UNIVERSITÀ
DI TORINO

Università
di Genova

DOCTORAL THESIS

Perspectives and Praxes for Socially Situated NLP: An Interdisciplinary Queer and Context-Aware Linguistic Approach to the Development of Linguistic Resources

---

Author:
Andrea MARRA

Supervisors:
Prof. Cristina BOSCO
Prof. Viviana PATTI
Dr. Angela ZOTTOLA

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy in the
Dipartimento di Lingue e Letterature Straniere e Culture Moderne
and Dipartimento di Informatica (Università di Torino)
and in the
Dipartimento di Lingue e Culture Moderne (Università di Genova)

Corso di Dottorato in
Digital Humanities. Tecnologie digitali, Arti, Lingue, Culture e Comunicazione
(curriculum *Linguistica, Linguistica applicata, Onomastica*)

Academic years: 2022/2023, 2023/2024, 2024/2025

Academic disciplines: L-LIN/01 and INF/01

Declaration of Authorship

I, Andrea MARRA, declare that this thesis titled, "Perspectives and Praxes for Socially Situated NLP: An Interdisciplinary Queer and Context-Aware Linguistic Approach to the Development of Linguistic Resources" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the University of Turin and the University of Genoa.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Some parts of this thesis include revised versions of the following papers:

- Marra, Andrea and Cristina Bosco (2024). "Orsù dunque... avvocato? Osservazione del maschile sovraesteso nei nomina agentis su Twitter". In: *Lingua inclusiva: forme, funzioni, atteggiamenti e percezioni*. Ed. by Anna-Maria De Cesare and Giuliana Giusti. Vol. 6. LiVVaL. Venezia: Edizioni Ca' Foscari – Venice University Press, pp. 203–248. DOI: 10.30687/978-88-6969-866-8/008. URL: http://edizionicafoscari.it/it/edizioni4/libri/978-88-6969-867-5/orsu-dunque-avvocato/

- Ferrando, Chiara, Lia Draetta, Andrea Marra, Angela Zottola, Cristina Bosco, and Viviana Patti (2025). "AEQUITAS 2025 Fairness and Bias in AI". In: *Proceedings of the 3rd Workshop on Fairness and Bias in AI co-located with the 28th European Conference on Artificial Intelligence (ECAI 2025)*. Vol. 4147. Workshop AEQUITAS 2025, October 26th, 2025. Bologna, Italy: CEUR Workshop Proceedings, pp. 1–12. URL: `https://ceur-ws.org/Vol-4147/`

- Marra, Andrea, Chiara Ferrando, Lia Draetta, Bianca Cepollaro, and Viviana Patti (2026). "How is the reclamation of slurs perceived in Italian?: A sociolinguistic survey to inform future NLP studies". In: *Linguistik online* 144.3, pp. 209–225. URL: `https://bop.unibe.ch/linguistik-online/article/view/13550`

Signed:

_____

Date:

_____

A Maria Giovanna, Elena, Enza e Monica:
da voi ho imparato
a far convivere i punti di visti,
anche, e soprattutto, se diversi.

A Carmela:
a ogni mio cambio di prospettiva
la tua voce risuona ancora,
più forte che mai.

"Y lo que cambió ayer
Tendrá que cambiar mañana
Así como cambio yo
En estas tierras lejanas"

Mercedes Sosa, *Todo cambia*

"Tutta qua la felicità sembra semplice
Sei felice o sei complice
Cosa c'è di difficile, non ti perdere
Sei felice o sei complice"

La Rappresentante di Lista, *Vita*

Abstract (English)

This thesis investigates how an interdisciplinary, queer, and context-aware approach can contribute to the study of potentially harmful language in Natural Language Processing (NLP). Drawing on the intersection of Sociolinguistics, Queer Linguistics, Philosophy of Language, and NLP, this work argues that the annotation of linguistically and socially sensitive phenomena cannot be reduced to a purely technical operation, but must rather be understood as a socially situated practice. Within this framework, the thesis introduces and tests the *FATA (First Ask Then Act)* paradigm, conceived as a methodological reorientation that integrates social reflexivity and community participation into the early stages of the research pipeline.

The research unfolds across three interconnected investigations. The first comprises a preliminary study on Italian *nomina agentis* and on the relationship between grammatical gender, reference, social meaning, and power dynamics. The second presents a case study on the discursive construction surrounding Francesco Schettino and Carola Rackete within a corpus of annotated Italian tweets, which transparently documents the iterative design of an annotation schema for gender stereotypes and their associated dimensions. The third constitutes a case study on the reclamation of LGBTQ+ slurs in Italian, developed within the *FAVOLOSA (Fair Automatisation and Visibility Of the LGBT+ Community On Slur (Re)Appropriation)* project, which integrates a preliminary focus group, a sociolinguistic survey, and a multi-layered qualitative-quantitative interpretive analysis.

Overall, the thesis shows that stereotypes and linguistic reclamation challenge the categorical simplifications often required by computational annotation. It therefore advances the proposal of more reflexive annotation schemas, sensitive to disagreement, social positionality, emotional charge, pragmatic complexity, and interpretive effort, thereby contributing to NLP research while simultaneously enriching the broader discourse within the social sciences.

Abstract (Italian)

La presente tesi indaga come un approccio interdisciplinare, queer e attento al contesto possa contribuire allo studio del linguaggio potenzialmente dannoso nell'ambito del *Natural Language Processing (NLP)*. Muovendosi all'intersezione tra Sociolinguistica, Linguistica Queer, Filosofia del Linguaggio e NLP, questo lavoro sostiene che l'annotazione di fenomeni linguisticamente e socialmente sensibili non possa essere ridotta a una mera operazione tecnica, ma debba invece essere intesa come una pratica socialmente situata. All'interno di questa cornice, la tesi introduce e mette alla prova il paradigma *FATA (First Ask Then Act)*, concepito come un riorientamento metodologico che porta la riflessività sociale e la partecipazione delle comunità nelle fasi iniziali del processo di ricerca.

Il lavoro si sviluppa attraverso tre studi tra loro connessi. Il primo è uno studio preliminare sui *nomina agentis* italiani e sul rapporto tra genere grammaticale, referenzialità, significato sociale e dinamiche di potere. Il secondo è un caso di studio sulla costruzione discorsiva che ruota attorno alle figure di Francesco Schettino e Carola Rackete in un corpus di tweet italiani annotati e che documenta in maniera trasparente la progettazione iterativa di uno schema di annotazione degli stereotipi di genere e delle dimensioni a essi correlate. Il terzo studio riguarda la riappropriazione di termini denigratori nel contesto LGBTQ+ italiano, sviluppato nell'ambito del progetto *FAVOLOSA (Fair Automatisation and Visibility Of the LGBT+ Community On Slur (Re)Appropriation)*, che integra un focus group preliminare, un questionario sociolinguistico e un'analisi interpretativa, sia di tipo qualitativo che di tipo quantitativo, multilivello.

Nel complesso, la tesi mostra come gli stereotipi e la riappropriazione linguistica mettano in discussione le semplificazioni categoriali spesso richieste dall'annotazione computazionale. Essa avanza pertanto la proposta di schemi più riflessivi, sensibili al disaccordo, alla posizionalità sociale, alla carica emotiva, alla complessità pragmatica e allo sforzo interpretativo, con l'obiettivo di contribuire alla ricerca in ambito NLP arricchendo al contempo la riflessione nelle scienze sociali.

Acknowledgements

Questo lavoro di tesi rappresenta il traguardo di un percorso denso e complesso, che non avrei mai potuto portare a termine da solo.

Il mio primo e più profondo ringraziamento va alla mia tutor, Cristina Bosco, che mi ha accompagnato fin dal primo giorno con impegno, dialogo e un'autentica e sincera apertura a proposte e prospettive diverse. E con lei un grazie sincero alle due persone che hanno seguito il mio lavoro come co-tutor, Viviana Patti e Angela Zottola, per il tempo prezioso che mi hanno dedicato, per il confronto continuo e per avermi trasmesso fiducia, guardando alle sfide del percorso con uno sguardo propositivo e ottimista.

Vorrei poi ringraziare Francesca D'Errico, prima discussant di questo lavoro, e con lei le persone che hanno sviluppato e portato avanti il progetto *STEREHOTYPE*, fondamentale fonte di ispirazione dal quale ha preso avvio questa ricerca.

Un ringraziamento particolare va a Marco Venuti e Sara Tonelli per aver dedicato tempo, competenze e attenzione alla revisione di questa tesi.

Un ringraziamento di cuore va a Martina Rosola, che mi ha accompagnato durante il mio periodo all'estero e ha rappresentato una guida attenta e un supporto fondamentale e incoraggiante per la stesura di questo lavoro. Insieme a lei ringrazio il gruppo di ricerca *LOGOS* di Barcelona per avermi accolto, offrendomi uno spazio di confronto ricco e stimolante, e le persone di *Egalia*, con le quali ho condiviso e condivido obiettivi, entusiasmo e l'impegno per un mondo più equo.

Questa tesi non avrebbe la forma che ha senza il gruppo di ricerca con cui l'ho costruita e pensata passo dopo passo: grazie a Lia Draetta, Chiara Ferrando e Marco Madeddu per aver condiviso con me la fatica e la bellezza di questo lavoro, mettendo in discussione insieme a me tante prospettive e provando a costruirne di nuove. Grazie a Bianca Cepollaro per aver contribuito ad arricchire la ricerca con le sue competenze e i suoi illuminanti punti di vista. Allo stesso modo, ringrazio Elisa Chierchiello, Soda Marem Lo, Simona Frenda, Wolfgang Sebastian Schmeisser-Nieto, Selenia Anastasi, Mara Floris e Matteo Pellegrini, che, in tanti modi diversi hanno contribuito a questo lavoro, nella condivisione di studi e progetti paralleli e negli scambi di idee che hanno allargato e arricchito la mia prospettiva.

Questo lavoro di ricerca non avrebbe senso senza le voci di chi vi ha contribuito. Per questo, un ringraziamento speciale va a chi ha lavorato con dedizione e coinvolgimento all'annotazione dei dati, a chi ha partecipato con entusiasmo al focus group e alle persone, tantissime e anonime, che hanno dedicato il loro tempo a compilare il questionario e a diffonderlo.

Il dottorato è fatto anche e soprattutto di condivisione quotidiana. Ringrazio per questo le colleghe dottorande per il supporto intellettuale, pratico ed emotivo. E con loro grazie a tutte le persone che hanno abitato, anche solo per poco, l'aula 3L: tra una pausa caffè e un aperitivo hanno reso questi anni decisamente più sostenibili e spiritosi.

Ci sono poi amicizie che si intrecciano con la ricerca. Ringrazio Sara Racca e Virginia Calabria, i cui consigli puntuali hanno nutrito e sostenuto questo percorso. E ringrazio, con tutta la mia energia, Alessandra Cignarella, amica, e per un po' anche collega, ma soprattutto amica: senza di lei, semplicemente, non stareste leggendo questa tesi.

Il mio percorso di dottorato non è stato un'isola, ma si è nutrito del mondo che continuava a esistere fuori dall'accademia. Dunque, non posso che ringraziare le persone che hanno popolato la mia vita artistica e comunitaria. Ringrazio i collettivi di cui faccio parte e grazie ai quali posso continuamente imparare e formarmi: il *No Hate Speech Movement Italia* e l'*Associazione Maurice*, spazi di attivismo—ma, ancora più importante, di umanità—irrinunciabili. E grazie a chi compone le realtà teatrali e artistiche che attraverso, che in questi anni hanno dato energia e senso al mio essere.

Il personale è politico, e non posso quindi non ringraziare le mie sfamiglie, geograficamente sparse, fatte da persone che con me si organizzano quotidianamente per cambiare la società, a colpi di pigiama party, cortei, aperitivi, chiacchiere sul pianerottolo, fughe romantiche e videochiamate, ma soprattutto intimità, solidarietà, collaborazione, ascolto e una costante tutela e valorizzazione della libera espressione di sé. Sono tante, e in questi anni hanno contribuito a creare nuove idee di casa, facendomi spazio nella loro e rendendomi vivibili momenti insopportabili.

E tra le sfamiglie, un grazie speciale e dovuto va a chi compone quella napoletana. Ringrazio Elvira, Nando, Marta e Laura, per il sostegno materiale ed emotivo, e per avermi accolto in quella casa dove, per la prima volta, ho iniziato a sviluppare le idee che sono poi confluite nel mio progetto di dottorato

e mi hanno aiutato a essere qui. E ringrazio le persone che compongono la mia sfamiglia di origine, per aver messo in campo tutti gli strumenti a loro disposizione per comprendermi e per avermi dato la certezza che, anche se ci perdiamo, proviamo sempre a ritrovarci.

Infine, grazie alla mia testa: ce l'abbiamo fatta anche stavolta!

Contents

List of Figures

List of Tables

Preliminary Remark

This PhD work is a result of a study which involves different disciplines, methods and approaches. Although it has been mainly developed within the Department of Computer Science and the Department of Foreign Languages at the University of Turin, the intrinsic interdisciplinary nature of the PhD program within which this thesis was born, carried forward and then finalised allowed me to absorb praxes, perspectives, and precious material from other departments and researchers.

Within this interdisciplinary setting, the perspective adopted in this thesis is nevertheless grounded in a linguistic background, and more specifically in sociolinguistics and applied linguistics. This training shapes the way language is approached throughout the work, with particular attention to social context and interpretative processes through which linguistic forms acquire meaning and value in use. From this standpoint, linguistic expressions are not treated as self-sufficient carriers of meaning, but as socially situated resources whose interpretation depends on speaker positioning, interactional context, and shared norms—including the ways in which such norms are reproduced, negotiated, or actively subverted in discourse.

Alongside academic training, the perspective developed in this thesis is also informed by previous and ongoing experience in LGBTQ+ activism and in initiatives aimed at countering hate speech, as well as by involvement in educational and training activities within international projects focused on the promotion of human rights and the development of alternative narratives and counterspeech to discriminatory discourse. These experiences do not simply function as external motivations, but as sources of situated knowledge that contribute to understanding how harmful language—including dangerous speech—is (re)produced, perceived, and contested in real-world settings.

Taken together, these academic and extra-academic experiences motivate a view of annotation and interpretation not as purely formal or technical operations, but as inherently interpretative and socially grounded practices.

This assumption underlies the methodological choices made in this thesis and informs the way interdisciplinary perspectives are integrated, while remaining attentive to the ethical and epistemic implications of investigating socially sensitive language phenomena.

As in any research work worthy of the name, the content of my thesis depends substantially on the numerous opportunities for interaction I have had with other researchers: some who have simply inspired me, and others who have contributed more concretely to my research. As this work aims at building up a methodological structure which may give a broader scope to the computational study of the social phenomena here investigated by putting together perspectives and praxes from different and diverse disciplines, it was developed accordingly by putting together researchers—which are first of all people—from different and diverse backgrounds and areas of expertise. Therefore, I will use the plural first person as plurality characterised the whole process, even though at the same time I take individual responsibility for what is written and presented here.

Chapter 1

Introduction

This thesis linguistically investigates two social phenomena which—in the perspective here adopted—is not widely explored to date: gender stereotypes and reclamation[1] of LGBTQ+[2] slurs. In the next sections we present the strand of research within which we position ourselves, our disciplinary framework, our research questions, and, subsequent, the objectives and the scope of our study. Lastly, an overview of the thesis is outlined.

1.1 Strand of Research: Socially Situated Automatic Detection of Harmful Language

This thesis is situated within the field of Natural Language Processing (NLP), and more specifically within research on the automatic detection of harmful language, including hate speech (HS). Over the past two decades, the

[1] As clarified by Cepollaro and López de Sa (2022), this phenomenon is also referred to in the literature as *appropriation* or *reappropriation*. These labels, however, may be confused with debates on *cultural appropriation*, which concern a distinct phenomenon. Throughout this thesis, we primarily use the term *reclamation* together with its corresponding adjectival forms, *reclamatory*. At the same time, when discussing scholarship that adopts a different terminology, we also use *reappropriation* and *reappropriative*. We also specify here that the implicit focus in our work is always on *linguistic* reclamation. However, this does not exclude the possibility that non-linguistic symbols may also be reclaimed (or, perhaps more precisely in such cases, reappropriated), although this falls outside the scope of our research.

[2] For the sake of terminological rigour, it is useful to clarify the use of acronyms relating to sexual and gender minority identities within this work. Whilst the broader acronym LGBTQIA+ is acknowledged as equally valid, LGBTQ+ has been adopted as the standard term throughout the thesis to maintain continuity with choices made in prior academic contributions. However, in all sections devoted to the analysis of the FAVOLOSA project, the LGBT+ label has been retained. This faithfully reflects the terminology agreed upon and utilised by those involved during the project's development phase, with the understanding that the '+' suffix ensures a fully inclusive scope across all formulations.

widespread adoption of social media platforms has intensified the circulation of discriminatory and harmful discourse online, raising concerns about its impact on public debate, social cohesion, and democratic processes. In response, NLP and Machine Learning have increasingly been employed to develop automatic systems for identifying and mitigating such content.

HS detection has developed into a well-established research area, as documented by influential surveys such as Fortuna and Nunes (2018). Their work highlights the absence of a universally agreed-upon definition of HS and the resulting overlap with related notions such as abusive or offensive language. Crucially, they emphasise that HS cannot be treated as a purely lexical phenomenon, but rather as a context-dependent form of language use that relies on social, pragmatic, and interpretative factors. From a complementary, resource-oriented perspective, Poletto et al. (2021) provide a systematic review of benchmark datasets and annotation schemas for HS detection. Their analysis reveals a strong prevalence of explicit and overt forms of hate, alongside persistent challenges related to annotation reliability, contextual ambiguity, and the treatment of borderline cases. Within this research tradition, established practices guide the collection, filtering, and annotation of data, typically relying on widely used label sets that reflect the field's accumulated knowledge and shared assumptions.

Moreover, recent advances in HS detection have largely focused on deep learning architectures, especially transformer-based models, which enable context-sensitive representation learning across languages and domains (Jahan and Oussalah, 2023). While these approaches have significantly improved classification performance, they do not, on their own, resolve the conceptual and methodological challenges identified in earlier work—particularly those concerning the social grounding of language and the interpretative nature of annotation.

These limitations are particularly evident when addressing phenomena such as stereotypes (see Chapter 3, Section 3.1) and the contextual use of slurs (see Chapter 3, Section 3.2), where harm and resistance intersect, and cannot be easily disentangled. Addressing these issues requires more than computational models alone: it calls for analytical frameworks that explicitly account for the social embeddedness of language, the interpretative nature of annotation, and the role of context in shaping meaning.

Recent empirical work in NLP has begun to investigate stereotypes as a distinct object of analysis, highlighting their role in sustaining harmful discourse and their relevance for automatic language technologies. In particular, Schmeisser-Nieto et al. (2024a), within the STERHEOTYPES project,[3] introduce the *StereoHoax* corpus, a multilingual dataset of social media conversations centred on racial hoaxes related to immigration, and annotated for the presence of stereotypes. The corpus adopts a fine-grained annotation schema grounded in social psychological models, presented for the first time in Bosco et al. (2023), and explicitly distinguishes between implicit and explicit stereotypes, as well as between cases whose interpretation depends on the availability of conversational context.

This line of work points to a broader issue concerning the role of stereotypes in harmful discourse. As Cignarella, Giachanou, and Lefever (2025) highlight, stereotypes provide the cognitive and social scaffolding for prejudice and discrimination, ultimately sustaining harmful discourse. While research on gender bias and hate speech has reached a certain level of maturity within NLP, stereotype detection remains comparatively under-explored despite its profound implications for both social harm and automatic detection systems. Drawing on insights from Psychology, Sociology, and Computational Linguistics, Cignarella, Giachanou, and Lefever (2025) argue that identifying implicit and intersectional stereotypes can function as a proactive strategy to prevent the escalation of bias into overt hate. Their survey further documents emerging multilingual datasets (such as the aforementioned *StereoHoax*), classification tasks, and debiasing strategies that extend beyond gender and race, capturing a broader spectrum of social groups and forms of marginalisation.

Similarly, NLP research on linguistic reclamation remains limited, despite its relevance for understanding context-dependent and community-specific

[3]STERHEOTYPES (STuding European Racial Hoaxes and sterEOTYPES) was an international three years (2021–2024) research project funded under the "Challenge for Europe" call by Carlsberg Foundation, Volkswagen Stiftung and Compagnia San Paolo, led by University of Bari "Aldo Moro" (Italy) in partnership with University of Barcelona (Spain), Uninettuno University (Italy), University of Torino (Italy) and University of Toulouse "Paul Sabatier" (France). "The project focuses on 'racial hoaxes', communicative acts created to circulate information that are an allegation of a threat posed against someone's health or safety associated to individual or a group because of race, ethnicity or religion. Aiming at understanding the social and psychological processes emerging from racial hoaxes in digital generations across three border Mediterranean European countries (Italy, Spain and France), it integrates different methodological approaches coming from psychology and computational linguistics" (https://www.irit.fr/sterheotypes/).

language use. As documented in Ferrando et al. (2025), the reclamation of slurs constitutes a major open challenge for HS detection, as it involves the reappropriation of historically derogatory terms by members of target groups to express identity, belonging or solidarity. While this phenomenon has been extensively discussed in Philosophy of Language and Sociolinguistics, it is largely overlooked in NLP, where models still struggle to distinguish between abusive and non-abusive uses of the same lexical items.

Ferrando et al. (2025) highlight that this limitation contributes to the risks of both over-moderation and under-moderation in automatic content moderation systems, potentially leading to silencing marginalised voices. To address this issue, they argue for the need to collect fair and community-informed data, and to explicitly account for speakers' intentions, social context, and community norms when modelling reclaimed slurs. From this perspective, slur reclamation emerges as a paradigmatic case illustrating the limits of surface-level approaches and the necessity of socially grounded annotation practices.

Building on these observations, this thesis approaches the automatic detection of potentially harmful language as a socially situated linguistic task. In line with recent calls for socially aware language technologies (Yang et al., 2025), the present study foregrounds the socially embedded nature of language use, and investigates how gender stereotypes and LGBTQ+ slur reclamation challenge standard HS detection frameworks, particularly at the level of data annotation.

1.2 An Interdisciplinary Framework

Research on HS has progressively expanded beyond the identification of hostile expressions to encompass a wider range of socially grounded linguistic phenomena. In this context, stereotypes and slur reclamation occupy a critical position: they are not inherently hateful, yet they may enable, mitigate, or reframe harm depending on how they are linguistically realised and interpreted.

To account for this complexity, this thesis adopts an interdisciplinary

framework that places language at its centre, drawing on social sciences,[4] in particular on three socially grounded linguistic and philosophical traditions that are rarely brought jointly to bear on these phenomena in NLP: Sociolinguistics, Queer Linguistics, and Philosophy of Language.

Within this framework, Sociolinguistics is employed to analyse language as a social practice, where meaning is shaped by context and communities, particularly concerning gendered language and ingroup versus outgroup dynamics. Queer Linguistics foregrounds how LGBTQ+ communities negotiate, contest, subvert, and reclaim stigmatising language, and how identity and community membership condition the interpretation of potentially harmful terms. Philosophy of Language, in turn, provides conceptual tools for examining identity, meaning, speaker intention, and interpretation. Additionally, research in Social Psychology informs the broader study of stereotypes.

What distinguishes the present work is the direction and the parity of the exchange. Rather than enlisting socio-linguistic and philosophical knowledge as an instrumental resource at the service of NLP—or, conversely, treating NLP as a mere measurement instrument for the social sciences—this thesis treats the two as equal partners.

Our work aligns with recent perspectives that advocate for socially aware language technologies (Yang et al., 2025), as well as with perspectivist approaches that recognise disagreement and partiality as meaningful dimensions of language annotation rather than mere noise (Frenda et al., 2024). From this standpoint, annotation is here understood as an interpretative and socially situated practice, grounded in the notion of situated knowledge (Haraway, 1988), in which annotators' positionality and background knowledge play a constitutive role. This perspective is particularly relevant for phenomena such as gender stereotypes and LGBTQ+ slur reclamation, whose interpretation cannot be dissociated from community norms, social perspectives, and lived experience.

Within this framework, this thesis aims to contribute to the development of annotation methodologies that are both computationally viable and socially informed, fostering a reciprocal dialogue between NLP and the socio-linguistic

[4]Throughout this thesis, the term "social sciences" is used in a deliberately narrow sense. It does not refer to the broad and growing field of Computational Social Science (CSS, Lazer et al., 2009; Edelmann et al., 2020), but to the socially grounded linguistic and philosophical traditions named above.

and philosophical disciplines, while remaining grounded in the analysis of linguistic data.

1.3 Research Questions

Starting from the premises so far presented, our research is driven by the following research questions:

- RQ1. How and to what extent can methods and practices informed by socio-linguistic and philosophical knowledge contribute to NLP research?
- RQ2: What are the possibilities and limits of adopting a queer perspective in a study aimed at developing an annotation schema?
- RQ3: How can stereotypes and reclamation inform both Linguistics and NLP?
- RQ4: How and to what extent can gender stereotypes and LGBTQ+ slur reclamation be framed so as to inform the development of a fair, community-centred, and participatory annotation schema for automatic detection?

1.4 Objectives and Scope of the Research

Once clarified the research questions, we enumerate the objectives of our study, which are multiple.

The broad objectives are both ethical and technical, as we want to:

- propose an original methodology for the computational study of pragmatic phenomena that takes into account the inherently social nature of language;
- integrate a queer perspective into NLP methodologies;
- connect the sociolinguistic and computational fields for the development and the validation of NLP resources.

The specific objectives lead our investigation in order to:

- lay the foundations for the creation of annotation schemas sensitive to human interpretation for the automatic detection of gender stereotypes and LGBTQ+ slur reclamation;
- design an annotation process where the participation of annotators is central and the positionality of the researcher is clear;
- collect quantitative and qualitative data for the definition—and the study—of the phenomena investigated, i.e. gender stereotypes and LGBTQ+ slur reclamation.

1.5 Overview

The remainder of this thesis is organised as follows:

- in Chapter 2 we provide the conceptual and methodological background of our work, defining disciplines and perspectives which enriched and directed this thesis;
- in Chapter 3, we outline the state-of-the-art of the two investigated phenomena, i.e. gender stereotypes and slur reclamation, mainly focusing on what has been done in NLP;
- Chapter 4 is conceived as a junction between the theoretical background and the case studies: we introduce here our paradigm, by clarifying how our research questions harmonise with this background and what the case studies presented in the next chapters may tell us;
- in Chapter 5, we present a preliminary literature survey and a preliminary study about gender and language in Italian that we have carried out before approaching the two case studies;
- in Chapter 6, we present the first case study, which investigates the annotation of gender stereotypes in Italian Twitter discourse through the Rackete-Schettino corpus;
- in Chapter 7, we present the second case study, which explores LGBTQ+ slur reclamation in Italian and reflects on its implications for fair and socially situated NLP;

- we conclude with Chapter 8, where we draw our conclusions and outline possible further research.

