 246
  pronouns 299n19
injurious insults 73–74, 77, 78
insider/outsider identities 37, 39, 40, 44, 46–47, 68
Instagram 62
insults
  aimed at actions rather than being of a person/group 32
  definition of hate speech 329
  embodied insults 224
  indirect uses of 42n11
  injurious insults 73–74, 77, 78
  insults in political debates on Twitter 57–109
  not necessarily hate speech 11, 64
  out-group construction 257
  responding with an insult 210
  speech acts 65–71, 102, 294n15
  see also derogatory speech
intangible harms 203–204
intederminacy 268
intensity of hate speech
  changing derogatory force over time 32–33
  continuum of 6
  explicit hate speech 10
  implicit versus explicit hate speech 14, 273
  intensification strategies 242, 256, 265–271, 273, 324
  tagset design 117
intentions
  see speaker intentions
inter-annotator agreement 145, 174n5, 188
interdiscursivity 258, 292
International Lesbian, Gay, Bisexual, Trans and Intersex Association (ILGA) 41n9
intersectionality 7
intertextuality 61
intonation 239n1
invented language 42, 88n18
  see also new word formation
inverted comma 264
irony
  annotation of corpora 174–175
  comedy 16
  contextual clues of offensiveness in homophobic hate speech 179, 181
  detection systems cannot pick up 113, 272, 285
  implicit meanings 13, 287, 305
  as marked lexicalisation 181
  nationality-based hate speech 308
  offensive words 121
  political discourse 94–97, 99, 113
  rhetorical insults 89
  tagset design 119
  TWEER corpus 121
irrealis 330
Italiani a Cracovia [Italians in Krakow] 291
Italiani in Polonia [Italians in Poland] 291
iterative adverbials 253

J

jamovi 212
Jeshion, Robin 31, 36, 43n14, 44n16
Johnson, Mark 90

K

Kennedy, Randall 46n19
keyness 123, 147
keyword-based approaches to corpus studies 116
Kim, Sun He Ok 201
kinetic elements 69n10, 70
Knoblock, Natalia 14n4
Köpke, Barbara 222
Krämer, Nicole C. 121
Kukla, Quill (Rebecca) 36, 37, 39, 46n19, 48, 48nn20–21

L

labelling tags 87
Labov, William 68
Lakoff, George 90
Langton, Rae 34, 35, 36n3, 37
language variation and change studies 187–188
large language models 112
lavender linguistics 114, 123
left-wing politics 92–93, 97–98, 126, 127
Lega 62n5, 95, 126, 253
legislative definitions of hate speech 5, 8, 284
length of residence (LoR) in L2 country 196–197, 216–217, 222, 223, 225, 226
Lepore, Ernie 42n10
Lewis, David 318
LGBTQIA+ framing 123
Liew, Jeffrey 204, 208, 224
linguistic creativity 88n18, 92
  see also new word formation
litotes 242
loaded words 16
loanwords 178, 181, 183
Locatelli, Davide 114, 121, 132, 137, 147
locutionary acts 34
  see also illocutionary dimension of speech acts; perlocutionary effects
Lombardi Vallauri, Edoardo 14, 59n1, 237, 238, 239, 254

M

M5S 71, 72, 79n17, 95
MacKinnon, Catherine 35
macro-topics 240
Maienborn, Claudia 252
malapropisms 127n8
Marzo, Stefania 165, 188
Mauro, Ezio 127
McNamara, Luke 204, 224
meaning clusters 125
mediation on social networks 63
Meloni, Giorgia 60, 73, 75–76, 84, 87, 91, 93, 94, 97, 98, 243–244, 245, 249, 254, 269–270
metadata 245, 291
metaphors
  animal metaphors 82–83, 90–91, 131
  conceptual metaphors 90
  COVID-19 anti-Chinese and anti-immigrant hate 272–273
  intensification/mitigation 268–269
  political discourse 89–92, 99, 241, 263–264, 272–273
metonymy 294, 295, 298
microlinguistic analysis studies 187–188
Middlesex (Eugenides) 28
migration/immigrants
  anti-immigrant discourse 233–279
  assimilation strategy 128–129
  degree of offensiveness 14
  epithets/slurs 32, 80