Chapter 2

Conceptual-Methodological Background: Disciplines and Perspectives

This chapter is organised into three sections. Section 2.1 presents a literature review outlining the challenges and emerging perspectives in applying Natural Language Processing for social good. Section 2.2, on the other hand, draws on insights from the social sciences to contextualise and frame the perspective underpinning our work. Section 2.3 presents the main techniques used in our work and their methodological implications.

Importantly, the chapter does not aim to provide an exhaustive review of either NLP scholarship or social-science theories, but rather to outline the specific conceptual and methodological perspectives that most directly inform the present research.

2.1 Challenges and Emerging Directions in Socially Aware NLP

Language represents a core component of human existence, serving as a vehicle for conveying information, articulating emotions, generating meaning, and shaping individual and collective identities (Jahan and Haque, 2023). Its social roles—ranging from maintaining societal structures and persuading others to providing entertainment and marking group affiliation—are equally vital, and cannot be overlooked (Hovy, 2018).

To date, the field of Natural Language Processing (NLP) has predominantly concentrated on the analysis and interpretation of language, with particular emphasis on developing datasets and computational models aimed at the automated detection of specific phenomena, such as irony, toxicity, abusive language, and hate speech[1]. The overarching objective remains the enhancement of model performance across a variety of linguistic tasks.

However, in recent years a growing body of research has examined the current limitations faced by NLP studies. Hovy (2018) argues that the field has increasingly centred on tasks that align well with statistical modelling techniques, often overshadowing the investigation of more nuanced linguistic phenomena. He advocates for greater attention to the social functions of language, and emphasises the necessity of acknowledging demographic variables that can shape NLP model behaviour. Although this perspective marks a significant step toward re-integrating sociolinguistic insights—by proposing novel neural architectures and methodological frameworks—there remains a notable scarcity of direct human involvement in the research process.

Other researchers have highlighted the potential benefits of drawing more explicitly from Sociolinguistics, positing that accounting for linguistic variation and social meaning is essential for building more robust NLP models. Despite advances in semantic representation, the social dimensions of meaning are still largely neglected. For instance, Nguyen, Rosseel, and Grieve (2021) argue that integrating social meaning and linguistic variability could significantly enhance the automatic detection of abusive language. Yet, as in previous approaches, the perspectives of those directly affected—both victims and perpetrators—appear conspicuously absent from the investigative process.

2.1.1 Involving Humans in the Pipeline

With the rise of Deep Learning (DL), the tasks of collecting and annotating labelled data have become increasingly labour-intensive and time-consuming. To address these difficulties, researchers have adopted strategies such as

[1]These labels are not always defined in a fully consistent way across the literature. In particular, categories such as *abusive*, *offensive*, *toxic*, and *uncivil* often partially overlap and may be operationalised differently across datasets and annotation frameworks (Pachinger et al., 2023).

generating synthetic datasets, accelerating human-model interaction, reducing annotation costs, and leveraging pre-trained models along with transfer learning techniques. Nevertheless, the need for domain-specific labelling and frequent updates often remains crucial to attaining high-performing and reliable models.

In response, Human-in-the-Loop (HITL) methodologies have gained ground by integrating human expertise directly into the learning and modelling pipeline. As noted by Kumar et al. (2024), researchers are progressively embedding existing human knowledge within Machine Learning (ML) frameworks. Their work underscores the significance of inherently human factors—such as emotional states and practical abilities—which may affect both teaching and learning outcomes. The paper outlines a range of strategies for incorporating human feedback into ML workflows.

HITL approaches have seen broad application in NLP, where they have contributed to improved model accuracy, interpretability, and user engagement. These systems harness the complementary strengths of humans and machines, resulting in enhanced decision-making, streamlined task automation, and more effective processes overall. While we recognise the value of such collaboration for refining model performance and generalisation, we also argue for the importance of involving humans at earlier stages of the research process, not merely as annotators or evaluators, but as active contributors to framing and contextualising the investigation itself.

However, involving humans in the loop does not by itself address the question of whose interpretations are represented and how disagreement should be understood.

2.1.2 Perspectives Matter

A growing theoretical movement within Artificial Intelligence (AI) research—called Perspectivism—advocates for enhancing machine learning models by incorporating annotations from individuals with diverse backgrounds and viewpoints. This paradigm challenges the conventional annotation practice, which tends to treat disagreement as statistical noise to be minimised or eliminated through forced consensus, either by harmonisation or automatic aggregation. As articulated in the *Perspectivist Data Manifesto*, "aggregation and harmonisation destroy any personal opinion, nuance, and rich linguistic

knowledge that come as a result of the different cultural and demographic background of the annotators"[2].

In a recent contribution, Casola et al. (2024) exemplify the perspectivist approach through the creation of a multilingual corpus of ironic short dialogues, collected from platforms such as Twitter and Reddit, and encompassing multiple languages and dialects. Their findings reveal that annotators from different demographic groups frequently disagree on what constitutes irony, underscoring the influence of cultural and social factors on both perception and annotation. Additionally, the study illustrates how maintaining disaggregated annotations and rich annotator metadata enables more nuanced evaluation of Large Language Models (LLMs), particularly regarding their ability to detect irony, their alignment with various social perspectives, and their responsiveness to prompts encouraging perspective-taking.

Complementing this empirical work, Frenda et al. (2024) provide the first comprehensive survey on perspectivism within NLP. Their review systematically examines datasets that retain individual annotator labels, methods for modelling human perspectives, and the broader methodological implications of perspectivist practices. The authors explore how subjectivity is treated in tasks like toxic language detection and highlight that even tasks traditionally perceived as objective can exhibit systematic annotator disagreement. The survey also discusses techniques for capturing and assessing diverse perspectives and reflects on the theoretical and practical impact of perspectivism on the future of NLP research.

Along similar lines, recent work has emphasised that disagreement in subjective annotation tasks should be analysed rather than suppressed. In particular, Sandri et al. (2023) show that annotator disagreement often reflects different interpretative strategies, sensitivities, and understandings of the task, and that collecting qualitative feedback from annotators can be crucial for identifying sources of ambiguity and for refining annotation guidelines. This perspective further supports the view that disagreement is an inherent and informative component of socially sensitive annotation tasks.

The present research is informed by the perspectivist perspective, particularly in its critical stance towards traditional annotation methodologies that treat disagreement as noise. Building on this view, the thesis explores the

[2]https://pdai.info/

possibility of extending perspectivist insights beyond the annotation phase, by considering how people's lived experiences, interpretations, and backgrounds may also inform earlier stages of inquiry, including the formulation of research questions and task definitions.

In line with this view, we also find the Inclusivity by Design framework proposed by Kurrek, Saleem, and Ruths (2020) to be a promising model. This approach rethinks annotation practices by deliberately encouraging diversity of opinion through team-based annotation processes, where individuals are grouped based on contrasting demographic characteristics. While we recognise the value of this approach, and support the inclusion of heterogeneous annotator perspectives, we emphasise the importance of consulting people directly, in some cases even before data collection begins. In our view, inclusive participation at the conceptual and design phases is crucial to capturing the full spectrum of social meaning in language technologies.

Yet, even perspectivist approaches often remain focused on annotation, leaving open the question of whether participation should also shape earlier stages of research design.

2.1.3 Participation in NLP

In AI and adjacent design-oriented fields, Participatory Design (PD) has emerged as a strategy for centring the voices of individuals and communities affected by technological systems (Field et al., 2021). Birhane et al. (2022) emphasise the value of participatory methods in fostering responsible AI development, highlighting not only potential enhancements in algorithmic performance, but also opportunities for collective reflection and exploration. Within the broader discourse on the AI developed for social good, participatory practices are increasingly framed as mechanisms to ensure that impacted communities are involved as stakeholders throughout the design and implementation of AI systems. One of the core aims of participatory methodologies is to democratise knowledge about technological infrastructures by bringing together technical expertise and the lived experiences of non-expert users and communities.

Nevertheless, as Field et al. (2021) caution, PD should not be seen as a universal remedy. Sloane et al. (2020) critically examine how PD initiatives may devolve into what they term "participation-washing", i. e. superficial

or tokenistic forms of involvement that lack meaningful engagement. Their framework outlines three distinct conceptions of participation: *Participation as Work*, *Participation as Consultation*, and *Participation as Justice*. Of these, we consider *Participation as Justice* especially relevant, given its deeper social and political implications. The Design Justice framework proposed by the authors moves beyond conventional value-sensitive design by centring the experiences of marginalised groups and challenging dominant power structures, including those embedded in capitalist systems that shape the deployment and governance of ML technologies. We strongly align with this perspective, advocating for a *designing with* approach that ensures NLP outcomes are genuinely beneficial and reflective of diverse communities' needs.

In this vein, the study by Caselli et al. (2021) reinforces the role of PD as a mutual learning process between participants and researchers, with a particular emphasis on the empowerment of under-represented groups. In their analysis of 100 papers from the ACL Anthology that mention the term "community", only 9 showed any form of community engagement, and, notably, none involved communities directly in the design phase of the NLP systems under development. This lack of meaningful inclusion underscores a critical gap in current practices.

The need for NLP research to move beyond occasional collaborations with Sociolinguistics or other social disciplines could proficiently result in a deeper integration of participatory principles, starting from the earliest stages of research. This would involve not only consulting with communities but co-constructing the very research questions, objectives, and solutions with groups most affected by language technologies.

2.2 Social-Science Frameworks on Language, Identity, and Interpretation

A significant shift in research paradigms often stems from a re-orientation or expansion of perspectives. Therefore, in order to gain a deeper understanding of the linguistic phenomena explored in NLP studies and to re-design some procedures, in this PhD work we propose a methodological move relying on the search heuristics as suggested by Abbott (2004), which are here specifically

based on the adoption of approaches not yet widely investigated in NLP but well established in social sciences.

Within this broader orientation, Sociolinguistics constitutes a key point of departure, together with related frameworks that make it possible to examine language in relation to identity and social interpretation. Among these, Queer Linguistics (QL)—discussed in detail below—offers a particularly relevant critical lens for analysing how linguistic phenomena are bound up with social norms, identity categories, and power relations. Closely connected notions such as *intersectionality* and *positionality* further specify how meaning and interpretation are shaped by differential social locations, lived experience, and epistemic standpoint. Finally, Social Philosophy of Language provides an additional conceptual grounding by clarifying how, in socially sensitive cases, interpretation depends on the interplay between semantics, pragmatics, and socially available contextual assumptions.

Therefore, this shift provides an important theoretical foundation for integrating insights from social sciences into computational research, as it allows linguistic phenomena to be examined in relation to social practices rather than as decontextualised textual elements.

2.2.1 Language as Social Practice: A Sociolinguistic Lens

Sociolinguistics has long explored the relationship between language and society, showing that linguistic forms and practices are patterned in relation to social differentiation, community norms, and the contexts in which they are used (Labov, 1973; Berruto, 2004). From this perspective, language cannot be treated as an autonomous object detached from social structure; rather, it must be analysed in relation to the communities and social values through which it becomes significant. This foundational insight is especially important for the present thesis, since the phenomena addressed here—such as stereotypes, gender nomination, derogatory expressions, and their reclamatory uses—cannot be adequately understood without reference to the conditions under which they are produced, interpreted, evaluated, and, perhaps, contested.

Recent sociolinguistic scholarship has further expanded this view by questioning traditional ontological assumptions about language itself. Instead of treating language as a bounded and stable system, contemporary work increasingly foregrounds it as *social practice*, that is, as inseparable from social

action, ideological processes, and situated meaning-making. A particularly useful account of this shift is offered by Wang, Jin, and Li (2023), who trace how Sociolinguistics has progressively broadened its understanding of what counts as language, emphasising its embeddedness in power relations, interactional conditions, and broader semiotic practices.

Importantly, this more recent orientation should not be seen as replacing earlier sociolinguistic concerns, but as extending them. Attention to variation and to the correlation between linguistic forms and social factors remains central; however, it is now situated within a broader account in which linguistic practices are also understood as participating in the production, interpretation, evaluation, and contestation of social meanings. In this sense, Sociolinguistics offers a reflexive framework for understanding how identities and categories acquire social salience through language use, and how this process is grounded in recurrent forms of participation in shared social practices.

A useful refinement of this perspective is offered by the notion of *community of practice*. The term was first developed by Lave and Wenger (1991) as part of a social theory of learning, and was later brought into Sociolinguistics (see Eckert and McConnell-Ginet (1992); Eckert and Wenger, 2005; Eckert and Brown (2006); Eckert and McConnell-Ginet (2007)) as a way of connecting broad categories such as gender to on-the-ground social and linguistic practice. In Eckert and Brown (2006)'s account, a community of practice is a collection of people who engage on an ongoing basis in some common endeavour. Its relevance lies in the fact that it identifies a social grouping not through shared abstract characteristics, nor through simple co-presence, but through shared practice. In the course of regular joint activity, participants develop ways of doing things, views, values, power relations, and ways of talking. For this reason, the community of practice becomes a particularly useful locus for the study of situated language use, since it offers an accountable link between the individual, the group, and the broader social order.

This re-conceptualisation of language also creates an important bridge to the critical approaches discussed in the following subsections. In particular, it resonates strongly with Queer Linguistics, since both challenge essentialist categories and foreground the role of linguistic practices in constructing and contesting social norms. Adopting a sociolinguistic lens on language as

socially situated practice therefore allows this thesis to establish a coherent connection between social theory and computational analysis, while also justifying methodological choices that seek to preserve context, interpretation, and perspective rather than abstracting away from them.

2.2.2 Queer Linguistics as a Critical and Methodological Framework

Building on the sociolinguistic perspective outlined above, Queer Linguistics (QL) is an interdisciplinary field of study rooted in feminist and LGBTQ+ scholarship, which investigates the interplay between language, gender, and sexuality through an intersectional framework (Davis, 2008). It fundamentally challenges hetero-cis-normative ideologies embedded in language and discourse (Cameron and Kulick, 2003), calling into question fixed and essentialist understandings of identity.

Within this framework, language is not merely a vehicle for expressing pre-existing identities but a constitutive site where identities are continuously negotiated, contested, and reconfigured. This view aligns with sociolinguistic approaches that reject static and enumerative conceptions of language in favour of an understanding of linguistic practices as fluid, context-dependent, and deeply intertwined with social power (Wang, Jin, and Li, 2023). From this perspective, QL can be seen as part of a broader ontological shift in the social sciences, one that emphasises language as practice rather than system.

Rather than simply analysing representations of queer identities, QL positions language as a site of resistance, transformation, and performative expression (Butler, 1990), which highlights the role of discursive agency in both the construction and disruption of gender and sexual norms (Bucholtz and Hall, 2004). This is especially relevant for the present thesis, since QL makes it possible to approach gendered and sexual categories not as stable descriptive givens, but as socially mediated and contested formations that are negotiated in discourse. Research within this field has explored linguistic practices within queer communities, the reclamation and re-contextualisation of slurs, and the disruption of dominant narratives through subversive language use (Leap, 1996). In this respect, Edmondson (2021)'s work is particularly useful, as it frames slurs targeting LGBTQ+ people not simply as

negative labels, but as expressions tied to broader relations of the Gramscian conception of hegemony, marginalisation, and exclusion; correspondingly, linguistic reclamation is treated not as a neutral semantic shift, but as a politically motivated process through which members of marginalised groups may use such slurs for non-derogatory purposes.

Therefore, by dismantling binary logics and essentialist assumptions, QL brings visibility to a spectrum of identities and linguistic strategies and, furthermore, provides a rigorous critique of the sociopolitical structures embedded in language itself (Motschenbacher, 2010).

QL does not exist in isolation but forms part of a wider constellation of approaches in Language, Gender, and Sexuality (LGS) studies, which together offer valuable insights into the methods and theories used in this thesis. As outlined by Motschenbacher (2019), LGS scholarship includes a range of approaches—such as Feminist Linguistics, Lavender Linguistics, Discursive Gender Linguistics, and QL—that, in different ways, examine how language shapes the social intelligibility of gender and sexuality. Within this broader constellation, QL is especially important here because it pushes more explicitly against fixed classificatory schemes, foregrounds non-normative practices and identities, and encourages critical reflection on the categories through which linguistic analysis is conducted.

In this sense, QL is not only a theoretical lens but also a methodological stance. It invites researchers to interrogate the categories they rely on, to attend to marginalised voices, and to treat overlaps, ambiguities, and contested meanings as analytically significant rather than as noise. As Motschenbacher notes, this can be pursued through a wide range of approaches—from sociolinguistic and discourse-oriented analyses to multimodal, mixed-methods, and corpus-based work (Motschenbacher, 2019).

For the purposes of this thesis, the main value of QL lies precisely here: it provides a principled basis for challenging rigid or binary representations of gender and sexuality, while also offering a framework for analysing how marginalised groups negotiate, resist, or transform dominant ideologies through language. This is particularly relevant to both the study of stereotypes and the analysis of reclaimed slurs. Edmondson's distinction between different possible contexts of reclamatory use—for instance, uses in reference to oneself or to others, and uses by ingroup or outgroup speakers—is

especially helpful in this respect, because it shows that reclamation is not a single undifferentiated phenomenon, but one whose interpretation depends on speaker position, target, and social relation (Edmondson, 2021).

By taking up this QL framework, this thesis follows Motschenbacher's call for diverse and reflexive methods that stay critical of their own assumptions. More specifically, it treats categories of gender and sexuality as analytically necessary but never self-evident, and therefore as requiring continuous reflection in both theoretical framing and methodological design. This approach makes it possible to capture the shifting, contested nature of gendered and sexual identities in discourse, and to show how language is used both to reinforce and to challenge normative ideas. QL thus provides not only a critical stance toward normative categories of gender and sexuality, but also a principled justification for questioning the linguistic abstractions that often underpin computational models.

Recent studies have indeed begun to integrate QL into computational and corpus-based approaches. For example, Youngquist (2023) uses distributional semantic models alongside collocation and concordance analyses to examine identity labels in the r/lgbt subreddit. This study shows how computational methods can reveal the persistence of binary relationships ("cis-trans", "gay-straight") while also highlighting areas where labels shift or overlap (such as "bi" and "pan"). Youngquist combines these quantitative tools with a critical QL perspective, demonstrating how online discourse both sustains and challenges normative categories.

Although my own work does not focus on developing distributional models, it aligns with this trajectory by proposing innovative and fair annotation schemas grounded in QL principles. In doing so, it extends the critical focus of QL into the design of computational resources, ensuring that annotation itself reflects the diversity and fluidity of gender and sexuality, with the attempt of not reproducing restrictive binaries.

2.2.2.1 Intersectionality

Over the past few decades, intersectionality has emerged as a central analytical framework in feminist theory and beyond. Originating in Black feminist thought, it offers a powerful critique of essentialist identity categories by emphasising how multiple dimensions of identity—such as race,

gender, sexuality, class, and disability—interact to shape experience and social position (Crenshaw, 1989; Crenshaw, 1991; Collins, 2000). This framework resonates with the broader discursive turn in Linguistics, which also challenges universalist and static conceptions of meaning and identity (Esposito, Pérez-Arredondo, and Zottola, 2024).

Intersectionality provides the tools to account for the complexity and simultaneity of social inequalities. As Davis (2008) notes, its conceptual strength lies in its ability to navigate intricately layered systems of oppression without reducing them to single dimensions. This has made it particularly influential in feminist and queer theory, where it allows for a more textured understanding of power relations and identity formations (Lykke, 2010).

In the context of linguistic research, intersectionality enables the analysis of how discourse both reflects and enforces hierarchical structures. It complements discursive approaches that view identity as dynamic and context-dependent, rather than fixed or intrinsic (Bucholtz and Hall, 2004). This perspective is especially relevant to the methodology proposed in this work. By incorporating an intersectional lens, we aim to critically examine how language and language practices are implicated in the reproduction-or contestation-of intersecting systems of privilege and oppression. In doing so, we gain a deeper understanding of the discursive mechanisms through which social inequality is maintained or challenged.

Adopting an intersectional perspective also entails a critical awareness of how analytical categories are constructed and operationalised. From a sociolinguistic standpoint, this implies recognising that linguistic forms and meanings cannot be interpreted independently of the intersecting social positions of speakers and hearers. Such an approach is consistent with the aforementioned sociolinguistic work that foregrounds the entanglement of language, power, and social practices, cautioning against analytical simplifications that obscure structural inequalities (Wang, Jin, and Li, 2023). In this sense, intersectionality functions as both an interpretive framework and a methodological safeguard against reductive analyses.

2.2.2.2 Positionality

Positionality refers to the recognition that all knowledge production is situated and shaped by the social, cultural, and epistemic positions of those involved

in the research process. Rather than striving for an illusory form of neutrality, contemporary qualitative and mixed-methods research emphasises reflexivity as a necessary condition for methodological rigour and ethical accountability (Yip, 2024). This perspective is particularly relevant for studies that engage with marginalised communities or sensitive identity categories, where the researcher's positional stance directly influences data collection, interpretation, and representation.

From an epistemological standpoint, positionality builds on feminist critiques of objectivist science, most notably the concept of situated knowledges (Haraway, 1988). Haraway argues that all forms of knowledge are partial and embodied, and that acknowledging one's position does not weaken scientific claims but instead strengthens their credibility. Similarly, Harding (1991)'s *standpoint theory* highlights how power relations shape what counts as legitimate knowledge, urging researchers to interrogate whose perspectives are centred and whose are marginalised. In this respect, Rolin (2009) offers a useful methodological clarification: the epistemic advantage associated with subordinated standpoints should not be understood as automatic, but as contingent on particular epistemic projects and on the concrete social conditions under which knowledge is produced. Rolin also stresses that, in the study of power relations, evidence may be suppressed, distorted, or rendered difficult to articulate precisely because relations of power shape what can be said, by whom, and under what conditions.

In the context of this thesis, positionality operates at multiple levels. Methodologically, it informs the design of sociolinguistic instruments such as focus groups and surveys, guiding decisions about participant selection, question framing, and the interpretation of metalinguistic judgments. At a later stage, positionality becomes crucial in the development of annotation guidelines and the selection of annotators for computational tasks. Since annotation necessarily involves interpretive choices, making positional assumptions explicit helps prevent the uncritical reproduction of normative or binary categories within NLP resources. This point is especially important if participants and annotators are understood not as passive sources of data, but as situated knowers whose perspectives help determine what becomes visible and analytically salient in the research process. From this angle, positionality confirms its methodological function: it helps create the conditions under

which socially sensitive evidence can emerge with greater reflexivity and awareness of the power relations that may otherwise inhibit or distort it.

By foregrounding positionality, this work aligns with QL in treating reflexivity not as an optional add-on but as an integral component of rigorous analysis. Explicitly articulating the situated nature of analytical decisions allows for greater transparency, and supports the development of more socially aware and ethically grounded computational models.

2.2.3 A Philosophical Perspective on Semantics and Pragmatics of Socially Sensitive Language

This subsection provides a philosophical consolidation of the Sociolinguistic and QL perspectives discussed above, by making explicit a set of assumptions about meaning, identity, and interpretation that are already operative in the preceding sections. Rather than introducing an independent philosophical framework, the aim is to clarify how Social Philosophy of Language helps account for phenomena that are methodologically central to this thesis, in particular the role of speaker identity and the systematic difficulties involved in interpreting socially sensitive language in online contexts.

A core motivation for adopting this perspective is that several linguistic phenomena typically addressed within "NLP for social good"—such as hate speech, slurs, and stereotypes—are characterised by a high degree of interpretative underdetermination and subjectivity. In these cases, meaning cannot be recovered from textual form alone, but depends on pragmatic and social factors, including assumptions about the identity of the speaker or writer, their relation to relevant social groups, and the expected audience. As argued by Cepollaro and Labinaz (2019), social media environments make this dependence particularly salient, as contextual cues are often sparse, and the identity of users is frequently inferred rather than known. Under such conditions, losing track of who is speaking may radically alter the interpretation of an utterance, with direct implications for the assessment of discrimination and for moderation or censorship practices.

2.2.3.1 Identity, Social Context, and Linguistic Interpretation

From a social-pragmatic standpoint, identity is not treated here as an essential or intrinsic property of individuals, but as a socially salient parameter that affects what an audience is warranted to infer from an utterance. Cepollaro and Labinaz (2019) emphasise that, in online interaction, the reduced availability of prosodic and contextual cues increases the likelihood of misunderstanding, thereby amplifying the role of speaker identity in guiding interpretation. In particular, assumptions about who is speaking may support or block non-literal readings, such as irony or humour, especially when these involve discriminatory language on the surface.

A crucial methodological consequence of this view is that disagreement in interpretation should not be dismissed as mere noise. Instead, divergent interpretations may reflect systematic differences in how interlocutors or annotators reconstruct identity, audience, and context. In computational settings, this challenges the assumption that socially loaded language can be assigned to a single, context-free label, and supports approaches that treat disagreement as a meaningful signal rather than an error to be eliminated.

Slurs and derogatory epithets constitute a paradigmatic case for illustrating the dependence of meaning on social context and identity. In Bianchi (2015)'s reconstruction of the philosophical debate, the offensive dimension of slurs has been treated within three broad perspectives: a *semantic* perspective, according to which derogatory content is part of the literal meaning of the expression; a *pragmatic* perspective, according to which such content is not literally expressed but conveyed through the use of the expression in context; and a *deflationary* perspective, according to which slurs are socially prohibited words whose offensiveness does not primarily depend on encoded or conveyed content, but on a kind of socially sustained prohibition. This theoretical divergence becomes especially salient in cases of reclamation, since uses by members of the target group are often treated as non-offensive, and therefore put pressure on explanations that assume that derogatory force contributes uniformly to content across contexts. In this context, and within the framework of *Relevance Theory*, she develops an *echoic* account of reclamatory uses. *Echoic uses* are described as "un sottoinsieme degli usi attributivi o interpretativi – in cui il parlante non si limita a riportare il contenuto dell'enunciato o del pensiero attribuito, ma vuole informare chi ascolta del proprio atteggiamento

verso di esso"[3] (Bianchi, 2015, p. 294). Reclamatory language falls within this account insofar as speakers *echo* derogatory uses, or the discriminatory perspective associated with them, while making clear their dissociation from that perspective. Bianchi clarifies, however, that this proposal does not require an actually uttered prior sentence to be present in the immediate context: what is echoed may also be a socially available norm. Nor does the account require the same dissociative attitude in every case, since the stance may range from distance to mockery or sharper criticism. She also notes that many of these uses still look fully descriptive, in that the speaker commits to the truth of the neutral counterpart while not endorsing the offence carried by the slur. Finally, the echoic proposal is not limited to semantic accounts alone, since, within *Relevance Theory*, even implicitly conveyed content can be echoed.

On this view, reclamatory uses can be read through both a semantic and pragmatic perspective, as they do not erase the offensive meaning of a slur, but may continue to evoke it while pragmatically signalling that it is not endorsed.

Importantly, as both Bianchi (2015) and Cepollaro and Labinaz (2019) observe, the interpretation of reclamation uses is particularly fragile in online settings, where speaker identity may be uncertain or contested. In these contexts, the same lexical item may plausibly be interpreted either as derogatory or as non-offensive, depending on the assumptions made about who is speaking and how the utterance is intended to be taken up. This reinforces the view that meaning is not fully determined by form, but emerges from socially situated practices of interpretation.

The relevance of identity for interpretation becomes especially evident in relation to gender. Drawing on Feminist Philosophy of Language and gender theory, Judith Butler argues that gender is not a pre-discursive property of individuals, but is constituted through repeated social and linguistic practices—a process they famously characterises in terms of performativity (Butler, 1990).

While Butler's work originates outside the analytic tradition, its core insights have been extensively taken up within Analytic Feminism, particularly in debates concerning the metaphysics and semantics of gender categories. In this literature, gender is analysed as a socially constituted category whose

[3]Our translation: "a subset of attributive or interpretive uses, in which the speaker does not merely report the content of the utterance or thought being attributed, but also intends to inform the hearer of their own attitude towards it".

linguistic articulation plays a central role in sustaining normative expectations and asymmetries of power (Haslanger, 2000; Haslanger, 2012).

This perspective also provides a philosophical framework for understanding the persistence of gender stereotypes by clarifying the role of *generics* in social interpretation. Drawing on Haslanger (2011), generics can be understood as generalisations of the form *Ks are F* that should not be treated as ordinary quantified statements—statements containing an explicit quantifier such as "all", "most", or "some" (e.g. "Most women are emotional", *Most Ks are F*)—since they concern a category in general rather than a determinate proportion of its members. On this view, they may pragmatically invite the inference that the predicated property holds by virtue of what members of the relevant kind are, thereby introducing into interpretation a claim of *generic essence* (Haslanger, 2011). This is especially important for the analysis of stereotypes, because apparently descriptive utterances may carry socially consequential implications without being overtly framed as normative or prejudicial.

Recent work by Cella and Rosola (2024) further sharpens this point by focusing on the role of generics in sustaining social essentialism. More precisely, Cella and Rosola highlight that the issue is not exhausted by the question of whether generics promote essentialisation more than quantified generalisations: generics are also especially insidious because of their *inferential asymmetry* and *slipperiness*. They may be accepted on the basis of limited evidence, interpreted as applying broadly to the members of a group, and remain difficult to counter even in the presence of strong counter-evidence.

Read together with Haslanger's account of generics, common ground, and ideology (Haslanger, 2011), this suggests that generics are socially operative generalisations that shape interpretation, credibility, and expectations about members of social groups precisely because their implications can become part of the interpretive background. This property makes them particularly effective vehicles for transmitting and stabilising stereotypes, while simultaneously rendering them resistant to counter-evidence. From the perspective developed here, generic statements about gender—such as "Women are emotional"—exemplify how linguistic form, social expectations, and interpretation interact to produce robust and socially consequential meanings. When such implications go unchallenged, they may enter the common ground and contribute to

the naturalisation of socially produced regularities, making them appear as if they were grounded in the nature of the category itself rather than in broader social relations. Even when not intended to be discriminatory, such utterances can contribute to the normalisation of essentialist assumptions about gender categories, thereby shaping how individuals are perceived and judged.

2.2.3.2 Methodological Implications for NLP

A Social Philosophy of Language perspective clarifies why speaker identity and group membership should be treated as context-sensitive evidential variables rather than as fixed ground-truth attributes. As discussed by Cepollaro and Labinaz (2019), in social-network interactions the identity of the writer is often reconstructed only indirectly, through cues such as names or nicknames, profile images, profile information, writing style, observance of netiquette, or other traces of self-presentation. Such reconstruction is inherently defeasible: it neither guarantees a unique interpretation nor allows straightforward verification. Moreover, online communication may involve profiles that mask individual identity, collective or fictional accounts, and repeated processes of sharing that make it difficult to keep track of who originally produced a given utterance. In such cases, interpretive uptake often depends less on independently verifiable demographic facts than on the presumed identity made salient by the available cues and its socially relevant alignment with a group or perspective.

This point can be further connected to Haslanger's account of ideology, common ground, and social meaning (Haslanger, 2011; Haslanger, 2012). On this view, socially produced schemas and background assumptions may become so entrenched that they come to appear natural and inevitable, even though they are sustained by broader social relations and practices. Linguistic interpretation, therefore, often relies on defeasible inferences shaped by sedimented stereotypes and implicit expectations, rather than on direct access to stable facts about individuals.

For NLP research, this has several methodological consequences.

First, identity-relevant variables should not be modelled as transparent or self-evident inputs, but as context-dependent and socially mediated interpretive cues. Recognising this evidential fragility is crucial for avoiding the

naturalisation of social categories and of the expectations attached to them, both in theoretical analysis and in computational modelling.

Second, annotation schemas must acknowledge that some interpretive variables are uncertain, inferred, or contested. Disagreement among annotators may reflect genuine differences in warranted assumptions about identity and context, rather than a lack of competence. As a consequence, dataset documentation should make explicit which contextual and identity-relevant cues are available, which are assumed, and which are deliberately excluded for ethical reasons.

In sum, adopting a Social Philosophy of Language perspective allows this thesis to articulate a principled account of why meaning, identity, and interpretation cannot be clearly separated in socially consequential domains. By foregrounding the role of identity, gender, context, and stereotypes in linguistic interpretation, this section provides a conceptual foundation for the methodological choices developed in subsequent chapters, particularly with respect to annotation design and the treatment of ambiguity and disagreement in computational models.

2.3 Considerations on Methods

In our work, we adopted two well-established techniques which are typically used in social sciences, in particular in Sociolinguistics, i.e. focus group and survey. Nonetheless, while Sociolinguistics traditionally favours naturally occurring language data, ethical constraints surrounding consent and data protection often require the adoption of elicitation methods that balance ethical responsibility with analytical validity.

In this respect, focus groups represent a well-established compromise, allowing researchers to observe relatively spontaneous interaction while maintaining transparency and informed consent. As Motschenbacher (2019) observes, such methods can be especially effective once participants have undergone a degree of routinisation with the research setting.

From an ethical perspective, the use of surveys also requires careful consideration of informed consent, data anonymisation, and the protection of participants' self-identification choices. This is particularly important in research

involving gender and sexuality, where categorisation practices may carry normative or exclusionary effects. Making participants' responses and positionalities central to the analytical process aligns with the reflexive methodological stance advocated in contemporary LGS research.

In the next subsections, the two techniques are presented.

2.3.1 The Focus Group

The focus group is a semi-structured group interview, which functions as a data collection procedure that brings together people to answer input questions facilitated by a research team, usually consisting of a facilitator and observers. The focus group provides direct access to language and concepts, facilitates the collective construction of meaning, and gives researchers a greater understanding of how a community structures and organises its social world. In our perspective, this method allows the involvement of people from the target community, as well as other groups, and leads to record the possible use of language strategies and to gather structured opinions on the topic under investigation. From group interactions, a number of fixed points, thoughts, or doubts may emerge which can be investigated on a large scale by means of a survey.

Focus groups are widely used in Sociolinguistics and social research as a means of eliciting interactional data and shared meaning-making processes (Litosseliti, 2003; Harrington et al., 2008; Frisina, 2010). Their value lies not only in the collection of individual opinions, but also in the observation of how participants negotiate meanings collectively, making them particularly suitable for exploring sensitive or contested linguistic phenomena. In studies concerned with language, identity, and power, focus groups allow researchers to capture emergent discourses that may not surface in individual interviews or surveys.

From a sociolinguistic standpoint, focus groups constitute a form of elicited language data that nonetheless allows for the observation of interactional dynamics and meaning negotiation among participants. Accordingly to Motschenbacher (2019), LGS research prioritises actual language use over abstract descriptions of the language system, viewing linguistic behaviour

in relation to language users and their social positioning. Focus group discussions are therefore particularly valuable for examining how identities, categories, and norms are jointly constructed in situated interaction.

Methodologically, focus groups are widely used in research on language, identity, and power precisely because they make visible how meanings are negotiated collectively rather than individually (Harrington et al., 2008). They enable the analysis of alignment, disagreement, categorisation practices, and the emergence of shared or contested interpretations. As emphasised by Litosseliti (2003), focus groups are particularly suitable for exploring sensitive or socially regulated topics, as the group setting can encourage reflection, mutual clarification, and the articulation of perspectives that may not surface in one-to-one interviews.

2.3.2 The Survey

A sociolinguistic survey is a structured research method designed to engage a representative segment of the population (Milroy and Gordon, 2008; Schleef, 2013). It allows for the involvement of a large number of participants, particularly through the use of online tools such as Google Forms, which facilitate the rapid and broad distribution of questionnaires. This approach makes it possible to collect diverse perspectives in a relatively short time, initiating a deeper reflection on the sociolinguistic variability of the phenomenon under investigation.

In the context of our study, the sociolinguistic survey serves as a strategic tool to activate respondents' metalinguistic awareness while gathering rich socio-demographic, perceptual-linguistic, and self-identification data. This method supports the collection of both quantitative and qualitative insights, without neglecting the participants' own interpretive frameworks and self-positioning. In addition, the survey can offer confirmation of specific linguistic usages, while providing data on their contexts and motivations and capturing explicit attitudes regarding their perceived acceptability.

Within LGS research, survey-based methods have also played a central role in documenting patterns of linguistic behaviour across social groups, particularly where naturally occurring data are difficult to obtain or ethically problematic. As outlined by Motschenbacher (2019), such methods are commonly used in perception studies and attitudinal research, enabling scholars

to investigate how linguistic forms are interpreted and socially indexed by language users.

Our decision to employ a sociolinguistic survey is guided by both broad conceptual considerations and specific practical aims. Methodologically, the validity of our approach is closely linked to the issue of positionality, which forms a key epistemological foundation of our work. A thorough understanding of the patterns and trends that emerge from sociolinguistic data empowers researchers to make more intentional and ethically sound decisions. This reflective process helps identify and mitigate potential biases—conscious or unconscious—that might influence subsequent phases of computational modelling. Articulating one's positionality is therefore essential, as it shapes every aspect of the research, from design and data interpretation to final conclusions. It also helps move the research beyond default assumptions or intuitive heuristics, which may compromise the integrity of NLP systems.

From a more targeted perspective, the sociolinguistic survey plays a dual role in shaping the next stages of the project. First, it helps identify which data sources are most relevant and appropriate, based on participants' responses and the meanings they attribute to the phenomenon. Second, by revealing attitudinal and perceptual differences across demographic groups, it informs the eventual selection of a more representative and socially attuned group of annotators. This contributes to greater reliability, inclusivity, and contextual sensitivity in the data annotation process—ultimately reinforcing the quality and fairness of the overall research.

Chapter 3

Gender Stereotypes and Slur Reclamation: A NLP-oriented Literature Review

This chapter reviews the literature on the two focal phenomena of this thesis—gender stereotypes and LGBTQ+ slur reclamation—with the aim of clarifying their conceptual foundations and mapping how they have been addressed within Linguistics, especially NLP. The discussion is organised to support the research questions introduced in Chapter 1, with particular attention to the interpretative and socially situated nature of these phenomena, and to the implications this has for data collection and annotation.

The chapter is divided into two sections. Section 3.1 summarises the state-of-the-art of research on gender bias and stereotypes in NLP, and reviews computational approaches to automatic stereotype detection, while outlining how gender stereotypes relate to sexism and hate speech. Section 3.2 then introduces slur reclamation from a philosophy-of-language perspective and discusses why reclamation poses a persistent challenge for automatic detection, especially in multilingual and community-sensitive settings.

Throughout, we emphasise how advances in modelling must be accompanied by careful operational definitions and annotation practices that account for context, identity, and power relations.

3.1 UnbIAsing Gender

Despite gender being widely discussed in many disciplines—such as in Linguistics—as a complex and socially constructed concept (as discussed in

Chapter 2), much of NLP research has traditionally treated gender as a binary variable (Sun et al., 2019; Stańczak and Augenstein, 2021). Practically, NLP systems often rely on grammatical, referential and lexical categories that do not always align with social understandings of gender, which may result in the under-representation of non-binary identities and contribute to issues such as misgendering, erasure, and the reproduction of gender stereotypes (Stańczak and Augenstein, 2021). Recognising these limitations is therefore an important step towards the development of more inclusive and responsible NLP methodologies.

3.1.1 Bye Bye Bias: Gender-Fairness in NLP

Research on gender in NLP has, so far, largely concentrated on identifying and mitigating gender bias in automatic systems. As Costa-jussà (2019) highlights, studies on gender bias in NLP serve a dual role. On one hand, NLP can be used as a tool to identify gender bias in various societal domains such as news media and advertisements. On the other hand, it often generates gender-biased outputs, thereby reinforcing and perpetuating existing societal stereotypes. This is largely due to NLP systems being trained on biased datasets, which results in the amplification of bias through the learning algorithms.

Multiple studies have examined this issue from different angles. For instance, Rescigno et al. (2020) investigated how three widely used machine translation systems—Google Translate, Bing Microsoft Translator, and DeepL—handle gendered language. Their analysis of translations from English into Italian, French and Spanish revealed varying degrees of bias: Google Translate exhibited the highest level of gender bias, Bing showed a lesser degree, and DeepL was generally more gender-neutral.

Some scholars (Costa-jussà, 2019; Sun et al., 2019; Stańczak and Augenstein, 2021) have conducted comprehensive literature reviews on the topic, exploring strategies to mitigate gender bias in NLP systems. These include dataset preprocessing, algorithmic interventions, and post-processing techniques. Their research emphasises the need to design algorithms that produce equitable and inclusive outputs, thereby promoting fairness in NLP technologies from the outset.

Despite these efforts, significant challenges remain. Stańczak and Augenstein (2021) identify four key limitations in current research. First, social gender is frequently conceptualised as a binary variable, ignoring its inherent fluidity. Second, most studies focus on high-resource languages such as English, often neglecting under-represented languages. Third, although the number of papers addressing gender bias is growing, many new algorithms are still insufficiently tested for bias and overlook ethical implications. Finally, existing methodologies often lack clear definitions of gender bias and are supported by limited evaluation baselines and pipelines.

Some other works focus on more specific aspects, such as the functioning of gender bias in word embeddings (Bolukbasi et al., 2016), disparities in co-reference resolution systems (Zhao et al., 2018), and—more related to our work—the incorporation of stereotype awareness into hate speech detection models (Sanguinetti et al., 2020).

Together, these studies underscore the complexity of gender bias in NLP and the pressing need for more inclusive, representative, and ethically grounded approaches to model development and evaluation. However, as noted by Chiril, Benamara, and Moriceau (2021), the role of gender stereotype detection as a feature or input in sexist hate speech classification remains under-investigated.

Recent work on Gender-Based Violence (GBV) in online discourse further highlights the limits of approaches that focus exclusively on bias-mitigation at the system level. In particular, Ferrando et al. (2024) analyse YouTube comments reacting to the femicide of Carol Maltesi, proposing a fine-grained annotation schema that captures a range of discursive dimensions beyond overt misogyny, including empathy, aggressiveness, and responsibility attribution. Their analysis shows that gendered harm frequently emerges through implicit evaluations, victim-blaming narratives, and stereotype-driven interpretations rather than through explicitly abusive language alone.

Building on this work, Pasini et al. (2025) extend the analysis by comparatively examining reactions to the cases of Carol Maltesi and Giulia Cecchettin. In doing so, they introduce an annotation task based on a fine-grained schema that adapts and refines the categories proposed in Ferrando et al. (2024). Their findings further suggest that patterns of empathy, blame, and moral evaluation are shaped by socially shared assumptions about gender roles and

respectability.

Taken together, these studies point to the relevance of analysing how gendered meanings are structured through implicit and evaluative discourse practices. By foregrounding the role of interpretation and socially shared assumptions, they invite a closer examination of stereotypes as a constitutive element of harmful and discriminatory language use.

3.1.2 Gender Stereotypes, Sexism, and Hate Speech

Stereotypes are ubiquitous phenomenon in human communication. We naturally create them to support our cognitive limits in dealing with people that collocate out of our social group or are simply different from us under some respect. They are often expressed in implicit and subtle form that makes them very hard to automatically detect (Hovy, 2015; Sap et al., 2019; Schmeisser-Nieto, Nofre, and Taulé, 2022; Schmeisser-Nieto et al., 2024a; Cignarella, Giachanou, and Lefever, 2025).

As outlined before, stereotypes play a central role in the emergence of sexism and hate speech, particularly through the way they generalise and attribute traits to social groups (Chiril, Benamara, and Moriceau, 2021). Defined by Lippmann (2004) as "pictures in our heads", stereotypes shape perception by providing simplified representations of others, which in turn influence attitudes and behaviours. They function both descriptively—indicating what group members are like—and explanatorily—suggesting why they are that way (Fiske, 1998). Although not inherently negative, stereotypes frequently underpin discriminatory ideologies, reinforcing intergroup hierarchies and legitimising social biases.

Gender stereotypes—which are framed as generalised views or preconceptions about attributes or roles considered appropriate for men and women by the Office of the United Nations High Commissioner for Human Rights (2013)—have been widely studied in Psychology and Linguistics. As we mentioned above, beyond their descriptive function such stereotypes operate as socially shared expectations that regulate behaviour and evaluation, thereby contributing to the maintenance of gendered hierarchies. It is precisely in this sense that gender stereotypes constitute a central component of sexist ideology.

In fact, Fiske's work—including the ***Stereotype Content Model*** (Fiske et al., 2002; Cuddy, Fiske, and Glick, 2008) and the ***Ambivalent Sexism Inventory*** (Glick and Fiske, 1997)—conceptualises sexism as a system of beliefs grounded in stereotypical representations of women. Within this framework, a distinction is drawn between *hostile sexism*—targeting women who challenge traditional gender roles—and *benevolent sexism*, which idealises and rewards conformity to those same expectations. Both forms rely on gender stereotypes to legitimise unequal treatment and thus contribute to gender-based discrimination.

In online contexts, these stereotypes frequently surface in sexist hate speech, which, as highlighted in the *Council of Europe Gender Equality Strategy*, aims to "humiliate or objectify women, to undervalue their skills and opinions, to destroy their reputation, to make them feel vulnerable and fearful, and to control and punish them for not following a certain behaviour" (Council of Europe, 2016, p. 5).[1] In particular, gender stereotypes appear to be closely intertwined with language (Bazzanella, Fornara, and Manera, 2006; Gygax et al., 2019; Jahan and Haque, 2023). In the field of Psycholinguistics, for example, it is widely accepted that stereotypes can be activated during language processing and that they can play a role in communication (Banaji and Hardin, 1996; Maass, Arcuri, and Suitner, 2014).

3.1.2.1 Fiske's Stereotype Content Model (SCM) and Related Works

As mentioned above, a widely used framework for describing stereotype structure is Fiske's Stereotype Content Model (SCM), which characterises stereotypes along two core dimensions: *warmth* (e.g., perceived friendliness, trustworthiness) and *competence* (e.g., perceived capability, agency) (Fiske et al., 2002; Cuddy, Fiske, and Glick, 2008). Within this model, different social groups may be positioned in distinct regions of the warmth-competence space, giving rise to ambivalent stereotypes (e.g., high warmth but low competence), which are particularly relevant for understanding forms of benevolent sexism.

In the context of gender, SCM-informed accounts have been used to explain why some stereotypes can appear superficially positive while still reinforcing hierarchical expectations and constraining social roles (Glick and Fiske, 1997). This is consistent with the view that stereotypes do not need to be explicitly

[1] https://rm.coe.int/1680651592

hostile to sustain discriminatory outcomes. For the purposes of this thesis, SCM is useful as a conceptual scaffold for linking social psychological theory to the linguistic expression of stereotypes; the same *warmth-competence* dimensions may surface in discourse through evaluative adjectives, presuppositions, generics, and role descriptions, often in implicit form.

At the same time, applying SCM categories to language data requires caution. Stereotype content and its linguistic realisations are culturally situated and historically contingent; moreover, operationalising such constructs in annotation schemas can risk reifying categories if contextual and interpretative factors are ignored. This motivates the methodological emphasis of the present work on context-aware and socially grounded annotation.

3.1.2.2 Detecting Stereotypes

While the link between gender stereotypes and hate speech is well-established in Psychology (García-Sánchez et al., 2019), computational approaches have only recently begun to address this connection, as for example, in the aforementioned works from Ferrando et al. (2024) and Pasini et al. (2025).

In general, stereotypes have been addressed in various studies within the NLP field, though they are rarely the central focus, and gender stereotypes in particular are often under-explored (Cignarella, Giachanou, and Lefever, 2025). In the context of the EVALITA campaigns,[2] for example, the *Hate Speech Detection (HaSpeeDe)* task series introduced datasets annotated for the presence of stereotypes, particularly those targeting immigrants. In the 2018 edition of the campaign (Bosco et al., 2018), participants were asked to detect whether such stereotypes were present in text messages, marking one of the few instances where stereotype recognition was an explicit task.

A further relevant Italian contribution is provided by Cignarella et al. (2024), who introduce the *QUEEREOTYPES* corpus, a social media dataset annotated for stereotypes targeting LGBTQIA+ people, together with related dimensions such as hate speech, offensiveness, aggressiveness, irony, and stance. Besides extending stereotype detection to a still under-explored target in Italian NLP, this work is also methodologically relevant because the

[2]EVALITA is the main evaluation campaign for NLP and speech technologies for Italian. It regularly organises common evaluation exercises (*shared tasks*) in which different research teams address the same language-related task on a shared dataset, making it possible to compare computational methods and results in a systematic way.

resource was developed through collaboration between NLP researchers and LGBTQIA+ activists.

Other works have further explored stereotype detection, primarily in relation to immigrant and racialised groups (Francesconi et al., 2019; Sanguinetti et al., 2020), rather than focusing on gender. A significant contribution in this direction is the study by Bosco et al. (2023), which presents an Italian social media corpus annotated with racial stereotypes using a novel, fine-grained annotation schema. Their methodology emphasises how stereotypes are expressed through specific forms of discredit, grounded in real-world contexts of discrimination. This work not only supports the development of linguistic resources for stereotype detection but also highlights the importance of capturing the multifaceted nature of hateful communication.

Building on this perspective, Schmeisser-Nieto et al. (2024a)'s *StereoHoax* was built, a multilingual corpus focused on racial hoaxes and social media reactions, annotated for stereotype presence, forms of discredit, and contextual framing. Their work further illustrates how stereotypes operate as rhetorical tools for delegitimising and discrediting targeted groups, offering valuable insights into the mechanisms by which hate speech spreads across online platforms. Rather than treating stereotypes as isolated textual phenomena, their approach explicitly accounts for the conversational structure in which stereotypes occur, showing that a substantial proportion of stereotypical content can only be identified through reference to prior discourse. Importantly, the analysis presented in *StereoHoax* demonstrates that stereotype content aligned with SCM-related dimensions, such as perceived lack of warmth or competence, is frequently conveyed implicitly. Stereotypes often emerge through entailment, evaluative framing or in-group positioning, rather than through direct attribution of traits to the targeted group. This finding has methodological implications for both annotation and automatic detection, as it suggests that SCM-informed categories remain analytically useful, but require a discourse-sensitive operationalisation that takes contextual and pragmatic factors into account.

Related evidence is reported in Schmeisser-Nieto et al. (2024b), a parallel corpus-based work conducted within the same research line, where *StereoHoax-IT* (a subset of StereoHoax, containing reactions to fake news in Italian) and *SterheoSchool*, are compared, both quantitatively and qualitatively. The

second dataset contains teenagers' comments to fake news generated in psychological experiments. This work highlights a different yet complementary role of implicit and context-dependent stereotype expression. Building on the criteria originally proposed in Schmeisser-Nieto, Nofre, and Taulé (2022), the study operationalises implicitness through linguistic and pragmatic labels. The annotation task—carried by three annotators with linguistic background—consisted in two different layers. Annotators first marked whether a comment contained an anti-immigrant stereotype and, if so, assigned it to one of *ten predefined topics* (e.g., xenophobia victims, economic resource, culture and religion, security). Then, they were asked to evaluate context—through contextual information from preceding messages reported when needed—and implicitness—through *a set of non-mutually exclusive strategies and linguistic devices* that aimed at capturing how the stereotype is conveyed without being overtly stated: *world knowledge*, *figures of speech*, *irony/sarcasm*, *humor/jokes*, *extrapolation*, *imperative/exhortative*, *entailment/evaluation*, and *other*. Quantitative results showed that entailment/evaluation was the dominant implicitness strategy (64% in *StereoHoax-IT*, 80% in *SterheoSchool*), while extrapolation, imperative and irony appeared less frequently but in specific contexts. The qualitative analysis examines concrete examples to reveal how these strategies operate. For example, in their interpretation, expressions like *"Governo di involtini primavera!!!* (eng. "Spring rolls government!!!") allow an ironic reading where the use of a metonym ("spring rolls", a traditional Chinese dish) triggers a negative evaluation of Chinese people, while sentences like *"Venezia, donne velate sputano al crocifisso"* (eng. "Venice, veiled women spit on the crucifix") combine extrapolation and entailment/evaluation to encode cultural/religious bias. Further cases illustrate that imperative/exhortative forms (e.g. *"Se non fate niente Fra 10 anni l'italia sarà tutta musulmana!"*, eng. "If you do nothing in ten years Italy will be completely Muslim!") and figures of speech (questions, rhetorical devices) resemble explicit statements yet remain implicit. The authors also note frequent co-occurrence of irony/sarcasm with figures of speech and humour, but the most of implicit stereotypes rely on logical relations that require contextual inference. Inter-annotator agreement (IAA), measured with Fleiss' κ, reached moderate to high levels both across ***topics*** (0.48–0.86) and ***implicitness*** (0.45–0.86) categories, suggesting that,

although implicit stereotypes are inherently inferential, their operationalisation through structured criteria can yield acceptable reliability. At the same time, the analysis underscores that context plays a decisive role: in many cases, stereotypical meaning becomes identifiable only when the surrounding discourse is taken into account. The study concludes that detecting implicit stereotypes demands models capable of capturing logical entailments and contextual knowledge, and it suggests extending the schema to other languages.

Taken together, all these studies support a view of stereotypes as evaluative constructs—which are dynamically negotiated in discourse and cannot be reduced to surface lexical cues—rather than as fixed representations that can be reliably recovered from decontextualised text alone. They support a discourse-sensitive and multi-layered approach, in which stereotype presence, target, topical content, and implicitness strategies are analytically distinguished but operationally interconnected.

3.2 ReclAIming the Queer

The term *queer* occupies a distinctive position in the study of slur reclamation, as it is both a historically derogatory label and a paradigmatic example of successful linguistic reclamation. Originally used in English as a pejorative term to stigmatise non-normative sexualities and gender expressions, *queer* has been progressively reclaimed since the late twentieth century through political activism and academic intervention, particularly within LGBTQ+ movements and critical theory (Brontsema, 2004).