non-cooperative contexts 329, 336–340
policies regarding 97–98
political discourse 7
political discourse in COVID-19 233–279
stereotypes 13
see also nationality-based hate speech
Miller-Idriss, Cynthia 235
mirroring paradigm 58
mitigation strategies 119–121, 241, 256, 258, 265–271, 273, 324
mockery 60
modality 14, 242, 330
moderation of online hate 272
monotonic updating 320
moral status 70
morphosyntax
annotation of corpora 175–176
derivational morphology 123–124, 149, 176, 261, 307
inflected and derived forms 123–124, 149
typical features of online texts 292
word formation strategies 14n4
word stemming 124, 168
see also nouns; pronouns; verbs
Morra, Nicola 100, 101
Motschenbacher, Heiko 123, 125, 126, 137, 138
multidata studies 160
multimodality 76n14, 114, 162
Murgia, Michela 38n5
Musk, Elon 62n4
mutual acknowledgement 322

N

naming 29
nationality adjectives 283, 289, 294, 295–298, 302
nationality-based hate speech 281–313
natural language processing (NLP) 10, 112, 113, 114, 162, 163, 203
nature-based semantic cluster 131–134
near-synonyms 161, 165–169, 170–188
negated assertions 80
negative connotation
apparently neutral terms 167
degree of offensiveness 186
emotions 179–181
political discourse 64, 70, 80, 82–87, 103–104
TWEER corpus 141
negative evaluation 33, 68, 69, 70, 294
neutralisation of slurs 45
neutrality, apparent
contextual clues of offensiveness in homophobic hate speech 179, 181
degree of offensiveness 186
dog-whistles 17
evolution of derogatory force 32
heteronormative paradigm 148
implicit meanings 164, 167, 170–171
nationality adjectives 283, 289
starting a chain of aggression with 309
new word formation 14n4, 88n18, 272, 292
nicknames 91n19
'nigga' 43, 46n19
'nigger' 47
Nitti, Paolo 102
non-canonical questions 325–326
non-cooperative contexts 315–346
non-derogatory uses of slurs 29, 42–45
see also appropriation of slurs
non-generic common ground 320
non-hate speech use of potentially hateful words 202–203

non-literal meanings 12
non-mediated fields 62
non-negotiable alternatives 327
non-targeted addressees 29
non-verbal elements 206
normalisation of hate speech 10, 234, 273
normativity of language 31, 45n18
nouns
annotation of corpora 172
homotransphobia 116
insults 82–83, 103
nominalisation 83
noun phrases as insults 70
numbers, improper use of 269–270
Nunberg, Geoffrey 32, 47
Nunes, Sérgio 285, 289

O

object pronouns 261
obligation modals 14
Octoparse 244
offensiveness
annotation of corpora 174, 177
attention-grabbing words/arousal 216
bilingual study 214–216
contextual clues of offensiveness in homophobic hate speech 159–191
degree of offensiveness 14–16, 42n11, 161, 163, 170–188, 214–216
detection of hate speech 113
feature of social networks 121
implicit meanings 163
L1 versus L2 200
perception tests 170–171
reporting offensive content 288
speaker intentions 69, 102
subversion of offensiveness 45
taboos 214
target's assessment of 67, 118
TWEER corpus 120, 139–140
variable offensiveness of words 113, 116, 118
omissions in corpus data 138
online disinhibition effect 121
online perceptual tests 170
opacity 242–243, 254–255
open issue questions 325, 326–327, 330–342
open versus latent hate speech 287
see also implicit meanings
opinion mining 144, 146
opposite meanings 12
oppression, networks of 41
orthography 172, 175, 188
'otherness' 135, 257
out-group construction 39, 128, 151, 206, 256–271, 273, 324
see also in-groups/out-groups
out-of-concept tokens, removing 172–173