This process of reclamation has led to a semantic and pragmatic shift: in many contemporary contexts, *queer* functions as a self-chosen identity marker, a term of solidarity, or an umbrella category encompassing diverse non-heteronormative identities. Importantly, however, this shift has not resulted in a complete erasure of its derogatory potential. The term remains context-sensitive, and its interpretation continues to depend on factors such as speaker identity, audience, intention, and sociocultural setting. This makes *queer* a canonical example of what Philosophy of Language describes as a *conventionalised yet reversible* case of reclamation (Bianchi, 2014; Cepollaro and López de Sa, 2022).

Beyond its status as a reclaimed slur, as we have seen in Chapter 2 *queer* has also been adopted as a critical and methodological lens within the humanities and social sciences, as for *Queer Linguistics*. This dual role of *queer*[3]—as both an object of linguistic reclamation and a theoretical framework—is particularly relevant for NLP research. It highlights why the same lexical form may function as a trigger for hate speech, identity affirmation, humour, or political stance, depending on context and speaker positioning. Consequently, automatic detection systems that rely solely on surface forms or decontextualised labels risk systematic misclassification. For this reason, a queer-informed perspective reinforces the need for socially aware and context-sensitive annotation practices, in which meaning is understood as situated and contested rather than fixed.

3.2.1 An Operational Definition of Reclamation for the Present Study

For the purposes of this thesis, and especially in view of the case study and its implications for NLP, we adopt an operational definition of reclamation rather than attempting to resolve all the philosophical controversies surrounding it. Consistent with recent work in Philosophy of Language, we understand **reclamation as the linguistic practice whereby speakers, typically though not necessarily members of the targeted communities—reclaim terms historically used in a derogatory manner against their group and reuse them for non-derogatory purposes, such as expressing belonging and identity, fostering solidarity and camaraderie, or contesting entrenched structures of discrimination** (Bianchi, 2014; Cepollaro and López de Sa, 2022; Zsisku, Zubiaga, and Dubossarsky, 2024).

More specifically, we also draw on Edmondson (2021)'s refinement of Bianchi (2014)'s definition. Bianchi defines appropriated[4] uses as "uses by targeted groups of their own slurs for non-derogatory purposes, in order to demarcate the group, and show intimacy and solidarity" (Bianchi, 2014,

[3]Although we also use *queer* in its reclaimed sense to refer broadly to the LGBTQ+ community, this is not its primary use in the thesis. In what follows, the term is generally used in its critical and methodological sense, in connection with Queer Linguistics, unless a different meaning is explicitly specified.

[4]In this publication, Bianchi refers to the phenomenon as *appropriation*, although she also acknowledges the broader terminological variability surrounding it.

p. 35). Edmondson explicitly takes up this formulation, but argues that the phrase "their own slurs" is potentially misleading, since it may suggest that the targeted group somehow possessed or controlled the term prior to its reclamation. He therefore reformulates linguistic reclamation as **"non-derogatory use, by members of marginalised groups, of the slurs used by members of more hegemonically powerful groups to derogate them"** (Edmondson (2021, p. 194)). This refinement is especially useful here because it makes the asymmetry of power fully explicit: slurs are not treated simply as offensive labels, but as expressions embedded in broader relations of hegemony and marginalisation. In this connection, Edmondson explicitly invokes the Gramscian notion of *hegemony* and recalls Baker (2008)'s formulation of it as "the exercise of power, whereby everyone acquiesces in one way or another to a dominant person or social group". He also stresses that reclamation should be understood as a politically motivated process, and that even within targeted groups it is not practised uniformly: not all members engage in reclamation projects equally, and the decision to engage or not may depend on personal and political reasons. For the present study, this definition is particularly helpful because it directs attention to speaker position, target group, and social power as analytically relevant variables for the interpretation of reclamatory uses and, later, for the modelling and annotation of socially sensitive language in NLP. Against this background, we briefly outline three conceptual points that frame the analysis and support its later methodological translation into NLP.

First, we take reclamation to be a heterogeneous phenomenon encompassing both highly conventionalised uses (for instance, the reclamatory use of *queer*, whose non-derogatory meaning has become widely established) and novel, less conventionalised uses that have yet to gain widespread acceptance. Moreover, reclamation arises across diverse contexts: from explicitly activist settings, where speakers are acutely aware of the political implications of their linguistic choices, to more informal, intimate interactions in which slurs may be employed playfully or affectionately (Bianchi, 2014). In this respect, Bianchi (2015)'s distinction between *amicable contexts*, in which members of the target group use a slur non-offensively to express solidarity and intimacy, and more explicit contexts of *reclamation proper*, in which the reuse of the slur becomes part of a conscious political practice, is especially helpful. Within

this type of context, Bianchi also includes uses by artists—writers, poets, songwriters, playwrights, and comedians—who reclaim derogatory epithets as a way of subverting dominant socio-cultural norms. For all these cases, then, in her echoic account of reclamatory uses Bianchi proposes the broader label of *community uses* to refer, more generally, to non-offensive uses by members of the target group without sharply separating intimate and explicitly political cases. This more inclusive category is useful here because it captures the fact that reclamation may range from local, relational—and not necessarily self-consciously political—interactions to overtly activist and movement-based practices. As briefly outlined in Section 2.2.3, Bianchi's echoic account is best understood as a theoretically useful proposal for explaining certain community uses: on this view, speakers echo derogatory uses—or the discriminatory perspective associated with them—while making clear their dissociation from that perspective. In such cases, reclamation erases the offensive background of the slur by evoking such a background and, contextually, subverting it, through mockery and re-signification. This account is particularly apt for non-fully conventionalised cases in which the offensive background remains active, whereas more stabilised cases such as *queer* or *gay* in English may no longer display a clearly echoic effect (Bianchi, 2015), as their original meaning could be considered both semantically and pragmatically obliterated[5]. This heterogeneity is methodologically relevant for the present study, since it cautions against treating reclamation as a single, context-independent class of uses.

Second, although discussions of reclamation are often framed in terms of social groups[6]—specifically ingroup versus outgroup speakers—this framing risks oversimplification. Social groups are not monolithic entities, as members' experiences and interests are shaped by individual psychological differences as well as by intersecting axes of identity. As Crenshaw (1989)'s seminal work on intersectionality makes clear, dimensions such as gender, ethnicity, socio-economic status, and education intersect to produce distinctive forms of identity and oppression. Consequently, experiences of, for example, being

[5]However, this affirmation should be treated with some caution, as it does not derive from recent data neither it was investigated in the present study. It is based rather on our experience of interaction with speakers across different contexts, and, in the case of *queer*, on the fact that, as we have seen at the beginning of this Section, the term has long been conventionalised in academic usage.

[6]Which, in turn, can be seen as communities of practice.

gay and encountering homophobia differ significantly across factors such as gender, class, and education (Lykke, 2010; Esposito, Pérez-Arredondo, and Zottola, 2024). Thus, speaking of "the LGBTQ+ community" as a single, homogeneous entity is necessarily reductive. In this work, we adopt such terminology for the sake of clarity and simplicity, while also explicitly applying an intersectional lens to our analysis of how participants' gender, sexual orientation, age, and other factors shape their perspectives on reclamation. This point is fully compatible with Edmondson (2021)'s insistence that reclamation is unevenly distributed within targeted groups and should be understood as a collective consequence of individual linguistic choices made with varying degrees of uptake and success.

Third, prevailing views in both scholarly and non-scholarly contexts typically assume *targetism*—the idea that only ingroup members (i.e., members of the targeted communities) may legitimately reclaim slurs. In our study, rather than presupposing targetism, we investigate how participants evaluate the role of speaker identity in shaping perceptions of reclamatory uses. Following Cepollaro and López de Sa (2022), we distinguish between *belonging to a target community* (e.g., being gay and thus belonging to the LGBTQ+ community) and *being appropriately related to it* (e.g., an ally who advocates for LGBTQ+ rights, and thereby stands in a relation of solidarity without membership). This point is especially relevant for NLP, since it suggests that speaker identity should not be modelled as a simple binary input, but approached instead as a socially mediated and context-dependent dimension of interpretation.

3.2.2 Language-specific Trajectories of Reclamation

Beyond the paradigmatic case of *queer*, reclamation processes display language-specific trajectories that reflect the morphological and semantic resources of individual languages. In particular, Italian provides several examples of reclaimed LGBTQ+ slurs whose development cannot be fully captured by direct reference to English models. As noted by Nossem (2019) and Pepponi (2024), each reclaimed term is embedded in a distinct historical, semantic, and sociocultural trajectory and cannot be adequately analysed independently of the language in which it circulates.

From this perspective, reclaimed slurs do not simply travel across languages as stable or equivalent labels, but are re-created through processes

of localisation and resemanticisation that reflect the linguistic resources and ideological tensions of the receiving context. This language-specific dynamic is central to the present thesis, whose empirical focus lies on Italian, and provides a useful background for understanding the contextual variability and interpretative instability of reclamatory uses.

Within the Italian linguistic space, this dynamic is particularly visible in the emergence of locally reclaimed terms such as *frocio* and *frocia*, which co-exist with the borrowed and partially domesticated use of *queer*. Building on Nossem (2019)'s notion of *(g)localisation*, Pepponi (2024) highlights how *frocia* can be analysed as a morphologically and semantically innovative form derived from the masculine slur *frocio*, functioning both as a morphological gender motion and as an explicitly political act of reclamation. Crucially, this morphological transformation is not merely interpreted as an alteration of the form of the slur, but seems to introduce new pragmatic possibilities, as grammatical gender itself becomes a resource for positioning, stance-taking, and subversion. Importantly, Pepponi distinguishes between adjectival and substantival uses of *frocia*, showing how this term may operate simultaneously as an identity flag and as a discursive stance, thereby expanding its pragmatic potential beyond that of its source form. In particular, the same lexical item may retain a derogatory value in certain adjectival uses while functioning as a self-affirming marker in substantival contexts, underscoring the role of syntactic and discursive configuration in shaping interpretation. In this sense, *frocia* exemplifies how grammatical gender and morphological creativity may become resources for subversion rather than mere reflections of social categorisation. As noted by Nossem (2019), this gender-marked reconfiguration opens up subversive potentials that are not directly available in the English term *queer*, whose lack of grammatical gender constrains comparable forms of morphological play.

Taken together, Nossem's lexicological analysis and Pepponi's account of the contemporary Italian LGBTQ+ lexicon foreground a crucial point for the present thesis: even when lexical history is not the primary object of investigation, attending to the semantic and morphological stratification of reclaimed slurs is essential for understanding their contextual variability, perception, interpretation, and political force within Italian language use.

3.2.3 The Challenge of Slurs for NLP

Building on the conceptual background presented in Chapter 2, in recent years research in language technologies has increasingly concentrated on the fine-grained dimensions of linguistic expression, including phenomena such as irony (Casola et al., 2024), sarcasm (Frenda et al., 2022), and hate speech (Nozza et al., 2023; Malik et al., 2024). The imperative to identify and mitigate the dissemination of offensive discourse directed at specific social groups—on the basis of gender, sex, sexual orientation, race, religion, language, or political affiliation—has become particularly pressing (Guillén-Pacho et al., 2024; Cuccarini et al., 2024). Nevertheless, the task of delineating hate speech through rigid categorical boundaries remains problematic. Contemporary scholarship in NLP underscores the necessity of accounting for contextual factors to reduce the risk of semantic misinterpretation (Pamungkas, Basile, and Patti, 2020; Pamungkas, Basile, and Patti, 2023), emphasising, for instance, that a given swear word may function either as an abusive or as a non-abusive expression depending on its situational usage.

The analysis of hate speech becomes even more intricate when considering the phenomenon of slur reclamation. As we have partially reported in the previous sections, this practice has been the subject of extensive scholarly inquiry across multiple disciplines, including Philosophy of Language (Bianchi, 2014; Anderson and Barnes, 2022; Jeshion, 2020; Cepollaro and López de Sa, 2023), Linguistics (Brontsema, 2004) and Sociolinguistics (Mutlak, 2024), as well as Experimental Psychology (Galinsky et al., 2013; Bianchi et al., 2024).

Within Computational Linguistics, and particularly in NLP, the phenomenon of slur reclamation remains markedly under-explored, with an even greater paucity of studies in the Italian context. While recent advances in AI have primarily focused on the construction of corpora and the development of models for detecting the abusive use of derogatory expressions in social media discourse (Fortuna and Nunes, 2018; Poletto et al., 2021; Jahan and Oussalah, 2023), research has largely framed hate speech as a classification task centred on overtly abusive language forms (Davidson et al., 2017). This orientation is further reflected in the widespread use of lexicon-based resources for hate speech detection, such as *HurtLex* (Bassignana, Basile, and Patti, 2018), a multilingual lexicon originally developed from an Italian resource and designed to support the automatic identification of inherently derogatory or

offensive lexical items. While *HurtLex* acknowledges that the interpretation of such terms may depend on context, its design and primary use remain focused on detecting abusive language rather than modelling socially situated or reclamatory uses of slurs. Within this dominant paradigm, reclaimed or identity-affirming uses of derogatory expressions tend to be treated as noise or false positives rather than as phenomena requiring dedicated modelling. As a result, research explicitly addressing reclamatory uses is still scarce (Zsisku, Zubiaga, and Dubossarsky, 2024).

To date, and to the best of our knowledge, only Cuccarini et al. (2024) have specifically investigated the Italian case in NLP. Their exploratory work—more widely described in Chapter 7—is illustrative for the central question that reclamatory language detection poses to NLP, i.e. the need to prevent the erroneous censorship of legally permissible speech in AI-supported moderation systems. This is a risk that could disproportionately affect activists and marginalised communities, thereby constraining rather than expanding their opportunities for representation and, ultimately, limiting their freedom of expression (Cepollaro and Labinaz, 2019). Against this scenario, a pressing challenge concerns the development of equitable datasets suitable for model fine—tuning, as well as the pursuit of a balance between the detection of hate speech and the safeguarding of free expression and the open circulation of ideas.

Chapter 4

Junction Chapter: The FATA Paradigm

Given the methodological background and the literature review, presented the methodological environment and the line of research within which this PhD thesis was developed, and highlighted the objectives of our research, it is now necessary to clarify how our research questions harmonise with this background and what the case studies presented in the next chapter may tell us.

This chapter functions as a junction between the theoretical background and premises developed so far and the empirical case studies presented in the next chapters. Rather than introducing new empirical material, its purpose is to make explicit the logic that connects research questions, theoretical commitments, and methodological choices, clarifying how the case studies are designed to contribute to answering the research questions introduced in Chapter 1.

This chapter aims to foreground the epistemological and practical assumptions guiding the present work. In particular, it articulates why an interdisciplinary, queer-informed, and community-centred approach is not merely compatible with NLP research on harmful language, but necessary to address phenomena such as stereotypes and slur reclamation in a principled and sustainable way.

In doing so, this chapter does not anticipate the technical details of the case studies, which are discussed in later chapters. Instead, it clarifies what kinds of insights the case studies are expected to yield, how they are anchored in the literature reviewed so far, and which dimensions of the phenomena under investigation they are designed to make visible.

4.1 Applying a Positioned, Queer Perspective

Applying a queer and positioned perspective to our work means treating theory not as an external interpretative layer, but as a set of constraints that shape research design. As discussed in Chapter 2, positionality and reflexivity are not add-ons to the annotation process, but conditions for its validity, particularly when working with socially sensitive language.

Within this framework, the case studies are conceived as situated investigations rather than neutral tests of pre-defined hypotheses. They are designed to surface tensions and ambiguities in interpretation, rather than to eliminate them through forced consensus. This orientation reflects a QL stance that treats instability and contestation as analytically meaningful, especially in relation to gendered and reclaimed language.

Crucially, adopting a positioned perspective also entails making visible the limits of one's own analytical standpoint. The present work treats the case studies as partial but principled explorations, whose value lies in their capacity to inform more socially aware annotation practices and to raise new questions for future research.

Following the analytic feminist framework introduced in Chapter 2 and after revising the limitations of NLP works in Chapter 3, in the present work we treat gender as a socially meaningful parameter that shapes interpretation and is itself shaped by norms and expectations. This view aligns with performative and display-oriented approaches, which understand gender as something speakers *do* in context rather than something they *are* in an essentialist sense (Butler, 1990). This perspective is particularly important for understanding and trying to frame gender stereotypes, which—as we have seen—frequently operate through implicit meaning, pragmatic inference, power relations, and culturally shared assumptions.

Operationally, this thesis distinguishes between (i) self-identification when participants explicitly report their gender identity (e.g., in surveys), and (ii) perceived or inferred gender when interpretation is performed by readers or annotators on the basis of linguistic and contextual cues. This distinction is crucial for our later annotation design: rather than assuming that gender is always available as a ground truth, we make explicit when it is reported, when it is inferred, and when it is unavailable or contested. In turn, this supports the inclusion of sociolinguistic variables—such as gender-role self-concept—

as factors that may help explain systematic differences in interpretative judgements within the annotation team or among survey respondents.

4.2 Interdisciplinarity and Methodological Sustainability

In the context of this thesis, interdisciplinarity is not conceived as the mere juxtaposition of insights from different fields, but as a methodological necessity arising from the nature of the phenomena under investigation. Gender stereotypes and LGBTQ+ slur reclamation cannot be adequately analysed within the confines of a single disciplinary framework: they are simultaneously linguistic, social, psychological, and political phenomena.

From this perspective, sustainability refers not only to computational scalability, but also to the epistemic and ethical sustainability of research practices. Approaches that prioritise efficiency or performance at the expense of interpretability or social awareness risk producing models that could be brittle, as they result unfair or misaligned with the realities of language use. The interdisciplinary framework adopted in this thesis—integrating NLP with Sociolinguistics, Philosophy of language, and QL—aims to counteract these risks by ensuring that methodological choices remain accountable to the social dimensions of language.

Importantly, this stance does not reject computational abstraction, but seeks to regulate it. Abstraction is treated as a necessary step in modelling, yet one that must be guided by theoretically informed decisions about which aspects of social meaning can be simplified and which cannot. The case studies presented later in the thesis are designed with this balance in mind; they explore how interdisciplinary insights can be operationalised without collapsing complex phenomena into reductive categories.

4.3 Community-centred and Participatory Praxes for a Fair Automatic Detection

The commitment to community-centred and participatory praxes constitutes a further unifying thread between theory and empirical work in this thesis.

As argued in the methodological chapter, communities affected by harmful language are not only data sources, but stakeholders whose perspectives are essential to understanding how language is interpreted and experienced.

In practical terms, this commitment informs how data are selected, how annotation schemas are designed, how disagreement and instability are treated, and how this information is then categorised. Rather than aiming to suppress variation in judgement, the present work treats such variation as an empirical signal that reflects differing social positions and lived experiences. This is particularly relevant for phenomena such as slur reclamation, where legitimacy and offensiveness cannot be determined independently of community norms.

The case studies that follow operationalise these principles to varying degrees, depending on their specific objectives and constraints. While not all forms of participation are feasible in every research setting, the overarching aim is to generate a fair automatisation process, that may definitely move beyond extractive models of data collection toward practices that recognise the epistemic contribution of participants and annotators. Throughout this work, we seek to attend closely to what participants and annotators themselves make relevant, treating as analytically significant the aspects that emerge from their choices, feedback, suggestions, and positions, which allow us to move beyond our own blind spots.

4.4 What Do We Learn About Our Chosen Topics?

This section clarifies the specific contribution that the following empirical chapters make to the study of *nomina agentis*,[1] gender stereotypes and LGBTQ+ slur reclamation. Rather than producing general claims about the use of gendered nomina, the prevalence of stereotypes or the frequency of reclamatory uses, and instead of providing clear-cut and ready-to-go categories for automatic detection, the case studies examine how these phenomena are interpreted, negotiated, analysed, and operationalised in concrete research settings. They show what emerges when socially sensitive language is approached as a site

[1] Here to be understood as agent nouns used to designate social actors through specific lexical or morphological forms, and which constitute the core of the preliminary study presented in the following chapter.

of situated interpretation and methodological choice. In particular, they analyse how annotation practices, context, and participant perspectives shape what counts as harmful, acceptable, taboo, or ambiguous language. They also identify the conditions under which particular distinctions become analytically relevant or unstable, especially when meaning depends on cues such as speaker identity, group belonging, social positioning, and community-specific norms.

By approaching stereotypes and reclamation as interpretative challenges rather than fixed categories, the present work seeks to contribute a methodological rather than purely descriptive form of knowledge. In this sense, the empirical chapters function as sites in which theoretical assumptions and research procedures are tested against actual interpretative practice. The value of the case studies thus lies in their capacity to inform the design of socially aware NLP resources and to highlight the trade-offs involved in modelling language that is deeply embedded in social life. More specifically, they bring into focus the tension between the operational demands of computational research and the interpretative complexity made visible by participants, annotators, and researchers themselves, relatively to the contexts of use.

4.5 The FATA paradigm

To address the challenges outlined above, this thesis draws on the **First Ask Then Act (FATA)** paradigm, a *proposed* community-centred and inclusive approach to the study of potentially harmful language in NLP. The paradigm has been presented in the form of a position paper by Ferrando et al. (2025), and has been developed and operationalised within the broader research trajectory of this doctoral work.

FATA advocates a methodological reorientation of the standard NLP pipeline by foregrounding the role of affected individuals and communities in the early stages of research design, data collection, and annotation. Rather than treating socially sensitive phenomena as objects that can be exhaustively defined *a priori*, the paradigm starts from the assumption that perceptions of harm and acceptability, are socially situated and unevenly distributed across speakers and audiences. For this reason, FATA proposes to *first ask* communities and individuals how specific linguistic practices are perceived and

interpreted, before *acting* through large-scale data collection, annotation, and, eventually, language model development. This sequential logic is explicitly reflected in the integration of established sociolinguistic methods—such as focus groups and surveys—into the NLP workflow (Ferrando et al., 2025).

The theoretical foundations of FATA are grounded in the interdisciplinary traditions that challenge essentialist and decontextualised accounts of language. In particular, insights from QL provide a critical lens through which language is analysed as a site of power negotiation, identity construction, and social positioning (Motschenbacher, 2010; Motschenbacher, 2018; Crenshaw, 1989; Esposito, Pérez-Arredondo, and Zottola, 2024). Within this perspective, linguistic forms are not treated as inherently harmful or benign; rather, their social meaning emerges from situated practices of use and interpretation.

Consistent with this view, positionality plays a central role within the FATA paradigm. Following Yip (2024), positionality is understood as researchers' explicit awareness of how their identities, social locations shape their perspectives and assumptions and, therefore, every stage of the research process. Instead of assuming neutrality, FATA promotes reflexive practices aimed at making interpretative frameworks visible and accountable, and at mitigating forms of epistemic harm (Field et al., 2021). This reflexive orientation informs both the design of sociolinguistic instruments and the subsequent phases of annotation and analysis.

Within this framework, slur reclamation constitutes a particularly revealing phenomenon. The survey-based study reported in Marra et al. (2026) represents an initial attempt to operationalise the FATA paradigm by foregrounding participants' interpretative judgements prior to any computational modelling. The questionnaire presented in that work is the same survey instrument discussed in the present thesis, where it is analysed in a more comprehensive and methodologically integrated manner as part of our second case study, in Chapter 7.

In Marra et al. (2026), the survey is primarily used to articulate and empirically inform the philosophical and sociolinguistic assumptions underpinning our analysis of slur reclamation, with particular attention to the role of speaker identity, social positioning, and contextual inference in shaping perceptions of acceptability and harm. The present thesis builds on that application by embedding the same survey within a broader research design, in which not only

the survey itself, but also the wider initial process of annotation, is discussed.

In our case studies—addressed in Chapter 6 and in Chapter 7—both gender stereotypes and LGBTQ+ slur reclamation are examined in different *Ask* phases in an attempt to pave the way for a following *Act* phase. They are explicitly connected to annotation practices[2]—while maintaining a qualitative focus on the meta-reflexive feedback provided by both annotators and the communities involved.

Altogether, both gender stereotypes and LGBTQ+ slur reclamation reveal a shared structural tension that is central to the present thesis. In both cases, potential harm does not reside in specific lexical items or forms *per se*, but emerges from socially situated processes of interpretation that mainly depend on context and social positioning. As a result, both phenomena resist reduction to surface-level detection strategies and challenge annotation practices that assume stable, context-independent labels.

From a computational perspective, stereotypes and reclamation therefore raise parallel methodological issues. In the case of stereotypes, implicitness and culturally shared presuppositions complicate automatic detection; in the case of reclamation, the same lexical form may alternate between derogatory and non-derogatory uses depending on who speaks and how an utterance is framed; similarly, in the case of nomina agentis—specifically discussed both in our Preliminary Study (Chapter 5) and in Case Study 1 (Chapter 6)—the choice of a particular form may either convey derogatory intentions or, conversely, function as a marker of identity and self-positioning, through alignment or contestation. In all these cases, current NLP approaches tend to prioritise classification performance while underestimating the interpretative labour required to assign labels in socially sensitive contexts.

Finally, what connects these otherwise different phenomena is their shared place within the strand of our research, as outlined in Section 1.1, which addresses the challenges of automatic detection of harmful language from a socially situated perspective. From this point of view, gender stereotypes, LGBTQ+ slur reclamation, and the socially meaningful use of nomina agentis are approached not as separate issues, but as complementary sites from where we can examine the limits of standard hate speech and dangerous

[2]Case Study 1 describes an actual annotation task, while Case Study 2 presents an initial step of the NLP pipeline.

speech detection frameworks and experiment with more socially aware and participatory annotation practices.

The preliminary study presented in the next chapter builds on this insight by grounding these considerations in the specific linguistic and sociocultural context of Italian, providing both a conceptual and an empirical background necessary for the case studies that follow.

Chapter 5

Gender and Language in the Italian Context: A Preliminary Study

This chapter is mainly based on a published work led by the present writer (Marra and Bosco, 2024). In Section 5.1 we outline the linguistic and socio-cultural background necessary to understand how gender is encoded and negotiated in the Italian language, particularly focusing on the role of *over-extended masculine* and the long-standing debate on *nomina agentis*. Section 5.2 presents the preliminary study reported in Marra and Bosco (2024), outlining our data, annotation procedure, and key observations.

Beyond its descriptive aim, the chapter also fulfils a methodological function within the thesis. By showing how gendered forms in Italian are shaped by social norms and ideological assumptions—which may in turn produce and reinforce gender stereotypes—this perspective highlights why the analysis of gender-fair language cannot rely on purely formal or morphosyntactic criteria. Instead, it motivates the need to treat annotation and interpretation as socially situated practices, a premise that directly informs the design of the case studies discussed in the subsequent chapters.

5.1 Gender and Language in Italian: A 40-year Research Background

For almost forty years now, there has been an ongoing debate in Italy—which has intensified over the last decade—on the over-extended use of masculine forms for roles and positions held by women. Changes in Italian society have gradually led to a greater presence of women in many work and professional

roles that were previously closed to them and to a growing and specific focus on the debate on nomina agentis, with particular attention to certain job titles and their masculine and feminine variants (Formato, 2016; Formato, 2019). The feminine variants are grammatically provided for in the Italian language and attested in a diachronic perspective and in various text genres; however, they still seem to encounter resistance to use, which is deeply rooted in a social fabric in which men and women share and perpetuate gender stereotypes (Formato, 2016; Formato, 2019; Azzalini and Giusti, 2019).

As explained by Formato (2016, p. 98):

> [...] there is room to argue that Italian is a gender-fair language, with gender-specific (masculine and feminine) and gender-free (epicene) morphemes that, most of the time, indicate whether we are referring to, addressing or talking about individual (or a group of) women or men. However, a cultural and social symbolism together with deep-rooted gendered stereotypes contributed to changing the understanding of grammatical rules, meaning that masculine job titles are also used for female professionals [...].

According to Fusco (2019, p. 49), these are:

> [...] usi linguistici che riverberano presupposti sociali e culturali obsoleti e un sistema di attese ancorato a una visione del mondo superata, densa di pregiudizi e di stereotipi verso la donna.[1]

In particular, this resistance finds support in the over-extended masculine, a practice that is still widespread and well established in Italian (see Somma and Maestri, 2020), which consists of using the masculine as the unmarked, neutral gender. As explained in Section 5.1.2, in our research we prefer the use of the adjective *over-extended*[2] (it. 'sovraesteso') rather than *unmarked* or *neutral*. These two terms, in our opinion, reflect the same language ideology that the literature here presented seeks to challenge. We recall here the notion of *Standard Language Ideology*. Following Milroy (2001), this can be described as the

[1] Our translation: [...] linguistic usages that reflect obsolete social and cultural assumptions and a system of expectations anchored in an outdated world-view, full of prejudices and stereotypes towards women.

[2] In particular, *over-extended* here is intended for all cases in which the masculine form is used to designate referents of any another gender or of an unknown gender.

belief that a language exists in a standard form, treated as the legitimate point of reference against which other varieties are judged. Given this definition, Cameron (2014)'s discussion of language ideologies in relation to gender is relevant here because it shows that such beliefs are inseparable from broader social assumptions, and are sustained in the interplay between representations of language and gendered linguistic practice. From this perspective, calling the masculine *neutral* or *unmarked* risks ratifying the ideological elevation of a dominant form by presenting it as self-evident. By contrast, in line with the QL approach adopted in this thesis, we want to unsettle that ideological naturalisation and to distance ourselves from a heteronormative gaze.

5.1.1 Gender, Morphology and Assignment

Following Thornton (2022), *grammatical gender* in Italian is best understood as an agreement-based category: nouns belong to gender classes insofar as they control the form of associated words such as articles, adjectives, pronouns, and, in specific environments, participles and other *agreement targets*.[3] Italian has two grammatical gender values, masculine and feminine. These values are not identified by noun endings or meaning alone, but by agreement behaviour. At the same time, the assignment of a gender value to nouns in Italian relies on a mixed system: *semantic* criteria are especially relevant for human and animate referents, while *formal* criteria—phonological and, in some cases, morpho-syntactic—also play an important role. As a result, semantic and formal criteria often converge, but they may also conflict. In nouns referring to humans, the semantic criterion tends to prevail, which is why forms such as *papa* [the pope] remain masculine despite their final *-a*.

It is therefore useful to distinguish between the two grammatical gender values and the different noun types that relate, in different ways, to the referential gender of human or animate referents.

First, *independent* or *heteronymic nouns* (it. 'nomi indipendenti' or 'eteronimi') express masculine and feminine through different lexical roots, as in *la madre/il padre* [the mother/the father] or *la sorella/il fratello* [the sister/the brother].

[3] Here, *grammatical gender* refers to the morphosyntactic category reflected in agreement; *referential gender* (it. 'genere referenziale') refers to the gender value associated with the intended referent in discourse; *social gender* (it. 'genere sociale') concerns gender as a socio-cultural construct; and *gender identity* refers to an individual's deeply felt sense of their own gender. See Giusti (2022), Brambilla, Crestani, et al. (2020), and Formato (2019).

Second, many nouns have distinct feminine and masculine forms without different lexical roots, as in *la maestra/il maestro* [the.F conductor/the.M conductor], *l'infermiera/l'infermiere* [the nurse.F/the nurse.M], or *la revisora/il revisore* [the.F reviewer.F/the.M reviewer.M]. These pairs may be more or less symmetrical in their derivational pattern.

Third, *common-gender nouns* (it. 'di genere comune') have a single lexical form but select masculine or feminine agreement according to the sex of the referent, as in *la/il docente* [the.F/the.M teacher] or *le/i cantanti* [the.F/the.M singers]. In dialogue with Formato (2019)'s discussion, some of these items may be described as *semi-epicene* when this common-gender behaviour is restricted to the singular, while the plural shows distinct masculine and feminine endings, as in *la/il giornalista* [the.F/the.M journalist], but *le giornaliste* [the.F journalists.F] and *i giornalisti* [the.M journalists.M].

Finally, *epicene nouns* (it. 'epiceni') have a fixed grammatical gender and trigger only one agreement pattern regardless of the sex of the referent, as in *la guardia* [the.F watch], *il pedone* [the.M pedestrian], *la sentinella* [the.F sentry], *la persona* [the.F person].[4]

This distinction matters because the literature does not always use these labels consistently. This thesis reserves *epicene* for nouns with a single agreement pattern (e.g., *persona*) and *common gender* for nouns with one form but variable agreement (e.g., *il/la comandante*),[5] and *mobile gender* for nouns which present distinct feminine and masculine forms (e.g., *capitana/o*)[6].

Nonetheless, there are cases of *incongruity* in which the assignment of grammatical gender to a noun does not follow the semantic criterion, and some of these cases also lead to a clear violation of the rules of agreement. These are what Formato (2019) defines as *semi-marked forms*, as in the sentence *"Marianna Madia, il ministro incinta: che super famiglia!"*, Formato, 2019, p. 56)[7], which are well attested in native speakers' linguistic productions (Cignarella et al., 2021). In light of Thornton (2022)'s discussion, these forms are best understood not simply as grammatical anomalies, but as sites where agreement, lexical choice, and gender ideology come into tension. In addition to specific

[4]For partially overlapping classifications and labels, see Sulis and Gheno (2022).

[5]Following Thornton (2022), this terminological choice preserves a structurally important contrast within the Italian system and avoids conflating agreement behaviour with broader semantic intuitions about inclusiveness or genericity.

[6]In this case following the taxonomy presented in Sulis and Gheno (2022).

[7]"Marianna Madia, the.M pregnant.M minister: what a super family!".

incongruous forms with regard to the designation of women, Formato (2019) provides a taxonomy of other linguistic uses influenced by gender assumptions, distinguishing, for example, between cases where masculine terms are used for mixed-gender groups and those where they are used for unknown or generic individuals. This broader issue corresponds to what Thornton reports as *maschile (cosiddetto) non marcato* (eng. 'the (so-called) non-marked masculine'), that is, the traditional use of masculine forms for referents whose sex (or gender) is unknown, irrelevant, mixed, or institutionally abstract. The need for more appropriate solutions therefore also involves references to mixed groups (e.g. *tutti gli studenti*, eng. "all students") or individuals whose sex or gender identity is not (yet) known (e.g. *dobbiamo assumere un nuovo impiegato*, eng. "we need to hire a new employee"), as well as references to abstract functions (e.g. *elezioni a sindaco*, eng. "mayoral elections").

Italian also displays cases in which morpho-syntactic form and referential properties do not align straightforwardly. Thornton discusses *hybrid nouns*, namely nouns whose formal properties and semantic interpretation may pull agreement in different directions. In the domain of professional titles, this is especially visible when a masculine lexical form is used for a woman, as in examples such as *il sindaco Letizia Moratti* [the.M mayor.M Letizia Moratti], where different agreement targets may oscillate between syntactic agreement with the noun and semantic agreement with the referent. This point is useful here because it shows that not all mismatches are simple violations of grammar: some involve variable interactions between lexical gender, referential sex, and the hierarchy of agreement targets.

5.1.2 Gender and Semantics

As is already clear so far, the category of gender—especially when related to language—manifests its complexity. As Violi (1986, p. 41) points out:

> Il genere non è soltanto una categoria grammaticale che regola fatti puramente meccanici di concordanza, ma è al contrario una categoria semantica che manifesta entro la lingua un profondo simbolismo.[8]

[8]Our translation: Gender is not merely a grammatical category that regulates purely mechanical facts of agreement, but is, on the contrary, a semantic category that manifests a profound symbolism within language.

In fact, there is the symbolic horizon within which the examples cited above operate and which has led the masculine gender to be considered neutral or unmarked. In the Italian linguistic literature the notions of markedness and neutrality have already been questioned from various points of view (among others, see Cavagnoli, 2013; Thornton, 2016; Voghera and Vena, 2016; Formato, 2019). As already demonstrated by Doleschal (2009) and several subsequent studies (Gygax et al., 2008; Gygax et al., 2019), even in natural gender languages such as English, the so-called "MAN principle" can be identified, i.e. "un bias culturale, per cui la struttura mentale, per default, associa a prescindere una figura maschile ad un termine anche nel caso in cui quest'ultimo si presenti di carattere neutrale, erodendo completamente l'identità femminile"[9] (Galeandro, 2021, p. 67).

Abbatecola (2016, p. 140) defines masculine neutrality as "una dimensione talmente incorporata da risultare opaca al nostro sguardo",[10] clarifying a concept already known since Sabatini (1986)'s *Raccomandazioni* and thoroughly investigated by studies on resistance to the use of the feminine gender in the Italian language (see Formato, 2019). Sabatini (1987) had already identified grammatical and semantic asymmetries, i.e. asymmetrical linguistic conventions operating at the grammatical level and at the discursive or lexical level with respect to gender. Examples of these two categories include, respectively, the use of masculine terms for mixed-gender groups, among other grammatical asymmetries, and the exclusive use of adjectives for a specific gender, among other discursive or lexical asymmetries, (such as "graziosa/o", eng. *pretty.F/.M*, almost always associated with female referents or more generally evoking an idea of femininity). To quote her:

> Molti di questi cambiamenti[11] non si possono definire 'spontanei', ma sono chiaramente frutto di una precisa azione socio-politica. Essi dimostrano l'importanza che la parola/segno ha rispetto alla realtà sociale ed il fatto che siano stati assimilati significa che il problema è veramente diventato 'senso comune' o che, per lo

[9] Our translation: a cultural bias whereby the mental structure, by default, associates a male figure with a term even if the latter is neutral in nature, completely eroding the female identity.

[10] Our translation: a dimension so embedded that it is opaque to our gaze.

[11] In the immediately preceding passage, Sabatini refers to ideological shifts reflected in lexical substitutions such as "netturbino" for "spazzino" and "nero" for "negro".

> meno, la gente ormai si vergogna al solo pensiero di poter essere tacciata di 'classista' o 'razzista'. Quando ci si vergognerà altrettanto di esser considerati 'sessisti' molti cambiamenti qui auspicati diverranno realtà 'normale'.[12] (Sabatini, 1993, p. 98).

Thornton (2022) usefully shows that the so-called 'non-marked masculine' has traditionally extended across at least four contexts in Italian: reference to a specific known person, reference to a specific but unidentified person, reference to a generic role holder, and reference to mixed plural groups. It is precisely these contexts that feminist interventions and institutional guidelines have progressively targeted, starting from Sabatini's *Raccomandazioni* and continuing with later administrative and editorial recommendations. Within this debate, two broad strategies have emerged. One privileges the visibility of women through feminine forms and explicit pairings; the other relies more heavily on neutralising or obscuring gender through collective nouns or genuinely epicene expressions. As Thornton notes, both strategies respond to the persistence of the over-extended masculine, which nevertheless remains deeply entrenched in actual usage.

The visibility of women in language is therefore interconnected with adequate visibility of women in society. However, it should not be forgotten that sometimes women themselves, driven by the desire to enter the male-dominated world of power, implement the opposite strategy, deliberately taking on male titles and seeking opportunities to reaffirm their choice in the media. This point becomes especially significant when considered in relation to the notion of language ideology discussed in Section 5.1.

An example of this phenomenon is the morphologically masculine assignment of the epicene *Presidente* by three Italian politicians in different eras.

The first one is the case of Irene Pivetti. In 1992, the politician was elected deputy and nominated President of the Chamber. During her inaugural speech, although not using the article *il*, she accompanied the title of president with the masculine *"cittadino"* [citizen.M] and *"cattolico"* [Catholic.M] to refer

[12]Our translation: Many of these changes cannot be described as "spontaneous", but are clearly the result of specific socio-political action. They demonstrate the importance that words/signs have in relation to social reality, and the fact that they have been assimilated means that the problem has truly become "common sense" or, at the very least, that people are now ashamed at the mere thought of being labelled "classist" or "racist". When people are equally ashamed of being considered "sexist", many of the changes hoped for here will become "normal".

to herself. She subsequently emphasised that "il ruolo di presidente della Camera, come ogni altro ruolo parlamentare, non è né uomo né donna"[13] (Villani, 2020, p. 112) and that "nella lingua italiana, anche se molti non se ne accorgono, esiste il genere neutro"[14] (Villani, 2020, p. 112).

The second, more recent example is that of Giorgia Meloni, our current Prime Minister, who, following her election (October 2022), issued a circular from Palazzo Chigi stating her wish to be referred to as "il Signor Presidente del Consiglio dei Ministri" (eng. "the.M Mister President of the Council of Ministers"), a statement that was later denied in a correction in which the use of the expression "il Presidente" (eng. "the.M President") was requested.

In contrast to these examples, we recall the case of another politician, the Democratic Party deputy Laura Boldrini. During her term as President of the Chamber of Deputies (2013–18), she fought for the implementation of Alma Sabatini's Recommendations within the institutions, among other things asking to be called "la Presidente" (eng. "the.F President").

Against this background, Meloni's choice can be read as ideologically significant. This is connected not only to her conservative political stance[15] but also to the fact that the issue became publicly salient through her own linguistic self-positioning. During Boldrini's term, Meloni had presented the question of gendered forms as marginal for her (see Villani, 2020). Years later, however, once in office, the preference for the masculine form was immediately explicitly foregrounded, so that the choice itself became part of the public construction of the role. As argued by Fumagalli and Rosola (2024), the choice of a gendered form in public communication is pragmatically loaded: it can signal alignment with a broader political orientation and make visible a stance towards gender-fair language, thus contributing to shaping how a role is publicly framed. In their discussion of Italian political communication, Boldrini's request to be addressed as *la Presidente* is treated as a marked choice that can associate the speaker with feminist claims, whereas Meloni's preference for the masculine is read as making salient a refusal of such claims. From this perspective, the choice of one form rather than another performs a

[13]Our translation: "the role of President of the Chamber, like any other parliamentary role, is neither male nor female".

[14]Our translation: "in the Italian language, even if many people do not realise it, there is a neutral gender".

[15]Which, we remind, is opposed to Boldrini's.

positioning in relation to existing ideological and political norms.

In this sense, this seems to be the point at which the issue clearly shift from being a morphosyntactic, lexical, or semantic matter to being also, and perhaps above all, a pragmatic one, since language not only signifies, but also *acts*.

5.1.3 The Debate on Nomina Agentis

Although several linguists have extensively explained the reasons for the incongrous use of masculine forms to designate women and its impact on the social representation of women in terms of reduced visibility in language (Thornton, 2016; Voghera and Vena, 2016; Adamo, Tigani Sava, and Zanfabro, 2019), there is still strong resistance to the use of gender-specific forms, especially in the case of prestigious roles and professions (Baldo, Corbisiero, and Maturi, 2016; Giusti, 2022) that were once closed to women.

This line of research, which has been ongoing for decades, has had (and still has) the merit of investigating and promoting gender issues in Italian in line with what Maturi (2020) defines as *pertinentizzazione del genere* (eng. 'gender specification'), conceptually opposed to the idea of *neutralizzazione del genere* (eng. 'gender neutralisation'). On the one hand, these studies have clarified the existence and use of feminine forms already provided for by the Italian language system and, on the other, they have made it possible to formulate suggestions and guidelines for a more gender-fair language (see Robustelli, 2012; Rosola et al., 2023).

Over the years, the cultural and social components of sexism and misogyny conveyed through the use of language have also been highlighted. Sexist language use and the stereotypes it conveys appear to be closely linked to offensive, discriminatory and hate speech, and can be expressed through more or less visible phenomena, on a continuum ranging from the over-extended use of the masculine form (e.g. "il Presidente del Consiglio Giorgia Meloni", eng. "the Prime Minister Giorgia Meloni") to explicit misogynistic insults (e.g. "ci sono troie in giro in Parlamento che farebbero di tutto, dovrebbero aprire un casino", eng. "there are whores in Parliament who would do anything, they should start a brothel" (Formato, 2017, p. 399).

In line with this perspective, therefore, the over-extended use of the masculine form (and its conceptualisation as 'unmarked use')—especially in cases

where the gender of the person being referred to is known and is female—is a phenomenon that goes beyond linguistic boundaries and leads to deeper reflection on the variants of nomina agentis and their social role, as well as on the impact that the use of such variants can have on the inclusion of women and the construction of gender stereotypes.

5.1.4 Gender-Fair Language (GFL) and Stereotypes in NLP

Several works in NLP tried to highlight the connection between stereotypes and specific gendered forms. Cassotti et al. (2021) demonstrated a correlation between changes in their use and related socio-cultural events, investigating job titles in a corpus of over 3 billion tokens extracted from two Italian newspapers over a period of 57 years and performing a diachronic analysis that exploits the frequency of specific gender forms. The corpus described in Cassotti et al. (2021) contains 3,529,820,155 tokens, covers a period from 1948 to 2005 and consists of two sub-corpora: the first, selected from the Italian newspaper L'Unità, is extracted from a previous work by Basile et al. (2020); the second is extracted from the digital archive of *La Stampa*, available to the public, using the same methodology indicated in the previous work.

A relevant study connecting the language use to stereotypes is that of Nardone (2016), which examines whether and to what extent certain feminine nouns denoting roles and professions still exhibit a pejorative semantic asymmetry compared to their masculine counterparts. The analysis was conducted using *itWaC*, one of the most extensive corpora available for the Italian language, which was consulted and analysed using *Sketch Engine* software. The survey focused mainly on the frequency of use and collocations of nouns that seemed to encounter greater resistance to use or that have often been at the centre of debate. The study confirmed that words such as *segretaria* [secretary.F], *direttrice* [director.F], *collaboratrice* [collaborator.F], *dottoressa* [doctor.F] and *professoressa* [professor.F], although now in common use, are still characterised by semantic asymmetry with respect to their corresponding masculine forms, while nouns such as *rettora* [rector.F], or *ingegnera* [engineer.F] were found very rarely, and none at all for *medica* [medical doctor.F], *notaia* [notary.F], and *ispettora* [inspector.F].

Starting from these considerations and bearing in mind the proposals for gender-fair language, the work of Rosola et al. (2023) elaborates the categories of masculine use, identifying an operational distinction between *incongruous*,

generic and *over-extended* use. The long-term goal is to develop a taxonomy that can be used to create an innovative annotation schema that, from a computational perspective, will enable the development of tools for the semi-automatic correction and reformulation of administrative texts from a gender-fair perspective.

5.2 The Preliminary Study on Nomina Agentis in Italian Social Media

To complement the linguistic overview provided above, we briefly summarise the preliminary study reported in Marra and Bosco (2024), which investigates gendered reference and the use of nomina agentis in Italian social media discourse. The study had the dual aim of (i) providing empirical evidence of how masculine and feminine professional titles are distributed in contemporary Italian usage, and (ii) testing an initial, operational annotation schema capable of making explicit the interpretative steps required to determine referential gender in context.

Crucially, the preliminary work did not aim to draw generalisable conclusions about the phenomenon as a whole. Rather, it was conceived as a methodological probe, designed to assess the feasibility and limits of fine-grained manual annotation in a socially and pragmatically complex domain.

5.2.1 Data and Sampling

The study analyses Italian *Twitter* (nowadays *X*) data extracted from the large-scale corpus described in Cignarella et al. (2021), which contains approximately 9,7 million tweets posted between 2006 and 2021 and filtered by professional role nouns and their gendered variants. From this resource, we selected a restricted temporal window (January–March 2021) and filtered all tweets containing five high-prestige nomina agentis, both in their masculine and feminine forms, in order to compare their distribution. Then, we extracted balanced samples only containing the following morphologically masculine forms: *avvocato* [lawyer.M], *ingegnere* [engineer.M], *ministro* [minister.M], *rettore* [rector.M], and *sindaco* [mayor.M]. These roles were chosen on the basis

of prior sociolinguistic literature documenting strong resistance to feminine forms in prestigious positions.

For each masculine form, a sample of 1,000 tweets was manually annotated, resulting in a dataset specifically tailored to observe cases of over-extended masculine usage. In a second step, the analysis zoomed in on a single case study—*avvocato*—and compared its masculine form with the two corresponding feminine variants (*avvocata* and *avvocatessa*), extracting additional samples of 500 tweets per feminine form. This two-step design allowed us to test both the general annotation logic and its application to a more fine-grained, contrastive analysis.

5.2.2 Annotation Schema and Categories

The annotation schema is deliberately simple yet theoretically informed. For tweets containing a masculine nomen, annotators assigned one of three primary labels: *masc*, when the form clearly referred to a male individual; *femm*, when the masculine form was incongruous as it explicitly referred to a woman (i.e., over-extended masculine); and *generic*, when the referent's gender was unknown, unspecified, or irrelevant (e.g. abstract reference to a role or office).

In addition, the schema included a residual category (*altro*, eng. 'other') to account for non-pertinent uses (e.g., homonymy, proper names) or cases in which referential gender could not be reliably reconstructed due to missing contextual information. Importantly, these latter categories were not treated as annotation errors, but as evidence of the intrinsic limits of text-only interpretation in social media environments.

A further transversal label (*dibattito*, eng. 'debate') was used to mark tweets that explicitly or implicitly engage in metalinguistic discussion about gendered forms themselves, rather than using them referentially. Although this dimension did not yield strong quantitative results at this stage, it anticipated later analytical developments and highlights the entanglement between linguistic usage and ideological positioning.

5.2.3 Results, Observations and Methodological Implications

The results of the study confirm a strong prevalence of masculine forms across all five nomina agentis, even when referring to women, with notable

variation across roles (Table 5.1). Considering only the tweets in which the masculine form referred to an individual of identifiable gender, the share used over-extensively to refer to a woman ranged from 5.9% for *rettore* to 23.7% for *avvocato*, with *ingegnere* (12.2%), *sindaco* (12.7%), and *ministro* (15.2%) in between. Feminine variants remain comparatively rare, particularly for highly prestigious positions, corroborating earlier sociolinguistic findings. However, the most relevant contribution of the study lies not in these distributional patterns *per se*, but in the methodological insights gained through annotation.

	Identifiable gender					Other	
nomen	masc	%	femm	%	generic	nr	np
avvocato	737	**76.3**	228	**23.7**	14	20	1
ingegnere	694	**87.8**	96	**12.2**	175	31	4
ministro	796	**84.8**	143	**15.2**	52	9	0
rettore	599	**94.1**	37	**5.9**	61	141	162
sindaco	789	**87.3**	115	**12.7**	54	42	0

TABLE 5.1: Distribution of the annotated categories across the five masculine nomina agentis (samples of 1,000 tweets each). The *masc* and *femm* percentages are computed on their joint subtotal, i.e. the cases in which the referent's gender is identifiable; *generic* and *altro* (*nr* = not recognisable, *np* = non-pertinent) are the residual, non-identifiable cases. Adapted from Marra and Bosco (2024).

In particular, the annotation process revealed that referential gender often cannot be determined on the basis of the text alone. Annotators frequently needed to retrieve additional contextual information, such as discourse history, named entities, or external world knowledge, and in some cases had to explicitly acknowledge that a reliable decision could not be made. The scale of this difficulty is itself telling: the proportion of tweets per sample that could not be assigned a reliable gender value (*nr*) ranged from 0.9% for *ministro* to 14.1% for *rettore*, the latter also showing a high number of non-pertinent occurrences (*np* = 162), largely owing to homonymy with the Italian singer Donatella Rettore. This confirms that gender assignment in Italian is not a purely

formal operation grounded in morphology, but a situated interpretative task shaped by social knowledge and pragmatic inference.

These findings directly support a core assumption of the present thesis: annotation in socially sensitive domains necessarily involves interpretative judgement, and uncertainty should be modelled rather than suppressed. This study thus provides empirical grounding for the socially situated and participatory orientation adopted in subsequent stages of this work, including the use of focus groups and surveys to inform annotation guidelines (Chapter 2).

nomen	masc	femm	generic
avvocato	0	20	0
ingegnere	5	21	1
ministro	1	1	0
rettore	2	4	2
sindaco	0	1	1

TABLE 5.2: Occurrences annotated with the transversal *dibattito* label, split according to the primary category (*masc*, *femm*, *generic*) assigned to the same tweet. Adapted from Marra and Bosco (2024).

A further methodological cue, which proves consequential for the rest of this thesis, emerged from the transversal *dibattito* label. Across the five masculine samples, metalinguistic uses were marginal in quantitative terms (Table 5.2): only when the masculine was used over-extensively to refer to a woman did such occurrences cluster at all, most visibly for *ingegnere* and *avvocato*. The picture changes drastically once the two feminine variants of a single *nomen* are examined. In a dedicated case study on *avvocato* and its feminine forms *avvocata* and *avvocatessa*, the annotation schema was slightly adapted: since an over-extended use of the feminine does not occur in the data, the *generic* label is not applicable, and referential occurrences were instead split into *persona*, when the nomen denotes a specific individual, and *ruolo*, when it is used abstractly for the role or office, with the transversal *dibattito* label retained as before. Under this schema, the share of occurrences falling under *dibattito* rose to roughly 66% for *avvocata* and 37.8% for *avvocatessa*

(Tables 5.3), as opposed to the handful of cases recorded for the masculine *avvocato*. In other words, the very act of choosing a feminine title—particularly the still-contested *avvocata*—frequently coincides with talk *about* that choice, rather than a neutral referential use.

	referente			
nomen	*persona*	*ruolo*	*dibattito*	total
avvocata	151	19	330	500
avvocatessa	269	42	189	500

TABLE 5.3: Annotation of the two feminine variants *avvocata* and *avvocatessa* (500 tweets each). Adapted from Marra and Bosco (2024).

This asymmetry indicates that, for socially loaded forms, metalinguistic and meta-discursive uses are not noise to be filtered out but a constitutive part of how the forms circulate, and that any annotation schema flattening them would discard precisely the dimension where social meaning is negotiated. This insight is taken up in the case studies that follow, where the entanglement between using a form and commenting on it becomes a relevant analytical concern.