P

pairwise comparisons 328–329
Palermo, Massimo 60, 68, 74, 100, 102
Pamungkas, Endang Wahyu 119
par condicio law 62
paralinguistics 69n10, 70, 206
parallelisms 89, 97–100, 249, 250, 262
parentheticals 14
participant recruitment 210–211
partition semantics 319, 325, 326
Partito Democratico 71–72, 95
pathos 59–60
Pavlenko, Aneta 197, 199
pedagogical context of slurs 43
pejoratives 31–32, 88n18, 261
perception tests 170–171
performative functions of language 9, 17, 18, 29, 33–37, 40, 48, 286

periphrasis 298
perlocutionary effects 17–18, 34, 67–68, 69, 102–103, 287
persecutive speech acts 36
personal pronouns
  first-person pronouns 74–76
  in-groups/out-groups 261
  insults 257
  nationality-based hate speech 294, 298–303
  political discourse 74–77
  reduced use of 292
  stereotypes 13
perspectivation 241–242, 265–271
persuasive devices 3, 15, 59–60, 237, 247
philosophy of language 6, 78, 113, 151, 223, 283–284, 286
phonetic realisations 42n10
Pillon, Simone 331
pink triangle 45–46
Pirandello, Luigi 100
Pistolesi, Elena 64n6, 94
place adverbs 252, 298–303
Poletto, Fabio 174
politeness 44, 66n8, 71, 74
political appropriation 44
political discourse 7, 57–109, 62–64, 126–131, 233–279
political slogans 11
politically correct speech 16
polyphony 89
populism 59–60, 81, 235
pornography 35
possessive pronouns 75–76
possessives 255
possible world semantics 318
poststructuralism 146
Potts, Christopher 43n14, 205
power imbalances 34, 40, 41, 48, 70
practice, language as social 234–235
pragmatics
  adverbials 252–255, 272
  automatic detection systems 147
  dislocations 176
  dynamic pragmatics 316, 318–320
  implicit meanings 235, 236, 237–239, 241, 272
  insults 65, 66, 71, 102–103
  linguistic triggers of hate speech 246–256
  pragmatic inferences 13
  question typologies 325
  slur appropriation 195, 205–206
  speaker intentions 16
  see also speech acts
predicative functions 67
predictor variables 177–178, 179, 185
prefixes 88n18
prejudice
  categorisation 241
  in definitions of hate speech 7, 202–203
  hierarchy of sexuality 138, 148
  implicit meanings 11
  insults 68
  labels 30
  out-group construction 256
  pronouns 299n19
  stereotypes 13
prepositional phrases 252, 253–255
prepositions 98
prestige 68, 83
presumptions, unwarranted 11
presuppositions 10n3, 11, 94, 102, 235, 237–239, 253, 303, 317
preterition 271
principle of accountability 166
problem-oriented approaches in corpus studies 124–125, 138
proficiency in language 291
  see also bilingualism
prohibition modals 14
prohibition of language 42n10
pronouns
  annotation of corpora 176
  generic pronouns 13

inclusive pronouns 14
object pronouns 261
possessive pronouns 75–76
pronominal contrast 282, 294, 299n19, 302, 305
subject pronouns 299, 302
see also personal pronouns
proof-of-concept studies 328–329
propaganda 234, 238, 243
propaganda-like speech acts 36–37, 38, 68
propositions 318, 321
'protected characteristics' 288
prototype theory 69, 102
proxemic elements 69n10, 70
psycholinguistics 223
psychological damage 33, 69, 204, 207, 223–224, 226
psychological distance 66, 200
punctuation 175, 292
Pyszczynski, Tom 33

Q

Qualtrics XM platform 207
quantifiers 269–270
quantitative research 84–86, 114–122, 160–161, 164, 178–187, 310
'queer' 43, 44, 45, 205, 206
queer linguistics 114, 123, 148
question types 315–346
questionnaire methods 207–208, 225
questions under discussion (QUDs) 319, 321, 323, 330–342

R

racism 33–34, 39–40, 79, 129, 147, 233–279, 281–313
Rackete, Carola 95
Raggi, Virginia 127
Rahman, Jacquelyn 46n19
Ramírez-Esparza, Nairán 197
reality and language/words 29, 240, 269, 273
reappropriation of slurs
see appropriation of slurs
Reddit 317, 329, 333–336
reference strategies 123
referential functions of insults 66–67
regional variation 140, 166, 167, 168, 175, 179, 183
register 58, 140
religious semantic sphere 136–137
Renzi, Matteo 73
reported speech 42n11, 146, 242
reporting offensive content 288
reporting versus using slurs 42
reputational damage 77
research field, hate speech as 2, 3, 8, 31, 112
researcher subjectivity 310
response variables 177–187
rhetorical discourse 58, 92
rhetorical insults 89–102
'rhetorical number game' 269–270
rhetorical questions 326–327
Richard, Mark 43n14
Richetti, Matteo 73
right-wing rhetoric 80, 92, 126, 235, 248, 253–254, 263–264, 269, 272
Ritchie, Katherine 43n14
Rösner, Leonie 121
Rwandan genocide 40, 133, 260