5.2.4 Connection to Annotation Design

The preliminary study presented in this chapter serves as a concrete baseline for developing more refined annotation schemas that explicitly encode ambiguity, context dependence, and annotator decision-making. While limited in scope and scale, it demonstrates how sociolinguistic insights can be operationalised within computationally oriented workflows, and it anticipates the need for annotation practices that are sensitive to social meaning rather than restricted to surface grammatical features.

In this sense, the work complements both the NLP approaches discussed in Chapter 2 and the community-centred orientation of the FATA paradigm discussed in Chapter 4, positioning gendered reference as a paradigmatic case in which linguistic forms, social norms, and their interpretation are inseparably intertwined.

An additional aspect that deserves attention concerns the explicitly meta-discursive dimension of socially sensitive language use. As observed by Thornton (2022), drawing on De Mauro (1982)'s reflections on linguistic awareness, speakers do not merely use grammatical categories, but also reflect on them, especially in moments of social and linguistic change. In this sense, metalinguistic commentary on gendered forms is not external to language use, but constitutes an internal and recurrent feature of linguistic practice itself. This observation extends beyond gendered morphology alone. As Thornton notes—following Cameron (2012)'s notion of *verbal hygiene*—debates around contested linguistic forms often function as sites of normative negotiation rather than as purely technical discussions. Such metalinguistic and meta-discursive practices are particularly visible in domains where language is closely tied to social identity and power dynamics.

Seen from this angle, the choice of a feminine form where the over-extended masculine would conventionally be expected can also be read as carrying a reclamatory force, at least in a broad linguistic sense. The point is not that every feminine title should automatically be classified as a case of reclamation, but that such choices may function as deliberate acts of re-positioning against a dominant norm, especially when speakers explicitly claim them for self-reference. And, in light of the discussion about the use of *Presidente* reported in Section 5.1.2, even masculine form can have a reclamatory intention, if we consider the positionality of the person reclaiming it as ideologically conservative, in which cases, often, the use of the feminine variant can be lived as an imposition of a new dominant norm.

From this perspective, utterances that comment on, question, or contest gendered forms should not be treated as marginal or noisy data. Instead, they provide valuable insight into how linguistic norms are actively negotiated and, eventually, re-articulated by speakers. For NLP applications, this implies that meta-discursive instances embedded in corpora are analytically meaningful; they reflect speakers' metalinguistic awareness of socially loaded language and the meanings attached to it. Ignoring this dimension risks flattening the data and obscuring the interpretative processes through which stereotypes and norms are reproduced or challenged.

Chapter 6

Case Study 1: Gender Stereotypes in Public Discourse: the Rackete-Schettino Corpus

"Ma Q U A N T O sarebbe diversa la narrazione di questi eventi, se al posto di una capitana ci fosse stato un capitano? Quali sarebbero stati i toni usati? #SeaWatch3 #CarolaRacketeLIBERA"
Anonymous user, from the corpus

Building on the theoretical and methodological framework developed in the previous chapters, this case study investigates how public discourse is socially and linguistically constructed around two central figures, occupying the same professional role, but a different gender identity (a man and a woman), and being protagonists of two markedly different events. This contrast also serves to test the possibilities and limits of a fine-grained, socially situated annotation process for the analysis of stereotypes and related evaluative dimensions in Italian social media discourse.

Starting from the lessons learned from the aforementioned STERHEOTYPES project (see Section 1.1) and more generally from the literature presented in Chapter 3 and in Chapter 5, we realised that designing an annotation schema for the automatic detection of stereotypes could not disregard the investigation of features and dimensions which—mostly in the scope of HS detection—contribute to the construction and reinforcement of stereotypes.

The whole annotation process lasted about seven months and was carried out by taking into account the complexity of the phenomena investigated and

the consequent need to deal with this complexity and represent it as fair as possible in all the steps of the annotation process itself. It was a participatory process and, even if it was consistent with well-established annotation practice in NLP, we will dwell on the description and explanation of the reasons why decisions and changes were made. Therefore, in this chapter we present the process step-by-step, in the most transparent way, highlighting the methodological and analytical turning points that progressively shaped the final schema and the resulting interpretation of the corpus.

6.1 Data Collection and Annotation Design

The process started with the identification of public characters and events that had occurred in the recent past and had been extensively debated on social media, with an assumed polarisation determined by the social gender of the involved targets. More precisely, the aim was not to compare two historically equivalent events, but to contrast two discursive configurations centred on a man and a woman occupying the same professional role, in order to observe whether and how this shared role was narratively and evaluatively reconfigured in public discourse.

In a team composed of three researchers, we had our first **Brainstorming Meeting** where we decided to focus our data collection and annotation on two specific events, one with a male and one with a female protagonist, to pave the way for possible gender-based comparisons. At the same time, we were looking for contexts where we could find similarities that would be analytically useful for comparing them and investigating possible patterns. We then came up with two mediatically relevant events.

The **first event** is the sinking of the Concordia cruise ship in 2012, which occurred in the Tirrenian Sea and had tragic consequences due to a sail-by salute (known, in Italian, as *inchino*, "bow") to the coast of Isola del Giglio (Tuscany, Italy). The protagonist of the event was the Neapolitan Captain Francesco Schettino. The shipwreck took place on January 13, 2012. The risky sail-by salute manoeuvre led to the ship hitting a rock, causing it to partially sink and resulting in the deaths of 32 people, as well as numerous injuries. Schettino's handling of the emergency was immediately heavily criticised, particularly for his delay in ordering the evacuation and for abandoning the

vessel while many passengers were still on board. The accident sparked international outrage and a debate on maritime safety. We do not report the exact dynamics and explanations of the accident here—as they were established in the trials that took place in the following months and years—since our analysis concentrates instead on a time span of two months right after the event, sticking to the facts as they were elaborated by public opinion during that time.

The **second event** concerns a political case that took place in 2019 off the coast of Lampedusa (Sicily, Italy), resulting from a migrant rescue operation in the Mediterranean Sea and from the tension between maritime rescue obligations and the then recently enacted Italian Security Decree. The protagonist of the event was the German Captain Carola Rackete. In June 2019, while commanding the Sea-Watch 3—a vessel flying the Dutch flag and operated by the German NGO Sea-Watch—the captain and activist Rackete rescued 53 migrants off the coast of Libya. After transferring some of the shipwrecked people for medical reasons, she remained on board with 42 people, waiting for days for a safe port that the Italian authorities refused to grant. Faced with worsening conditions, she decided to enter the port of Lampedusa without authorisation. The episode took place in the context of the debate about the *Decreto Sicurezza Bis*—a decree promoted by the then Interior Minister Matteo Salvini—which tightened sanctions against NGOs involved in rescues. Rackete was arrested for violating maritime laws but was later released when the judge recognised that she had acted in a state of necessity, complying with international conventions on rescue at sea. The case sparked a heated debate on migration and humanitarian policies in Europe.

A detailed description of the events and the characters involved was later included in every version of the guidelines, so as to provide annotators with a stable contextual frame while still allowing the annotation task to focus on the tweets themselves.

6.1.1 Collecting the Data: The Rackete & Schettino Corpus

Once the frame of the study had been established, we gathered tweets concerning each of the two events from **TWITA** (Basile, Lai, and Sanguinetti, 2019), a resource consisting of a collection of Italian Twitter posts that span over a decade. Our data collection was achieved via keyword search; more

specifically, we selected tweets containing the surnames of the two protagonists, *Schettino* and *Rackete*. As the two episodes occurred years apart, we used different time ranges for the selection: January–February 2012 for Schettino and June–July 2019 for Rackete. In total, we gathered 5,933 tweets for Schettino and 550,824 tweets for Rackete. We hypothesise that the disparity in size is caused by the fact that **TWITA** intensified its efforts in collecting tweets over time. Afterwards, we cleaned the data by eliminating all duplicate texts, retweets, replies and quotes.

After this phase, the **Schettino** and **Rackete** sub-corpora were respectively composed of 1,100 and 13,202 tweets. As our work aims at comparing the two characters—in order to investigate online treatment reserved to men and women as a mirror of the underlying gender stereotypes—we balanced the data by selecting all filtered tweets from **Schettino** (1,100) and by randomly selecting 1,100 tweets from the larger **Rackete** sub-corpus. This choice was made in order to maximise cross-case comparability at the corpus level, even though it necessarily meant sacrificing the natural quantitative imbalance between the two events in the original source data.

6.1.2 Designing the Annotation: Creating a Draft of the Annotation Schema and its Guidelines

Once the dataset had been defined, the second **Brainstorming Meeting** took place. The research team manually examined a representative sample of 100 tweets to identify recurrent themes and discursive strategies, in a way that our choices would be as much data-driven as possible. During this preliminary observation, several tendencies emerged: a general **tendency to diminish or mock Captain Schettino**, a more **serious or empathetic tone when referring to victims**, and a **significant presence of humorous and ironic formulations** in both sub-corpora. Furthermore, the **attribution of blame appeared to be central** to the discourse, although the intensity and direction of responsibility attribution seemed to differ between the two protagonists.

This second Brainstorming Meeting was mostly guided by the aforementioned previous studies on social perception and stereotyping. In particular, we drew on Fiske's seminal work on the *Stereotype Content Model* (Fiske et al., 2002) to distinguish between perceived *cognitive/competence-related* dimensions

and *affective/warmth-related* dimensions in the tweets. In addition, we referred to recent contributions on the automatic detection of online hostility and stereotyping (Chiril, Benamara, and Moriceau, 2021), and to the framework proposed by Ferrando et al. (2024), where moral responsibility, emotional reactions, and gender stereotypes are observed in their interplay in public discourse.

Based on these observations, we decided to develop a multi-layered and multi-target annotation schema designed to capture both the explicit and implicit dynamics of Twitter messages.

The first layer included a total of nine dimensions, categorised as follows.

The first category was the ***Target*** of the message (*main character*, i.e., the protagonist of the event, vs. *other actors* such as victims, politicians, or institutions). This label enabled us to measure whether the discourse gravitates around the individual at the centre of the event or shifts towards secondary figures (those described in the texts provided to the annotators with the guidelines).

Then, five categories were explicitly related to the main character (Rackete or Schettino):

- the ***stance*** towards the target, for which we agreed on a tripartite classification *pro / against / neutral*, with the additional *I don't know* label;
- ***professional***, ***emotional***, and ***moral capacities*** of the target, which refer to their cognitive (e.g., skills, expertise, decision-making), affective (e.g., empathy), and moral (e.g., values) traits, to be evaluated on a *0-to-4 scale*;
- the presence of ***stereotypes***, within gendered or other stereotypical framing. For this category, if the stereotype was present we further distinguished a second-level dimension concerning the ***sphere to which the stereotype refers***, with one or more labels to choose among *body*, *behaviour/attitude*, *professional role/activity*, and *sexuality*.

Alongside them, three further parallel categories with related second-level multi-labelled options were conceived. They were binary categories (with an additional *I don't know* option) aimed at capturing key discursive dimensions identified during the exploratory phase:

- ***responsibility***: whether the tweet attributes blame, duty, or accountability to the target;

- ***irony***: presence or absence of humorous, sarcastic, playful, or derisive formulations;
- ***aggressiveness***: presence of explicit or implicit hostility towards the target.

If present, each of these three categories opened up to a second-level annotation allowing annotators to indicate one or more possible targets among the following: *main character*, *organisation*, *politics*, *society*, *victims*, and *other*. This decision allowed us to enlarge the scope in order to check for eventual phenomena of dehumanisation or hyper-humanisation of the actors involved (Ferrando et al., 2024), as both episodes involved victims, in one case migrants (see Schmeisser-Nieto et al., 2024a).

During the elaboration of the schema and, later, in the post-hoc analysis, special attention was also given to the expected use of specific nomina agentis (in particular, *capitana/o* [captain.F/captain.M] or *comandante* [commander]) in order to capture whether the protagonist-related role labels would appear (both in feminine or masculine form) and be discussed or contested in the tweets.

6.1.3 Researchers' Testing Pilot: Refinement of the Schema and Definition of the First Guidelines

To test the preliminary version of the schema, we selected a pilot sample of tweets to annotate and assigned it to each member of the research team. The **Researchers' Testing Pilot** aimed at producing a qualitative evaluation of inter-annotator agreement, refining unclear definitions, and identifying categories that may require merging or splitting (e.g., differentiating affective and moral capacities). The feedback from this stage would guide the production of the first version of the annotation guidelines to be used in the next annotation phase.

The Testing Pilot Corpus was annotated by the three members of the research team, using a Google Sheet table. The annotation involved the full set of categories described above and served to test their applicability and clarity across the two sub-corpora. From the outset it became clear that annotating the **Rackete** sub-corpus would be more complex than annotating the **Schettino** sub-corpus. In particular, the category of ***responsibility*** proved especially

difficult to apply consistently, confirming the need for more precise definitions and examples.

The Researchers' Testing Pilot also revealed notable asymmetries between the two datasets. Tweets about Schettino displayed a marked prevalence of humour and derision, with frequent jokes or ironic references turning the protagonist into a symbol of negligence or incompetence. By contrast, the Rackete-related set contained comparatively fewer stereotypes and a more polarised stance distribution. This suggested that humorous discourse and seriousness needed to be more explicitly captured in the annotation schema, either as independent dimensions or in interaction with stance. In fact, the research team debated whether to introduce a measure of ***tone of speech*** (e.g., *grave*, 'severe', vs. *leggero*, 'facetious') or to enrich the ***stance*** category so as to account for attenuating effects of humour. Moreover, looking at some ironic texts, mostly regarding Schettino, we realised that often the ***target*** of the whole message was not exactly one of the main characters, but rather a *figurative* reference to them. For this reason, we decided to introduce this third label alongside *main* and *secondary*.

Another important outcome of the Researchers' Testing Pilot was the emergence of gendered and stereotypical references to secondary figures—such as Domnica Cemortan, a character appearing in the Schettino sub-corpus—thus confirming the need to allow a dynamic selection of the *target* for the dimension of ***stereotype***. Thereby, we opted to permit multiple target labels instead of a single one. This decision also reflects an intersectional approach, consistent with Ferrando et al. (2024), acknowledging that stereotypes and blame attributions extend beyond primary actors. Moreover, in this direction, the label *geographical origin/nationality* was added to the ***sphere to which stereotype can refer*** dimension.

Regarding the ***competence-related*** categories, the researchers' annotation showed that treating cognitive, affective, and moral capacities as separate scalar dimensions was confusing. Moreover, moral actions (such as Rackete's rescue efforts) appeared also to constitute a demonstration of both professional competence and empathy. Conversely, humorous or ironic treatments of Schettino implicitly undermined his perceived professional competence. Consequently, we converged on a simplified annotation for ***these capacities***,

reducing the evaluation to a non-scalar *positive / negative / neutral / not applicable* format and merging affective and moral aspects. This choice was also grounded in Fiske's competence-warmth model (Fiske et al., 2002), which underlines how these dimensions co-occur in social perception.

Finally, the researchers' annotation led to a reconsideration of how ***stance***, ***aggressiveness***, and the presence of insults interact. We realised that the preliminary label ***aggressiveness*** did not really fit this dataset, as it was hard to recognise consistently, while what we found was a larger presence of insults. For this reason, we decided to replace that exploratory label with a more explicit dimension of ***offensiveness***, aligned with approaches to harmful language detection such as Chiril, Benamara, and Moriceau (2021). Moreover, we observed that *against* stance combined with humorous tone often produced attenuated blame rather than overt hostility. This refinement also aligns with the observation in Ferrando et al. (2024) that implicit and indirect forms of misogyny or hostility are especially prevalent in digital discourse and require nuanced annotation.

On the basis of these considerations, we produced a refined version of the annotation schema. Following the creation of the schema, we prepared a description of the events and the characters involved, together with a detailed presentation of each dimension and its categorisation, and we drafted a first version of the guidelines. This included clearer definitions, illustrative examples from both sub-corpora, and some highlights for selecting targets and applying categories consistently. These guidelines were later proposed to the annotators for their pilot annotation.

As our corpus is divided into two different sub-corpora that concern different events, we had to decide in which order to submit the tweets for annotation. We considered two options that lie between two extremes: (a) having annotators annotate first all the tweets related to one protagonist and then move on to the other one, and (b) alternating one tweet of one protagonist with one tweet of the other one and so on. We found that, in setting (a), annotators run the risk of not taking sufficient account of the difference between the two discursive contexts respectively linked to the Schettino and Rackete affair. Meanwhile, approach (b) can be very fatiguing, as annotators have to constantly switch their attention from one event to the other. In the end, we decided to adopt a compromise between the two

approaches: alternating subsets of ten tweets regarding each protagonist.

Here follows the description of the drafted annotation schema[1] as described in the first version of the guidelines (which can be checked in their original form in Appendix A), reporting a complete presentation of events and characters and explanations and examples for each category and label.

- ***Target*** indicates whether the target of the tweet is the *main* character (Rackete or Schettino) or a ***secondary*** character or *figurative*, in case a reference is figuratively made to the protagonist, i.e. the character is mentioned as a benchmark for other events or situations, even within lists of characters or facts, or even as a symbol of a phenomenon (e.g. Schettino as a symbol of Italian moral decadence). Examples:

  (1) a. Schettino disperato per il ritrovamento del corpo della bimba di 5anni...?!? Coglione dovevi svegliarti prima, troppo facile adesso!! [*main*]
  b. Il racconto di Rackete: "Dovevo entrare in porto, temevo suicidi. Ho avvisato ma nessuno parlava inglese" [*main*]
  c. @DeBortoliF La notizia della relazione extra coniugale di Schettino come quinta del corriere.it non fa onore alla grande testata che dirige [*secondary*]
  d. "Le decisioni della magistratura vanno sempre rispettate. Sia quando piacciono sia quando non piacciono. Questo è il senso della divisione e dell'indipendenza dei poteri dello Stato". @Roberto_Fico su #SeaWatch3 #CarolaRackete [*secondary*]
  e. Bellissimo servizio su #leiene sulla guerra al pizzo. C'è chi dice no alla mafia, l'Italia non è solo il paese di Schettino. [*figurative*]
  f. i #MondialiFemminili #CarolaRackete e #GretaThunberg ci dicono una sola cosa : Noi uomini abbiamo fatto il nostro tempo. #SeaWatch3 [*figurative*]

- ***Responsibility*** indicates whether reading the tweet reveals an attribution of responsibility in relation to what happened (the bow and the

[1] In the order that was then proposed to the annotators.

abandonment of the vessel by Schettino; the violation of the Security Italian Law and the entry without authorisation into the port of Lampedusa by Rackete). If present, it indicates whether responsibility is attributed to one or more entities (multiple response allowed) among *main character* (Schettino or Rackete), *organisation* (holding/vessel or NGO/vessel), institutions, *society*, *victims*, *other*. Examples (both regarding the *main character*):

(2) a. Trovato il corpo della piccola Daiana ....ma Schettino che ci fa a casa?in galera deve stare e' li il suo posto
b. #CarolaRackete (che ha speronato una motovedetta della #Gdf) è agli arresti domiciliari. Lo ha deciso la Procura di #Agrigento. L'accusa è resistenza e violenza a nave da guerra e tentato naufragio.

- ***Offensiveness*** indicates if there is any offensive statement in the tweet. If present, it indicates if it is directed to one or more entities (multiple response allowed) among *main character* (Schettino or Rackete), *organisation* (holding/vessel or NGO/vessel, in both cases to be intended as ship crew), *institutions*, *society*, *victims*, *other*. Examples (both regarding the *main character*):

(3) a. nuova parolaccia "schettino": irresponsabile senza cervello, pronto a negare anche l'evidenza, egoista, codardo; uomo di merda.
b. Ma la crucca #CarolaRacKete c'è l'ha i documenti o è clandestina anche lei #arrestatela

- ***Stance*** indicates whether the tweet is *pro*, *against* or *neutral* with respect to the activity or more generally to the person of the main character (Schettino or Rackete). In the annotation of this dimension, it must be taken into account the possible unaccountability of the actions performed by the target person or the attenuation of some of its negative characteristics as an element in favour (pro), for example through empathy towards the person or the event. Examples:

(4) a. @F_Schettino FORZA COMANDANTE, SONO CON TE !!!!! [*pro*]
b. Caro schettino ancora parli,tu devi stare zitto che dici solo stronzate,sei un raccomandato e basta,hai fatto morire bambini innocenti [*against*]
c. E con #CarolaRacketeLIBERA siamo un po' più liberi anche noi [*Pro*]
d. #CarolaRacketeINGALERA e i politici complici italiani e stranieri. [*against*]

- ***Humour*** indicates if the comment conveys a humorous content (using irony, puns, hyperbole, sarcasm etc.). If present, it indicates whether it is directed to one or more entities (multiple response allowed) among *main character* (Schettino or Rackete), *organisation* (holding/vessel or NGO/vessel), *institutions*, *society*, *victims*, *other*. Examples (both regarding the *main character*):

(5) a. i neutrini vanno veloci ma non quanto Schettino a scender dalla Concordia
b. #CarolaRackete Vuoi salvare e aiutare? Vai in corsia di una geriatria. Oppure c'è poca politica e poca grana che gira?

- ***Tone of speech*** indicates whether the discourse concerning the characters and/or the event is approached in a *light*, *severe*, or *neutral* manner. Examples:

(6) a. Dopo documentario Channel 4 su Costa Concordia non so come #Schettino sia tuttora ai domiciliari #schifo [*severe*]
b. @__StarGirl_ aspetta,io parlo di schettino, tu della moldava x caso? [*neutral*]
c. In che razza di mondo viviamo se alla prima difficoltà si scappa un esempio il Signor Capitano Schettino! Chiamarlo schettino è tanta roba ! [*light*]
d. La differenza tra Carola Rackete e Matteo Salvini sta in come incarnano le responsabilità che i loro ruoli comportano. [*severe*]

e. Non convalidato l'arresto di Carola Rackete. Escluso il reato di resistenza e violenza a nave da guerra. Per il reato di resistenza a pubblico ufficiale, il gip ha riconosciuto la discriminante dell'adempimento del dovere. [*neutral*]

f. Oltre ad aver portato in salvo una quarantina di migranti, a Carola Rackete va, senza alcun dubbio, riconosciuto un secondo grande merito: aver fatto diventare simpatica la Guardia di Finanza in un paese popolato da evasori fiscali. [*light*]

- ***Professional and cognitive abilities*** indicates whether a *negative* or *positive* judgement is present, through which the main character (Schettino or Rackete) is assessed or perceived with respect to professional or intellectual competence, or whether this is *not applicable*. Examples:

(7) a. Comunque se è vero che è #CarolaRacKete a guidare la #SeaWatch3 la affonderà tentando il parcheggio a S [*negative*]

b. #FF @nataliacavalli Non seguirla è come fare la luna di miele a Venezia, e trovarsi Schettino come gondoliere.' [*negative*]

c. Il Parlamento europeo vuole conoscere dalla voce di Carola Rackete, la capitana della nave #SeaWatch3 che ha deciso di forzare il blocco al porto di Lampedusa per salvare i migranti, se ci sono state violazioni delle libertà civili. [*not applicable*]

- ***Emotional and moral abilities*** indicates whether a *negative* or *positive* judgement is present, through which the main character (Schettino or Rackete) is assessed or perceived with respect to emotional, affective, and/or ethical-moral competence, or whether this is *not applicable*. Examples:

(8) a. Quanti di noi avrebbero fatto come Schettino? Non servono migliaia di anni di carcere ma ntra cinque e dieci secondo me. [*positive*]

b. Carola Rackete per me è un'eroina, non solo perché porterà in salvo 42 persone allo stremo delle forze dopo 15 giorni in mare aperto ma anche perché lo fa consapevolmente ed

assumendosene tutte le responsabilità. Non può piacere ai politici nostrani senza morale e responsabilità [*positive*]

c. Il tg5 quanto ha pagato per quel video? #Concordia #costa #schettino #giglio [*not applicable*]

- ***Aesthetic judgement*** indicates whether the tweet shows an aesthetic judgment (on body, parts of the body, clothing etc.) against the main character. Examples:

(9) a. Secondo me #Berlusconi e #Schettino sono andati nella stessa scuola.Di certo peró non vanno più dallo stesso barbiere. #glipiacerebbe

b. Visto che il @pdnetwork ha raccolto tanti soldi per la capitana #KarolaRackete potrebbe almeno vestirla? O la lasciano andare in tribunale seminuda? La forma é sostanza, diceva Benedetto Croce

- ***Stereotype*** indicates the presence of any stereotypical content, whether this is implicit or explicit (*yes/no/I do not know*).

  If the stereotype is present, ***to which group of people it refers***, among a set of possible targets. For the Rackete case: *women*; *women at the helm of a vessel*; *migrants*; *African people or Africa or African countries in general*; *NGOs*; *other*. For the Schettino case: *men*; *men at the helm of a vessel*; *women*; *women from Eastern Europe*; *cruise companies or people going on a cruise*; *other*.

  If the stereotype is present, ***to which sphere it refers***, for both cases the choice was among: *body*; *behaviour (or attitude)*; *activity (or role)*; *sexuality*; *origin/nationality*.

  Examples:

(10) a. La moldava sarebbe finita a letto con Schettino. Nulla é più pericoloso di un uomo che vuole fare impressione su una per farsela. [*women from Eastern Europe*; *sexuality*]

b. Mentre le femministe au caviar si stracciano il #reggiseno per la Rackete (infatti è 'ricca bianca tedesca'), lasciano marcire al loro destino le infibulate, le segregate, le spose bambine e le malmenate dai compari muslim che la #Rackete vuole alloggiare in italia #bastaong [*migrants*, *women*; *origin/nationality*, *behaviour (or attitude)*]

In the guidelines, **targets of *responsibility*, *offensiveness*** and ***humour*** were clarified as it follows:

- the *main character*, i.e. Schettino or Rackete;
- *the vessel or the holding/NGO, i.e. organisation*, to be understood as crew or group of people who worked or carried out activities on the Costa Concordia or the company Costa for Schettino, and as crew or group of persons working or carrying out activities on the Sea-Watch 3 or the NGO Sea-Watch for Rackete;
- *institutions*, whatever public institutions or bodies, such as the judiciary and its members or political figures—possibly indicating in the notes if it refers to specific people—and parties, including if the reference is to the political debate in general;
- *society*, in a broad sense (e.g. education, media, public debate, web users etc.);
- *victims*, to be understood as persons who were on the vessel and suffered from the decisions of the target (regardless of whether they caused death or physical damage), but also as subjects or groups that were not on the vessel but have suffered negative consequences as a result of what happened (e.g. relatives, minority groups, etc.). The Moldovan woman, Domnica, who was with Schettino during the incident is to be considered as a victim on an equal footing with the other people present on the ship;
- *other*.