S

salience 123
Salvini, Matteo 60, 73, 75, 77, 81, 86, 91n19, 95, 96, 97, 98, 99, 100, 126, 243–244, 245, 247, 248, 249, 250, 251, 253, 254, 255, 257, 258, 260, 262, 263–264, 268, 269, 270
Sanguinetti, Manuela 116–117, 119, 174n5

Santanchè, Daniela 73
sarcasm 8, 10n3, 174, 285, 287
Sbisà, Marina 238
scalar implicatures 12
Schäfer, Martin 252
Schmid, Monika 222
Schmidt, Anna 162, 165
second language acquisition 199–200
second person constructions 36, 74–75, 302–303, 306
Seglow, Jonathan 151
self-addressed questions 326, 330–342
self-application of slurs 45
self-empowering groups 48
self-praise 61–62
self-respect 151
self-worth 31
semantic families 123–124, 125, 126, 130
semantic frames 123
semantic spheres 123, 143, 164
sentic computing sources 203
sentiment analysis 114, 144–150
sexism 34, 38–39, 127
short texts 61
Sibomana, Andre 40n8
silencing effects of hate speech 204
silentism 42n10
Slagle, Mark 223
slur appropriation
  see appropriation of slurs
slurs
  action-engendering slurs 40
  aimed at entire groups 31–32, 78
  avoiding 11
  construction of identity 29–49
  contempt 31
  definition of hate speech 329
  evolution of derogatory force 32
  homotransphobia 149
  and identity 28
  impact of bilingualism on perception 193–230
  implicit versus explicit hate speech 14
  as indicator of hate speech 294, 329
  insults 78–79
  lexical fields 307
  linguistic behaviours 31
  neutralisation of 45
  non-derogatory uses of 29, 42–45
  performative functions of language 29, 33–37, 40, 48, 286
  in political discourse 80, 84–85
  social control functions 38–41
  and social labels 30–33
  TWEER corpus 139–140
'slut' 38, 39
social identity 123, 306n21, 309
social labels 30–33
social networks, and political discourse 58–65
sociocultural contexts 45–46, 114, 148, 199
sociology 66, 74
sociopragmatic norms 199
soft hate speech 15–16, 164
solidarity 43
song lyrics 42
speaker intentions
  and effects 16–18
  explicit hate speech 10
  implicatures 238
  implicit meanings 8
  insults 67n9
  irony 94
  performative functions of language 17
  perlocutive effects 34
  speech acts 10
  transitivity 143
spectrum/continuum, hate speech as 6, 115, 237, 289

speech acts
assault-like speech acts 36, 38
authoritative subordinating speech acts 36
cooperative interactions 317
culture-specificity 199
degrading speech acts 36
emotions 199
harm 40
illocutionary dimension of speech acts 10n3, 34, 35, 36, 66, 242, 287
implicit hate speech 11
insults 65–71, 102, 294n15
non-cooperative contexts 323
performative functions of language 17
persecutive speech acts 36
propaganda-like speech acts 36–37, 38, 68
slurs as 29, 33–37
speaker intentions 10
speed of response 2, 292, 316
spelling 124, 165, 172, 175, 188, 272
Stalnaker, Robert 316–317, 318
Starks, Donna 201
statistics, improper use of 269–270
stereotypes
annotation of corpora 174
anti-immigrant discourse 260–261
categorisation 241
COVID-19 anti-Chinese and anti-immigrant hate 259–260, 266–267, 272–273
in definitions of hate speech 7
Facebook community standards 288
implicit versus explicit hate speech 11, 12, 13
insults 68
labels 30
media perpetuation of 204
and nationality 293n12
nationality-based hate speech 298
personal pronouns 13
pronouns 299n19
semantic spheres 123
stylistics 186–187
tagset design 119
TWEER corpus 121, 122, 144
Storace, Francesco 126
structural presuppositions 238
structurally non-cooperative contexts 323, 324–325
Stuart-Smith, Jane 187
stylistics 161, 166–167, 186–187
subject pronouns 299, 302
subjunctives 143
subordination 13–14, 35–36, 37, 41, 46–47
suffixes 88n18
Sulpizio, Simone 32n1
swearing 163, 200, 329, 334
see also vulgar words
symbols 45–46, 287
synecdoche 241
synonyms 166
syntax
annotation of corpora 175–176
construction of stereotypes 13
deviations indicating emotional content 163–164
insults 70, 83
nationality-based hate speech 298
parallelism 97–100
see also morphosyntax