6.2 The Annotation Task

At the beginning of the process, we recruited four annotators: a tenure track researcher, a post-graduate research fellow and two master's students. The annotators had different backgrounds, but all of them being familiar to different degrees with at least one among Corpus Linguistics, Language Technologies, or Computational Linguistics. However, only two of them annotated the entirety of the dataset (and an additional third person was recruited at a later time).

Annotators' participation was voluntary and, for some of them, that was part of their work on their thesis or internship. As the schema included many dimensions, the guidelines were very detailed to improve their shared knowledge of the context.

In addition to the guidelines, we planned frequent meetings with the annotators to clarify any possible doubts on the schema as their annotation work progressed. Their feedback about the categories and related labels were used during the whole process as precious indications to make adjustments to the annotation schema.

The annotation was carried on the same platform used in Ferrando et al. (2024) (see Figure 6.1), designed by one of the members of the research team.

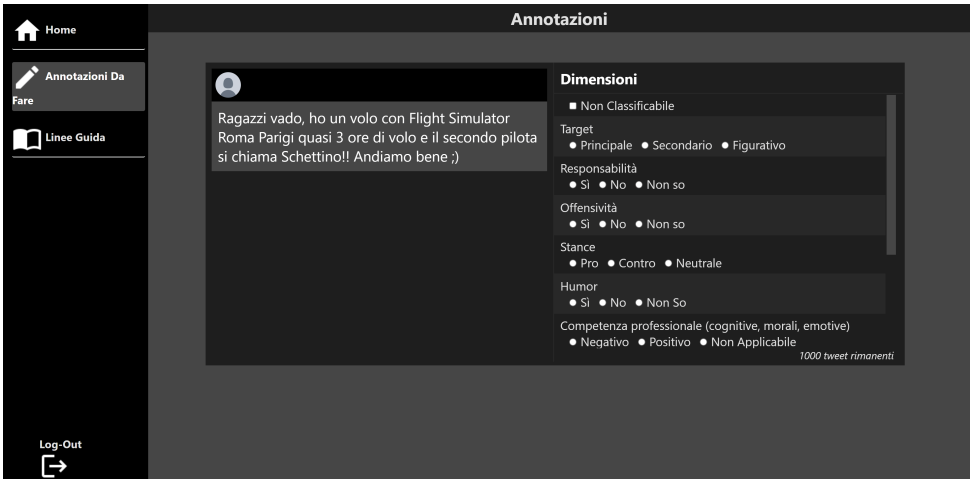

FIGURE 6.1: An illustrative screenshot of the annotation platform interface.

6.2.1 Pilot Annotation: Introduction and Check

Before starting the full-scale annotation, we organised an **Introductory Meeting** to present the project, the annotation schema and the preliminary version of the guidelines. In this meeting, we clarified the main objectives, discussed potential ambiguities, and offered the annotators the opportunity to raise any doubts. After this preparatory stage, we launched a pilot annotation phase on a random sample of 110 tweets drawn from both the **Schettino** and **Rackete** sub-corpora.

After two weeks, when the pilot annotation was concluded by all the participants, we held an **After-Pilot Meeting** to discuss the updated schema, resolve any remaining ambiguities and set a partial deadline (1,000 tweets) for the next phase of annotation. This iterative approach, combining pilot testing and immediate refinement, helped us to stabilise schema and definitions before the large-scale-data annotation phase began.

The After-Pilot Meeting confirmed that certain categories were more challenging than expected. For example, ***responsibility*** emerged as a particularly complex dimension, especially for tweets about Rackete. Similarly, annotators noted the high incidence of stereotypes not only when referring to the main characters but also originated while talking about secondary characters (such as Domnica in the Schettino set) and groups (e.g. people from Southern Italy). This observation reinforced our previous decision to allow annotators to select more dynamically the targets for several categories rather than forcing annotation on the two main characters only.

Other issues emerged as well: (i) a tendency to conflate cognitive, professional, emotional and moral capacities in the evaluation of the protagonists, (ii) uncertainty about how to handle tweets mentioning institutional actors such as magistrates or politicians, (iii) the need to differentiate between *institutions* and *politics*, and (iv) requests to mark the tone of discourse more explicitly (moving from *grave*, eng. 'severe', to *serio*, 'serious').

Following these comments, we applied a series of revisions both to the annotation schema and to the guidelines.

We added the label *other* to the ***target*** list, allowing annotators to select it whenever a tweet referred to a person or group not listed among the main or secondary characters. We merged cognitive, moral and emotional capacities into a single dimension of ***professional competence*** with a simpler *positive*

/ negative / neutral / not applicable coding. We also split *institutions* into two distinct labels (*institutions* and *politics*) and clarified that politicians, parties and public debates were to be coded under the latter. Finally, we revised the description of the ***stereotype*** dimension to include the recognition of stereotypes also when implicit, and instructed annotators to use the *other* label if a specific stereotype sphere or group was not present in the list.

In addition to the four annotators, a new annotator—a master's student—joined the project. A separate meeting was arranged by the present writer (which from now on will be referred to as MR, i.e., Main Researcher) with her to ensure she was aligned with our previous work. At the same time, two of the annotators (the tenure track researcher and a Master's student) discontinued their participation. Therefore, as anticipated, the annotation task was performed in its entirety by a team of three annotators.

In any case, all pieces of feedback—gathered respectively during the After-Pilot Meeting and during the Check Meeting, which took place after a small amount of actual annotations (see Section 6.2.2)—were included in our qualitative evaluation of the schema and contributed to the next round of revisions.

6.2.2 First Block of Annotation

When the After-Pilot annotation schema and guidelines resulting from the meeting were ready (see Appendix B), the actual annotation got started. We fixed the amount of tweets for the first block to 1,000. During the first block of annotation feedback was collected from the annotation team, and was taken into account for the Preliminary Analysis and for the following and crucial Feedback Meeting.

In particular, three weeks after the start we held a dedicated **Check Meeting** to address any doubts, ambiguities, or challenges that the annotators encountered in this initial phase. The main purpose of this meeting was to ensure a shared understanding of dimensions, labels and guidelines, without introducing changes to the annotation schema itself.

The feedback collected during the Check Meeting confirmed that most categories were now clear and manageable. It was noted that merging emotional-moral and professional capacities into a single ***professional competence*** label had been effective, and that, in general, annotators encountered few

difficulties. Some issues remained with the selection of ***targets***, particularly for the *other* label, and with the category ***responsibility***, which occasionally required deliberation, especially in the **Rackete** set. One annotator observed challenges in identifying stereotypes, particularly when they were expressed indirectly, as a reported speech (e.g., through quotation marks or subtle critique) without an overtly stereotypical intent. The MR clarified that, following the schema, all such instances could be then taken into account in a broad conceptualisation.

One annotator noted that stereotypes were especially difficult to identify when they appeared indirectly, for instance in reported speech or in quotations, without a clear stance associated with the current writer. It is worth noting that this is a difficulty inherent to annotation itself, since it is often unclear whether reporting a discourse also means endorsing it.[2] However, the MR clarified that, following the adopted schema, such cases could still be included within a broad conceptualisation of stereotypes.

6.2.2.1 Preliminary Analysis

Two months later, all members of the annotation team had completed the first block of annotation. Therefore, we conducted a preliminary analysis on the resulting dataset of 1,000 tweets, evenly split between the two protagonists (500 for Schettino and 500 for Rackete). All tweets were annotated by the three annotators and, in addition, a sample of 600 of these tweets (300 per subset) was independently annotated by the MR as well. This partial overlap wanted to make it possible to examine agreement not only within the annotation team, but also between the team and the schema designer, the MR, thereby providing a concrete basis for reflecting on researcher positionality and potential sources of bias. Therefore, a dataset was generated for the full set of 1,000 tweets to support descriptive analyses, to which we will refer as **Pre_Dataset**, and then respectively to the two subsets as ***Pre_Rackete*** and ***Pre_Schettino***.

Therefore, in this phase three complementary analyses were carried out.

[2]The same issue, as we will see in Chapter 7, is central to the study of slur reclamation, where reported and meta-discursive uses are likewise difficult to classify because reproducing a term does not automatically reveal whether its offensive force is being sustained or contested.

First, corpus statistics were computed on **Pre_Dataset**, ***Pre_Schettino*** and ***Pre_Rackete***, after aggregation, i.e., after the results had been elaborated through majority voting.[3]

Second, inter-annotator agreement (IAA) was calculated among the three annotators on the **Pre_Dataset** and the related subsets, using Fleiss' κ (see Landis and Koch, 1977), which corrects for chance agreement among multiple raters.

Third, on the subset of 600 tweets annotated both by the team and by the MR, IAA was calculated by means of Cohen's κ (see Landis and Koch, 1977) so as to allow pairwise comparisons between the MR and each individual annotator. This made it possible to compare the person who designed the schema with those who carried out the annotation.

Consistent with the model of continuous collaboration with annotators adopted in this project, this preliminary analysis served to evaluate the robustness of the annotation schema and to determine whether further adjustments were necessary.

6.2.2.2 Corpus Statistics (Pre_Dataset)

We begin with a descriptive overview of the preliminary aggregated dataset (**Pre_Dataset**), derived from the first block of 1,000 tweets annotated by the three annotators. A methodological clarification is needed at the outset: tweets marked as *non-classifiable* were excluded prior to aggregation. In the full dataset, 14 tweets were excluded following majority voting, resulting in 986 valid tweets.

[3]A technical specification is needed here with regard to the aggregation of the dataset and the majority voting. For the first-level dimensions, majority voting was applied by retaining as *true* only those categories selected by at least two annotators. When no category reached this threshold—that is, when each annotator selected a different label—the case was recorded as a *tie*. It should be stressed that *tie* does not yet amount to a measure of agreement or disagreement; it simply indicates that no first-level category obtained a majority vote in the aggregated dataset. For the second-level dimensions, the procedure was conditional on the corresponding first-level category being selected. Once a first-level label was retained through majority voting, the related second-level labels were included as *true* whenever they had been selected by at least half of the annotators who had selected that first-level label. In practice, this means that, when the first-level label (in our cases, the *Yes* label) had been selected by two annotators, a second-level label was retained if at least one of them had chosen it; when it had been selected by all three annotators, the threshold was two.

The statistics reported below are therefore *de facto* based on 986 tweets, with 489 tweets concerning Francesco Schettino and 497 concerning Carola Rackete.

As we explained above, aggregation was performed by majority voting. In cases where no majority label was reached, the outcome was recorded as a *tie*. For some dimensions the annotation schema explicitly allowed annotators to select an option corresponding to *I don't know* (from now on, *Idk*) or to *other*; when these labels received a majority of votes, they were correctly retained as a distinct aggregated outcome.

For first-level dimensions, annotators were required to select a single label. By contrast, multi-label selection was possible only for second-level categories. In particular, when ***Responsibility***, ***Offensiveness***, or ***Humour*** were annotated as present, annotators could select multiple targets indicating whom or what these evaluations referred to. Similarly, when *Stereotype* was marked as present, two second-level categories were activated, allowing annotators to label both the ***targets of the stereotype*** and the ***sphere to which it pertained***.

Throughout the analyses, values are reported as proportions (in decimal form) and raw frequencies. In tables, raw frequencies are provided in parentheses, whereas in the text they are reported following a semicolon (e.g. 0.43; n = 427). For the second-level categories, proportions may sum to more than 1.00, as they are interpreted as distributions of selected labels rather than as mutually exclusive dimensions.

6.2.2.2.1 Target Across the full dataset, the Target dimension shows a clear predominance of references to the main character (0.43; n = 427). However, the two subsets diverge in how often the discourse expands beyond the protagonist. In ***Pre_Schettino***, *figurative* references are remarkably frequent (0.35; n = 173). By contrast, ***Pre_Rackete*** exhibits a substantially higher proportion of *secondary* targets (0.32; n = 161) and a relevant amount of *other* (0.11; n = 55), indicating the activation of additional actors such as political figures, institutions, or broader collective entities.

Target	**Pre_Dataset (986)**	***Pre_Schettino (489)***	***Pre_Rackete (497)***
Main	0.43 (427)	0.44 (213)	0.43 (214)
Secondary	0.19 (191)	0.06 (30)	0.32 (161)
Figurative	0.18 (177)	0.35 (173)	0.01 (4)
Tie	0.12 (117)	0.11 (54)	0.13 (63)
Other	0.08 (74)	0.04 (19)	0.11 (55)

TABLE 6.1: Distribution of the **Target** dimension in the aggregated dataset of 986 tweets (annotation team only). Proportions are computed within each set.

This asymmetry suggests that, while both events generate highly personalised discourse, the communicative framing differs: Schettino often functions as a symbolic or metaphorical figure beyond the specific shipwreck narrative, whereas the Rackete case more often expands into a wider discursive arena.

6.2.2.2.2 Responsibility At the overall level, responsibility is slightly more often annotated as absent (0.49; n = 486) than present (0.41; n = 403), while *Idk* outcomes remain marginal (0.01; n = 13).

Subset differences are substantial: ***Pre_Schettino*** shows a higher proportion of *Yes* (0.45; n = 218) alongside a higher proportion of ties (0.12; n = 60), whereas ***Pre_Rackete*** is more frequently annotated as *No* (0.57; n = 281), with fewer ties (0.05; n = 24).

Dimension	**Set**	**Yes**	**No**	**Idk**	**Tie**
Responsibility	**Pre_Dataset**	0.41 (403)	0.49 (486)	0.01 (13)	0.09 (84)
	Pre_Schettino (489)	0.45 (218)	0.42 (205)	0.01 (6)	0.12 (60)
	Pre_Rackete (489)	0.37 (185)	0.57 (281)	0.01 (7)	0.05 (24)

TABLE 6.2: Distribution of the **Responsibility** dimension (first level; majority voting).

As shown in Table 6.3, across the dataset responsibility is predominantly attributed to the *main character* (0.87; n = 350). This tendency is particularly strong in ***Pre_Schettino*** (0.96; n = 209), where responsibility appears overwhelmingly individualised. In ***Pre_Rackete***, responsibility remains frequently

directed at the *main character* (0.76; n = 141), but it also extends with non-negligible selection rates to *politics* (0.28; n = 52).

Dimension	Set	Main char.	Organisation	Inst.	Polit.	Soc.	Victims	Other
Responsibility	**Pre_Dataset**	0.87 (350)	0.06 (23)	0.04 (16)	0.14 (58)	0.06 (23)	0.00 (0)	0.02 (9)
	Schettino	0.96 (209)	0.04 (9)	0.01 (3)	0.03 (6)	0.04 (9)	0.00 (0)	0.02 (5)
	Rackete	0.76 (141)	0.08 (14)	0.07 (13)	0.28 (52)	0.08 (14)	0.00 (0)	0.02 (4)

TABLE 6.3: Targets for **Responsibility** (second level; computed only where Responsibility = *Yes*). Multi-label selection allowed.

Overall, these distributions suggest that responsibility attribution in the Schettino discourse is primarily anchored in individual fault, whereas in the Rackete discourse it more readily expands towards political accounts, consistent with the broader politicisation of migration-related debate.

6.2.2.2.3 Offensiveness Overall, most tweets are annotated as not offensive (0.53; n = 525), while a substantial portion is classified as *offensive* (0.34; n = 340). A non-trivial share results in ties (0.10; n = 102), while *Idk* outcomes remain limited (0.02; n = 19), indicating that offensiveness is very often interpretable but not always unequivocal.

Subset differences are pronounced. ***Pre_Schettino*** is slightly more frequently annotated as *not offensive* (0.56; n = 275) and shows a higher proportion of ties (0.14; n = 70). By contrast, ***Pre_Rackete*** exhibits a higher incidence of offensiveness (0.41; n = 206) and fewer ties (0.06; n = 32).

Dimension	Set	Yes	No	Idk	Tie
Offensiveness	**Pre_Dataset**	0.34 (340)	0.53 (525)	0.02 (19)	0.10 (102)
	Schettino	0.27 (134)	0.56 (275)	0.02 (10)	0.14 (70)
	Rackete	0.41 (206)	0.50 (250)	0.02 (9)	0.06 (32)

TABLE 6.4: Distribution of **Offensiveness** (first level; majority voting).

At the overall level, offensiveness is most frequently directed towards the *main character* (0.46; n = 157) and *politics* (0.38; n = 128), with additional targeting of *society* (0.17; n = 59) and *other* (0.21; n = 70). The subsets diverge sharply: offensiveness in the Schettino discourse is overwhelmingly personalised towards the *main character* (0.75; n = 100), whereas in the Rackete discourse it

is predominantly redirected towards political actors (0.59; n = 122), with the main character selected much less often (0.28; n = 57).

Dimension	Set	Main char.	Organisation	Inst.	Polit.	Soc.	Victims	Other
Offensiveness	**Pre_Dataset**	0.46 (157)	0.02 (7)	0.04 (12)	0.38 (128)	0.17 (59)	0.05 (18)	0.21 (70)
	Schettino	0.75 (100)	0.02 (3)	0.01 (2)	0.04 (6)	0.15 (20)	0.00 (0)	0.32 (43)
	Rackete	0.28 (57)	0.02 (4)	0.05 (10)	0.59 (122)	0.19 (39)	0.09 (18)	0.13 (27)

TABLE 6.5: Targets for **Offensiveness** (second level; computed only where Offensiveness = *Yes*). Multi-label selection allowed.

These patterns suggest that offensiveness functions as individual-centred blame in the Schettino case, while in the Rackete case it more often shifts towards a broader political arena, targeting actors implicated in migration governance, consistent with a more confrontational and polarised debate.

6.2.2.2.4 Stance At the overall level, *against* stances constitute the largest proportion of the corpus (0.47; n = 462), followed by *pro* (0.25; n = 251) and *neutral* (0.20; n = 196), with ties accounting for 0.08 (n = 77).

When considering the subsets separately, stance sharply differentiates the two events: ***Pre_Schettino*** is predominantly condemnatory (*against*: 0.61; n = 296; *pro*: 0.02; n = 10), with a non-negligible portion of neutrality (0.30; n = 148), whereas ***Pre_Rackete*** contains a substantial amount of supportive positioning (*pro*: 0.48; n = 241), alongside a lower share of *against* stances (0.33; n = 166) and a smaller portion of neutrality (0.10; n = 48).

Dimension	Set	Pro	Neutral	Against	Tie
Stance	**Pre_Dataset**	0.25 (251)	0.20 (196)	0.47 (462)	0.08 (77)
	Schettino	0.02 (10)	0.30 (148)	0.61 (296)	0.07 (35)
	Rackete	0.48 (241)	0.10 (48)	0.33 (166)	0.08 (42)

TABLE 6.6: Distribution of **Stance** (first level; majority voting).

These distributions confirm stance as a highly discriminative dimension between the two cases: Schettino-related discourse is largely oriented towards condemnation towards the captain, while Rackete-related discourse again appears to be more polarised as it includes a large supportive component and a less relevant evaluation of neutrality, consistent with the ideological contestation surrounding migration and border politics.

6.2.2.2.5 Humour Overall, humour is slightly more often annotated as absent (0.48; n = 476) than present (0.45; n = 443), while *Idk* outcomes (0.01; n = 13) and *ties* (0.05; n = 54) remain definitely negligible, thus suggesting that it was easy for the annotators to make a decision.

The subset comparison reveals a strong asymmetry: humour is prevalent in ***Pre_Schettino*** (*Yes*: 0.67; n = 329) and comparatively marginal in ***Pre_-Rackete*** (*Yes*: 0.23; n = 114), where non-humorous tweets dominate (*No*: 0.73; n = 363).

Dimension	Set	Yes	No	Idk	Tie
Humour	**Pre_Dataset**	0.45 (443)	0.48 (476)	0.01 (13)	0.05 (54)
	Schettino	0.67 (329)	0.23 (113)	0.01 (7)	0.08 (40)
	Rackete	0.23 (114)	0.73 (363)	0.01 (6)	0.03 (14)

TABLE 6.7: Distribution of **Humour** (first level; majority voting).

At the overall level, humour is predominantly directed at the *main character* (0.75; n = 334), with lower but non-negligible targeting of *politics* (0.19; n = 86) and *other* (0.24; n = 107).

Nonetheless, when looking at subset-specific patterns, we notice a sharp divergence: in ***Pre_Schettino***, humour is almost exclusively protagonist-centred (0.92; n = 303), whereas in ***Pre_Rackete*** humour primarily targets political actors (0.67; n = 76), with the *main character* selected comparatively rarely (0.27; n = 31).

Dimension	Set	Main char.	Organisation	Inst.	Polit.	Soc.	Victims	Other
Humour	**Pre_Dataset**	0.75 (334)	0.02 (9)	0.04 (16)	0.19 (86)	0.08 (36)	0.00 (1)	0.24 (107)
	Schettino	0.92 (303)	0.02 (8)	0.01 (2)	0.03 (10)	0.07 (23)	0.00 (0)	0.27 (89)
	Rackete	0.27 (31)	0.01 (1)	0.12 (14)	0.67 (76)	0.11 (13)	0.01 (1)	0.16 (18)

TABLE 6.8: Targets for **Humour** (second level; computed only where Humour = *Yes*). Multi-label selection allowed.

Taken together, these distributions suggest that humour in the Schettino discourse plays a pivotal role, as it contributes to personalisation and trope-building around the protagonist, which could also explain the relevant portion

of figurative representation seen in Table 6.1. Conversely, humour in the Rackete discourse—when present—seems to be more often integrated into political contestation and directed towards institutional or ideological opponents.

6.2.2.2.6 Tone of Speech At the overall level, tone is most frequently annotated as *light* (0.51; n = 498), followed by *serious* (0.43; n = 423), while *neutral* remains marginal (0.01; n = 11).

The subset comparison reveals a pronounced divergence: ***Pre_Schettino*** is strongly associated with a light tone (0.76; n = 370), whereas ***Pre_Rackete*** is predominantly serious (0.69; n = 343).

Dimension	Set	Light	Neutral	Serious	Tie
	Pre_Dataset	0.51 (498)	0.01 (11)	0.43 (423)	0.05 (54)
Tone of speech	Schettino	0.76 (370)	0.01 (7)	0.16 (80)	0.07 (32)
	Rackete	0.26 (128)	0.01 (4)	0.69 (343)	0.04 (22)

TABLE 6.9: Distribution of **Tone of speech** (first level; majority voting).

This contrast reinforces earlier findings for humour and target of the tweet as well as for offensiveness: the Schettino discourse more often mobilises light registers that enable irony, mockery, insults, and figurative reconfiguration, while the Rackete discourse is more frequently articulated through serious, morally charged, and politicised framing. As we originally conceived it—and how it seems to be confirmed by annotation—tone of speech does not vary independently, but aligns with these other dimensions as a coherent register-level signal of how each event is narrated and evaluated.

Taken together, the distributions of target, offensiveness, humour, and tone of speech consistently point to two different discursive framings: a more personalised and trope-oriented construction in ***Pre_Schettino***, and a more debate-oriented arena in ***Pre_Rackete***.

6.2.2.2.7 Professional Competence At the overall level, negative evaluations of professional competence are the most frequent (0.49; n = 482). A substantial portion of tweets is classified as *not applicable* (0.23; n = 223), suggesting that many tweets do not primarily frame the protagonists in terms of

competence. Positive evaluations are comparatively less frequent (0.12; n = 122), while ties account for 0.16 (n = 159), indicating non-trivial ambiguity in how competence-related judgement is inferred from the text.

The subset comparison highlights a marked asymmetry. In ***Pre_Schettino***, negative evaluations dominate (0.68; n = 334), with positive assessments almost absent (0.01; n = 6) and a smaller share of ties (0.05; n = 25). In ***Pre_-Rackete***, negative ones are less frequent than for Schettino (0.30; n = 148), whereas Rackete-related discourse shows a significantly higher presence of positive competence evaluations compared to ***Pre_Schettino*** (0.23; n = 116), and ties are substantially more common (0.27; n = 134).

Dimension	Set	Positive	Negative	Not applicable	Tie
Professional competence	**Pre_Dataset**	0.12 (122)	0.49 (482)	0.23 (223)	0.16 (159)
	Schettino	0.01 (6)	0.68 (334)	0.25 (124)	0.05 (25)
	Rackete	0.23 (116)	0.30 (148)	0.20 (99)	0.27 (134)

TABLE 6.10: Distribution of **Professional competence** (first level; majority voting).

These distributions suggest that competence-related framing is not only polarising but also case-sensitive: Schettino-related discourse strongly associates the protagonist with professional failure, whereas Rackete-related discourse more often supports positive competence evaluations. At the same time, the larger share of ties in the **Rackete** subset indicates non-trivial interpretative dispersion in how competence-related judgements are inferred in that discourse.

6.2.2.2.8 Aesthetic Judgement Across the full dataset, aesthetic judgement is rarely activated (*Yes*: 0.02; n = 22), while the vast majority of tweets are annotated as not involving aesthetic evaluation (*No*: 0.97; n = 959). This distribution is stable across subsets, with only minor differences (Schettino: 0.03; n = 13; Rackete: 0.02; n = 9), suggesting that appearance-based evaluations are marginal within this corpus.

Dimension	Set	Yes	No	Tie
Aesthetic judgement	**Pre_Dataset**	0.02 (22)	0.97 (959)	0.01 (5)
	Schettino	0.03 (13)	0.97 (472)	0.01 (4)
	Rackete	0.02 (9)	0.98 (487)	0.00 (1)

TABLE 6.11: Distribution of **Aesthetic judgement** (first level; majority voting).

The overall results are not surprising in the light of the pre-annotation considerations—as in the researchers' pilot it was already noticed that aesthetic judgements did not appear so often as expected—but the subset pattern is still informative in that the difference between the two events remains small.

Therefore, methodologically this result can help us focus on two aspects. First, designing a schema—in particular the choice of which dimensions are useful for the task—is necessarily influenced by the researcher bias and world knowledge, and this facet of positionality needs to be heeded. Second, a fine-grained multi-label annotation schema (most importantly in the first phases of the process) bears the capacity to distinguish dimensions that structure the discourse (e.g. stance, tone, offensiveness, responsibility) and dimensions that are only sporadically activated, thereby helping to avoid over-interpreting rare evaluative phenomena.

6.2.2.2.9 Stereotype At the first level, stereotypical content represents a minority of the overall discourse (*Yes*: 0.10; n = 100), as most tweets are annotated as not containing stereotypes (*No*: 0.86; n = 846). The **Rackete** subset shows a higher proportion of stereotypes (0.13; n = 66) than the **Schettino** subset (0.07; n = 34).

Dimension	Set	Yes	No	Idk	Tie
Stereotype	**Pre_Dataset**	0.10 (100)	0.86 (846)	0.00 (2)	0.04 (38)
	Schettino	0.07 (34)	0.89 (434)	0.00 (0)	0.04 (21)
	Rackete	0.13 (66)	0.83 (412)	0.00 (2)	0.03 (17)

TABLE 6.12: Distribution of **Stereotype** (first level; majority voting).

At second level, stereotype targets must be interpreted event-specifically, as the available options differed across subsets (Schettino vs. Rackete) during annotation. This prevents a fully symmetric cross-case comparison at the label level, but it also makes the within-subset distributions more informative about which social categories were salient in each discursive context.