T

taboos 199, 208, 212, 214, 216, 222
tacit acceptance of propositions 321, 322
tag questions 242
tagset design 116–119
targets of hate speech
in definitions 5–6
distancing from the target group 33, 299, 302–303, 306

female left-wingers 127
in-group members 264
individuals versus groups 31–32
individuals versus institutions 7
insults 67n9
minority groups 7
non-targeted addressees 29
offensiveness judgements 67, 118
policing of target groups 39
'protected characteristics' 288
public persons as representatives of out-group 324
role of ethnicity in responding to hate speech 219–221
role of target's sexuality in responding to homophobic slurs 218–221, 223
tagset design 117
target detection 112
target-dominant group gap 33
victim/perpetrator reversals 128
see also bystanders; in-groups/out-groups
temporal connectives 251–252
Testa, Annamaria 60, 62
textual cohesion 292
"them" and "us"
see in-groups/out-groups; insider/outsider identities; "us" and "them"
Theory of Mind 206
third-person constructions 37, 74–75, 301, 306
threatening language 38, 69, 332
Tirrell, Lynne 38–39, 40, 41, 45
Titscher, Stefan 241
Toninelli, Danilo 92
topic effects 328, 341–342
topicalizations 235, 237–239, 248, 250–252
topic/focus structures 239, 253
topoi 242–243, 257, 258, 260, 262, 272–273
transitivity 143
translations to English 79, 245, 286n4
Treccani dictionary 166, 167, 298
trivialisation of evil 134
trolling 64
Tropedia 42n12
truth claims 17, 94, 266, 318, 321
see also reality and language/words
TWEER corpus construction 115–116
Twitter (X)
Code of Conduct on Countering Illegal Hate Speech Online (European Commission 2016) 285
contextual clues of offensiveness in homophobic hate speech 159–191
homotransphobia 111–157
insults in political debates on Twitter 57–109
political discourse in COVID-19 243–244
as research tool 161, 164–165
type-token ratios 59, 122, 169

U

Una Europa 3n1
United Nations definition of hate speech 5, 8, 237
unstable contexts 321–323
"us" and "them"
distancing from the target group 33, 265–266, 299, 302–303
insider/outsider identities 40
nationality-based hate speech 128
ours/theirs 255
out-group construction 255, 257
polarisations 62n3, 143, 241, 246
pronominal contrast 294
pronouns 299
reinforcing stereotypes 68
see also in-groups/out-groups; insider/outsider identities

user reports as detection system 289
utterance as essential element 10n3

V

vagueness 242–243, 254–255, 262
valuative tags 87
Van Dijk, Teun A. 62n3, 128, 246, 269
variationist sociolinguistics 160, 165, 177
verbs
  first-person verb forms 265–266, 267
  insults 85, 98, 103
  irrealis 330
  modality 14, 242, 330
  nationality-based hate speech 298
verdictives 36n3, 66, 69
victim/perpetrator reversals 128
violence
  incitement to 40–41, 112, 117, 118, 144, 256
  insults 68, 71
  irrealis modality 330
  TWEER corpus 120, 143
  verbal abuse 36
virtual communities 290
vocatives 70
vulgar words 167–168, 282, 294, 307, 329, 332, 334

W

white replacement theory 268
Wiegand, Michael 162, 165
Wikipedia 166, 167
Wilson, Rosemary 198, 199
Wodak, Ruth 236, 240, 242, 269
women as objects of hate speech 38–39
word formation strategies 14n4, 88n18, 272, 292
word pair tasks 195, 207–208, 212–218
word stemming 124, 168
World Health Organization 133

X

X
  see Twitter (X)

Y

YouTube 112, 285

Z

Zaia, Luca 243–244, 245, 259, 265
Zan, Alessandro 331n18
Zan bill 150
Zeman, Dan 42
Zimmerman, Klaus 66n8
Zipfian distributions 169, 177, 178