When analysing stereotype targets, both subsets frequently involve group-based categories connected to the event narrative. In the **Schettino** subset, the most frequent targets are *women* (0.44; n = 15) and *Eastern European women* (0.32; n = 11), suggesting that stereotyping is often anchored in gendered and nationality-indexed characterisations of secondary female figures. By contrast, in the **Rackete** subset, stereotypes most frequently target *migrants* (0.47; n = 31), followed by *women* (0.35; n = 23) and *Africa/people from Africa* (0.09; n = 6), indicating a stronger activation of migration-related and origin-based group framings. In both subsets, the recurrent selection of *other* (Schettino: 0.38; n = 13; Rackete: 0.45; n = 30) signals that a non-trivial share of stereotypical content falls outside the predefined taxonomy in this preliminary schema.

Targets of Stereotype (Schettino)	**Proportion**
Men	0.15 (5)
Male captains	0.06 (2)
Women	0.44 (15)
East. Europ. women	0.32 (11)
Cruising	0.06 (2)
Other	0.38 (13)

TABLE 6.13: Targets of **Stereotype** in the **Schettino** subset, computed only where Stereotype = *Yes* (second level; multi-label selection allowed).

Targets of Stereotype (Rackete)	**Proportion)**
Women	0.35 (23)
Female captains	0.02 (1)
Migrants	0.47 (31)
Africa/people from Africa	0.09 (6)
NGOs	0.00 (0)
Other	0.45 (30)

TABLE 6.14: Targets of **Stereotype** in the **Rackete** subset, computed only where Stereotype = *Yes* (second level; multi-label selection allowed).

Regarding stereotype spheres, *behaviour/attitude* dominates in both subsets (**Pre_Dataset**: 0.47; n = 47), followed by *origin/nationality* (0.31; n = 31). ***Pre_-Rackete*** shows a higher emphasis on origin-related stereotypes (0.36; n = 24 vs. 0.21; n = 7), whereas ***Pre_Schettino*** includes comparatively more sexuality-related stereotyping (0.24; n = 8 vs. 0.09; n = 6). Interestingly, the option *other* is the second-chosen label both overall (0.34; n = 34) and in the two subsets (***Pre_Schettino***: 0.26, n = 9; ***Pre_Rackete***: 0.38, n = 25).

Spheres of Stereotype	***Pre_Schettino* (34)**	***Pre_Rackete* (66)**	**Pre_Dataset (100)**
Body	0.18 (6)	0.09 (6)	0.12 (12)
Behaviour/attitude	0.50 (17)	0.45 (30)	0.47 (47)
Activity/role	0.09 (3)	0.15 (10)	0.13 (13)
Sexuality	0.24 (8)	0.09 (6)	0.14 (14)
Origin/nationality	0.21 (7)	0.36 (24)	0.31 (31)
Other	0.26 (9)	0.38 (25)	0.34 (34)

TABLE 6.15: Spheres of **Stereotype** (second level; computed only where Stereotype = *Yes*). Multi-label selection allowed.

Intersecting these data with Table 6.13 and Table 6.14, the prominence of behavioural and origin-related spheres for both subset, and the relevance of *sexuality* in **Schettino** supports the view that stereotypes in this corpus are triggered asymmetrically in the two subsets. In **Schettino**, a closer check to the data reveals that a substantial part of stereotypes regarding *behaviour* are associated with women several times, most of them developed around the figure of Domnica Cemortan. The higher proportion of *sexuality* in the

Schettino subset seems to confirm this, which at the same time seems to make the pair with the high proportion of the label *Eastern European women* as target.

Finally, the recurrent selection of *other* suggests that the current taxonomy may not fully capture the range of stereotypical framings activated in politically charged online discourse, pointing to potential refinements in future iterations.

6.2.2.3 IAA (Pre_Dataset)

IAA was computed on the complete set of 1,000 tweets annotated by the three members of the annotation team. Agreement was measured using Fleiss' κ (see Landis and Koch, 1977). All values reported in this section are derived from the full 1,000-tweets dataset, thus including tweets labelled as *non-classifiable*, to offer a complete picture.

Following conventional interpretative benchmarks, κ values between 0.21 and 0.40 are interpreted as indicating *fair* agreement, while values between 0.41 and 0.60 indicate *moderate* agreement. Lower values are expected for dimensions involving highly subjective evaluative judgement or requiring contextual interpretation.

6.2.2.3.1 First-level dimensions Agreement varies across dimensions. The highest values are observed for ***Humour*** ($\kappa = 0.438$) and ***Target*** ($\kappa = 0.420$), both reaching the moderate range. ***Stance*** ($\kappa = 0.335$), ***Tone of speech*** ($\kappa = 0.235$), and ***Professional competence*** ($\kappa = 0.235$) fall within the fair range.

Lower agreement characterises ***Offensiveness*** ($\kappa = 0.196$), ***Responsibility*** ($\kappa = 0.158$), ***Stereotype*** ($\kappa = 0.139$), and ***Aesthetic judgement*** ($\kappa = 0.085$), all of which require interpretative or evaluative judgement. The slightly negative value for ***Non-classifiable*** ($\kappa = -0.027$) that decisions about exclusion were particularly unstable across annotators.[4]

[4]This may reflect the difficulty of delimiting the boundary between problematic yet classifiable cases and cases that were genuinely impossible to annotate.

First-level dimensions	**Fleiss' κ**
Non-classifiable	-0.027
Target	0.420
Responsibility	0.158
Offensiveness	0.196
Stance	0.335
Tone of speech	0.235
Humour	0.438
Professional competence	0.235
Aesthetic judgement	0.085
Stereotype	0.139

TABLE 6.16: Fleiss' κ values for first-level dimensions computed on the full 1,000-tweet dataset (three annotators; *non-classifiable* included).

Overall, dimensions grounded in relatively explicit textual cues display greater stability, whereas socially and pragmatically situated categories show greater dispersion.

Table 6.17 reports first-level agreement computed separately for the Schettino and Rackete subsets.

First-level dimensions	**Schettino**	**Rackete**
Non-classifiable	-0.038	-0.016
Target	0.410	0.357
Responsibility	0.072	0.236
Offensiveness	0.067	0.319
Stance	0.158	0.356
Tone of speech	0.147	0.076
Humour	0.345	0.324
Professional competence	0.168	0.195
Aesthetic judgement	0.060	0.121
Stereotype	0.067	0.193

TABLE 6.17: Fleiss' κ values for first-level dimensions computed separately for Schettino and Rackete (three annotators; *non-classifiable* included).

The subset comparison reveals two differentiated agreement profiles. In ***Pre_Schettino***, higher agreement emerges for ***Target*** (0.410 vs. 0.357), ***Tone of speech*** (0.147 vs. 0.076), and ***Humour*** (0.345 vs. 0.324), suggesting relatively greater stability in annotating interactional and protagonist-centred dimensions. In contrast, ***Pre_Rackete*** shows substantially higher agreement for ***Responsibility*** (0.236 vs. 0.072), ***Offensiveness*** (0.319 vs. 0.067), ***Stance*** (0.356 vs. 0.158), and ***Stereotype*** (0.193 vs. 0.067), indicating that evaluative and politicised dimensions are more consistently identified in that discourse. These differences align with the descriptive statistics and support the interpretation of agreement as partially event-sensitive rather than purely annotation-driven.

6.2.2.3.2 Second-level dimensions Agreement for second-level dimensions was computed only on tweets where the corresponding first-level dimension was annotated as present.

Targets of *Responsibility* Agreement remains predominantly within the slight range. Responsibility targets display low stability overall, with marked asymmetries across subsets. In Rackete, agreement is relatively balanced but slightly higher for *Politics* (0.205), *Main character* (0.203), and *Society* (0.200). In Schettino, convergence is slightly better for *Organisation* (0.206) and *Society* (0.154), whereas *Main character* yields negative agreement (-0.170), signalling systematic annotator divergence when attempting to isolate the protagonist as a specific target among multiple options. The value $\kappa = 1.000$ for *Victims* in Schettino is compatible with a degenerate distribution (systematic agreement on absence within the filtered set) and should not be interpreted as strong discriminative reliability.

Targets of Responsibility	Schettino	Rackete
Main character	−0.170	0.203
Organisation	0.206	0.052
Institutions	−0.009	0.118
Politics	0.075	0.205
Society	0.154	0.200
Victims	1.000	−0.009
Other	0.048	−0.018

TABLE 6.18: Fleiss' κ values for second-level targets of *Responsibility*, computed separately for Schettino and Rackete (only where Responsibility = *Yes*).

Targets of *Offensiveness* Offensiveness targets show clearer differentiation across events. In Rackete, agreement reaches moderate levels for *Victims* (0.449) and *Main character* (0.441), remaining substantial for *Society* (0.359) and *Politics* (0.327), indicating a rather stable identification of the actors involved in the politicised debate. In Schettino, convergence is highest for offensiveness directed at *Society* (0.407) and *Organisation* (0.235), while agreement on the *Main character* is surprisingly low (0.061), reflecting potential ambiguity in differentiating insults directed solely at the captain from broader discursive attacks. As above, $\kappa = 1.000$ for *Victims* in Schettino likely reflects very low prevalence.

Targets of Offensiveness	Schettino	Rackete
Main character	0.061	0.441
Organisation	0.235	0.256
Institutions	−0.020	0.138
Politics	0.219	0.327
Society	0.407	0.359
Victims	1.000	0.449
Other	0.217	0.196

TABLE 6.19: Fleiss' κ values for second-level targets of *Offensiveness*, computed separately for Schettino and Rackete (only where Offensiveness = *Yes*).

Targets of *Humour* Humour targets further highlight discursive differences. In Rackete, agreement reaches moderate levels for humour directed at the *Main character* (0.462) and *Institutions* (0.432). Conversely, in ***Pre_Schettino***, the highest convergence appears for *Politics* (0.349), while agreement on the *Main character* is remarkably low (0.064). This suggests that although humour is highly prevalent overall in ***Pre_Schettino*** at the first level, annotators struggled to systematically agree on the protagonist as its primary target when multi-label selection was active, possibly due to the pervasive and diffuse nature of the irony in that specific discourse.

Targets of Humour	**Schettino**	**Rackete**
Main character	0.064	0.462
Organisation	0.242	0.327
Institutions	0.081	0.432
Politics	0.349	0.151
Society	0.240	0.121
Victims	−0.004	0.327
Other	0.229	0.104

TABLE 6.20: Fleiss' κ values for second-level targets of *Humour*, computed separately for Schettino and Rackete (only where Humour = *Yes*).

Targets of *Stereotype* Agreement for stereotype targets is predominantly in the slight to fair range, reflecting the implicit nature of stereotyping in this dataset. In ***Pre_Schettino***, convergence is highest for *Eastern European women* (0.329), while the label *women* yields lower agreement (0.212), indicating interpretative divergence regarding the gendered framing of secondary figures. In ***Pre_Rackete***, agreement reaches the moderate threshold for *migrants* (0.423) and approaches it for *women* (0.348), suggesting a somewhat clearer identification of these social categories when stereotyping is activated. The perfect agreement ($\kappa = 1.000$) for *NGOs* in **Rackete** should not be interpreted as evidence of substantive convergence, since it is produced by the total absence of positive cases.

Targets of Stereotype (Schettino)	**Fleiss' κ**
Men	0.113
Male captains	−0.020
Women	0.212
East. European women	0.329
Cruising	−0.020
Other	0.292

TABLE 6.21: Fleiss' κ values for second-level targets of *Stereotype*, computed for ***Pre_Schettino*** (only where Stereotype = *Yes*).

Targets of Stereotype (Rackete)	**Fleiss' κ**
Women	0.348
Female captains	0.179
Migrants	0.423
Africa/people from Africa	0.328
NGOs	1.000
Other	0.266

TABLE 6.22: Fleiss' κ values for second-level targets of *Stereotype*, computed separately for Schettino and Rackete (only where Stereotype = *Yes*).

Spheres of *Stereotype* Agreement across stereotype spheres is systematically low, frequently dropping near zero or into negative values (e.g., *Behaviour/attitude* in Rackete, −0.017). The only relative peaks occur for *Sexuality* in ***Pre_Rackete*** (0.263) and *Body* in ***Pre_Schettino*** (0.185). This profound interpretative dispersion confirms that while annotators may partially agree on who is being stereotyped, they find it extremely difficult to reliably classify the semantic domain of the stereotype. The boundary between a stereotype (e.g., related to behaviour) and a general negative evaluation or personal attack is highly ambiguous in online discourse, making this dimension intensely susceptible to subjective interpretation and potential over-interpretation.

Spheres of Stereotype	Schettino	Rackete
Body	0.185	0.113
Behaviour/attitude	0.019	−0.017
Activity/role	0.051	0.054
Sexuality	0.006	0.263
Victims' origin	0.051	0.063
Other	0.089	0.016

TABLE 6.23: Fleiss' κ values for second-level spheres of *Stereotype*, computed separately for Schettino and Rackete (only where Stereotype = *Yes*).

6.2.2.4 MR-Annotators Comparison (600 tweets)

In addition to the agreement analysis conducted on the full set of 1,000 tweets annotated by the three annotators, an overlapping subset of 600 tweets was independently annotated by the MR. This configuration makes it possible to compare agreement between the MR and each annotator, measured through Cohen's κ (see Section 6.2.2.1), with the agreement patterns already observed among the three annotators in the 1,000-tweet dataset.

Agreement between the MR and each annotator was computed separately for each annotation dimension on the 600-tweet overlap. These values were then read against the pairwise and overall agreement patterns observed in the annotation team's 1,000-tweet dataset, in order to assess whether the researcher's inclusion introduced systematic divergence or recurrent dimension-specific alignment.

Across first-level dimensions, agreement between the MR and each annotator falls broadly within the same range as the agreement observed among the annotators themselves in the 1,000-tweet dataset. No dimension shows a clear and systematic drop that could be attributed to the researcher's inclusion. Dimensions such as ***Target*** and ***Humour*** remain among the more stable ones, whereas ***Responsibility***, ***Offensiveness***, ***Aesthetic judgement***, and ***Stereotype*** continue to show lower and more variable agreement, much as they do in the team-only analysis.

Table 6.24 reports the exact pairwise agreement scores between the MR and each annotator for all first-level dimensions.

First-level dimension	MR–Annotator1	MR–Annotator2	MR–Annotator3
Non-classifiable	−0.042	−0.053	−0.045
Target	0.391	0.416	0.423
Responsibility	0.187	0.152	0.225
Offensiveness	−0.006	0.201	0.097
Stance	0.284	0.242	0.379
Tone of speech	0.336	0.160	0.037
Humour	0.374	0.345	0.385
Professional competence	0.275	0.127	0.331
Aesthetic judgement	−0.096	−0.137	0.006
Stereotype	0.022	0.078	0.160

TABLE 6.24: Pairwise Cohen's κ values between the MR and each of the three annotators on first-level dimensions, computed on the 600-tweet subset.

At second level, agreement remains strongly dimension-dependent. As in the 1,000-tweet dataset, ***Spheres of Stereotype*** constitute the most interpretatively unstable layer. At the same time, the inclusion of the researcher does not produce a generalised worsening of agreement, nor does it point to a stable interpretative bloc involving the MR. What emerges, rather, is a small number of localised convergences on specific dimensions.

These local peaks are summarised in Table 6.25.

Dimension	MR-Annotator	Cohen's κ
Spheres of Stereotype – Body	MR–Annotator3	0.493
Spheres of Stereotype – Behaviour	MR–Annotator3	0.323
Tone of speech	MR–Annotator1	0.336
Targets of Humour – Organisation	MR–Annotator1	0.431

TABLE 6.25: Selected peaks of MR-annotator agreement in the 600-tweet subset.

The value of $\kappa = 0.493$ for *Spheres of Stereotype – Body* is among the highest agreements observed in the MR-annotator comparison and approaches the moderate range. By contrast, *Spheres of Stereotype – Behaviour*, although showing a relatively stronger convergence between the MR and Annotator3,

remains within the fair range and is consistent with the broader instability of this layer.

The same applies to the alignment between the MR and Annotator1 on ***Tone of speech*** and ***Targets of Humour*** – *Organisation*. These are dimension-specific convergences rather than evidence of a broader clustering effect. Across the section, agreement varies by annotation layer and by category, not in a way that would suggest a systematically distinct profile on the part of the researcher.

Overall, agreement between the MR and each annotator is comparable in magnitude and distribution to the agreement observed among the annotators themselves. The researcher's inclusion does not appear to alter the general structure of agreement, nor to introduce a qualitatively different pattern of interpretation.

The comparison between the 1,000-tweet dataset and the 600-tweet overlap nevertheless remains informative. First, the relative ordering of dimensions is preserved. ***Target*** and ***Humour*** remain easier to annotate, whereas ***Responsibility***, ***Offensiveness***, ***Stereotype***, and especially ***Spheres of Stereotype*** remain more unstable. Second, where some dimensions show slightly higher or lower values in the 600-tweet subset, these fluctuations stay within ranges already visible in the team-only analysis. Third, the more noticeable cases of convergence do not accumulate around a single annotator or around the MR across dimensions, which suggests that interpretative proximity is local rather than person-specific.

From the point of view of researcher positionality, this result is methodologically relevant. If the annotation schema merely reflected an idiosyncratic framing imposed by its designer, one might expect the MR either to diverge systematically from the annotators or to align with them only through a rigid application of the scheme. This is not what emerges here. The more difficult dimensions remain difficult for the researcher as well, and the same general areas of fluctuation reappear. This suggests that at least part of the variability is tied to the interpretative difficulty of the task itself rather than to the researcher's presence alone.

Subset-specific dynamics point in the same direction. When MR-annotator agreement is examined separately for the two events, the broader trends already observed among the annotators reappear. In particular, *Stance* remains

easier to identify in ***Pre_Rackete*** and more unstable in ***Pre_Schettino***. Agreement between the MR and individual annotators reaches higher values in the Rackete subset (up to $\kappa = 0.501$ with Annotator1 and $\kappa = 0.491$ with Annotator3), while it drops substantially in Schettino, in some cases even below zero (e.g. $\kappa = -0.214$ with Annotator1). This pattern should be interpreted cautiously, but it supports the idea that the discursive properties of the two events—namely, the clearer ideological polarisation surrounding Rackete and the more pervasive irony characterising Schettino—weigh more heavily on agreement than the specific theoretical position of the schema designer.

In general, these results support a cautious conclusion: the researcher's inclusion does not seem to generate a distinct agreement profile, whereas the main areas of instability remain the same ones already identified in the annotation team's work. In this sense, the comparison strengthens the view that the most problematic dimensions are structurally difficult to annotate, and that their instability cannot be attributed only to who designed the schema.

6.2.2.5 Preliminary discussion

Overall, these preliminary findings suggest that agreement structure is primarily shaped by the semantic and pragmatic complexity of the annotation dimensions rather than by the positionality of the researcher. Dimensions requiring fine-grained semantic discrimination (such as stereotype spheres) remain inherently more unstable, whereas dimensions relying on more explicit textual cues display greater stability across annotators and researcher alike. Crucially, this instability is not uniformly distributed but heavily event-dependent: both the independent annotators and the schema designer experienced the exact same interpretative friction when confronted with the pervasive, blurring irony of ***Pre_Schettino***, which starkly contrasted with the more explicit ideological polarisation of ***Pre_Rackete***.

From a methodological perspective, the comparison supports the robustness of the annotation schema: the researcher's participation does not introduce structural distortion of agreement patterns, and observed divergences reflect the intrinsic interpretative demands of the categories rather than systematic bias. Far from indicating a failure of the coding guidelines, these

systematic divergences align with the emerging paradigm of Data Perspectivism, confirming that phenomena like offensiveness and stereotyping are highly context-sensitive constructs subject to legitimate human label variation.

With this in mind, we prepared the outline for a semi-structured group interview and organised the Feedback Meeting, aiming to qualitatively explore these areas of systemic disagreement, understand the cognitive heuristics employed by the annotators, and collaboratively refine the annotation framework without forcing an artificial consensus.

6.2.3 Feedback Meeting

The Feedback Meeting was conducted in the form of a focus group interview, led by the MR. The discussion followed a predefined set of questions aimed at eliciting (i) a general evaluation of the annotation experience, (ii) detailed feedback on each category of the schema, (iii) an assessment of the clarity and adequacy of the guidelines, and (iv) any additional remarks emerging from the annotators' practice.

The meeting opened with an overall assessment of the annotation experience in terms of perceived difficulty, workload and cognitive effort. All annotators rated the task as **4 out of 5** on an ease scale (1 = very difficult, 5 = very easy). They reported that the first phase of exposure to the schema and the interface required a higher cognitive load due to the necessity of internalising the categories and annotation routines. After annotating the first few dozens of tweets, however, the process became progressively more routinised. Annotators also converged on the idea that annotating approximately 30–40 tweets per day constituted a sustainable rate, reducing fatigue and fostering consistency. A recurrent observation concerned the structural asymmetry between the two sub-corpora: tweets related to Schettino were described as shorter, more repetitive, and strongly oriented towards humour and "meme-like" content, while tweets about Rackete tended to be longer, more information-rich and thematically denser, often requiring multiple readings to be fully understood. This difference was attributed both to the evolution of Twitter's character limits over time and to the more politically and morally charged nature of the Rackete case.

Following the interview outline, the MR guided the annotators through a systematic reflection on the categories of the annotation schema.

With regard to the label ***non-classifiable***, it was used only in a very small percentage of cases (approximately 1–2% of tweets). Annotators reported employing it mainly for posts that referred to homonymous individuals (e.g. a football player named "Schettino") or for tweets written in languages they could not understand.

Concerning the selection of the ***target***, annotators reported persistent ambiguity in distinguishing between *secondary* and *other* actors. Several political actors and collective entities were recurrently misidentified as secondary targets despite not being listed among the predefined secondary figures. This feedback motivated the expansion and clarification of the list of secondary targets in the guidelines and the recommendation to resort to the *other* option whenever uncertainty arose.

The ***responsibility*** category was generally described as conceptually meaningful but challenging. Annotators noted difficulties in tweets where responsibility was simultaneously acknowledged and morally mitigated by the tweet's author (particularly in the **Rackete** subset). They agreed on treating implicit attributions of responsibility in the same way as explicit ones, yet several "borderline" cases emerged, particularly in the **Rackete** subset. For tweets that acknowledged that the protagonist had violated a rule while at the same time morally justifying her behaviour ("she is responsible, but she did it for a good reason"), some annotators tended to hesitate between *yes*, *no* and *I don't know*. They also pointed out the frequent "duality" of tweets, where a negative judgement on one actor was followed by a mitigating comparison with others. In the discussion, we agreed to clarify in the guidelines that responsibility can be annotated independently of its moral valence (i.e. both "blame" and "accountability without condemnation" fall under responsibility) and that explicit statements should be prioritised when present. We considered but ultimately discarded the idea of adding an additional sub-label distinguishing implicit vs. explicit responsibility, as this was judged to increase complexity without providing a clear added value at this stage.

By contrast, ***offensiveness*** was unanimously perceived as intuitively easy, especially in overt cases. Annotators found overt insults and degradative expressions "*lampanti*" ("lampant"), although they also discussed more ambiguous cases such as "marcisci in galera!" (eng. "rot in jail!" referred to

Schettino), which prompted reflection on whether invectives aimed at punishment rather than at personal attributes should be considered offensive. Notably, annotators perceived more direct and serious forms of verbal hostility in the **Rackete** subset, while humour in the **Schettino** subset often attenuated expressions that would otherwise be coded as aggressive. This observation further confirmed the strong interaction between offensiveness, stance and humour.

The ***stance*** category (*pro / against / neutral*) did not raise major issues. Annotators indicated that neutral stance was mostly assigned to tweets reproducing news headlines and contents or factual statements. Some initial ambiguity emerged when tweets expressed strong positions towards secondary actors rather than the protagonist. The discussion clarified that stance must always refer to the main character, irrespective of evaluations directed at other figures.

The category that generated the greatest uncertainty was ***tone***. Annotators reported systematic difficulty in identifying a clear boundary between *light* and *serious* and, to a lesser extent, between *serious* and *neutral*. Many tweets employed apparently serious stylistic resources to convey humorous or facetious content, thus producing ambiguity. Annotators expressed the impression that tone was largely redundant with the combination of stance and offensiveness, and that they often annotated it "*con un senso di colpa*" ("with a sense of guilt"), choosing one of the options while being unsure that it captured the pragmatic nuance of the tweet. In general, ***tone*** was identified as the most fragile dimension of the schema and a candidate for simplification or removal in the following revision, even though it was considered to be maintained for the second annotation block in order not to disrupt the annotators' acquired routine. This perception was actually consistent with the relatively lower inter–annotator agreement observed for this dimension.

The ***humour*** category, instead, was reported as straightforward to annotate, especially in the **Schettino** subset, where humorous and derisive content was pervasive. Doubts arose only in a minority of cases, typically when tweets contained "inside jokes" or were replies to other posts whose context was not visible in the interface; in such situations annotators occasionally resorted to the *I don't know* option. The Meeting confirmed the usefulness of humour as a separate dimension, particularly for capturing implicit stereotypes and

attenuated forms of hostility that would otherwise escape a purely stance-based or offensiveness-based analysis.

A similar pattern was described for ***professional competence***: annotators tended to derive their judgements (*positive / negative / neutral / not applicable*) directly from stance and offensiveness, considering competence almost "automatic" once the other dimensions had been annotated. Only in cases where the tweet focused primarily on secondary actors they report spending more time assessing whether an indirect evaluation of the main protagonist's competence was present.

The categories ***aesthetic judgement*** and ***stereotype*** emerged as particularly informative during the Feedback Meeting. With regard to ***aesthetic judgement***, annotators confirmed its very low frequency in the dataset. Contrary to initial expectations, they reported slightly more aesthetic remarks concerning Schettino (e.g. comments on baldness) than Rackete (e.g., references to dreadlocks or hair), suggesting that physical appearance was only marginally involved in the discursive construction of both protagonists.

With respect to ***stereotypes***, annotators reported that the category had been initially underused. Prior to the Meeting, they tended to identify stereotypes only when explicitly gendered or strongly conventionalised. Through collective reflection, however, annotators recognised a broader set of stereotypical contents, including implicit stereotypes and those directed at nationality-based or regional groups (e.g., references to Germans as "crucchi", stereotypes concerning Southern Italians), as well as stereotypes involving secondary actors such as Domnica. The discussion also brought to light several practical issues requiring revision: the need for clearer guidance—or a dedicated option—concerning stereotypes about "society" or national groups; the limited use of certain predefined secondary actors (e.g. NGOs); partial overlaps among some stereotype subcategories (e.g. *activity, role, behaviour*). Annotators therefore agreed to maintain the overall structure of the stereotype dimension while refining and expanding the examples in the guidelines, encouraging more systematic use of the *other* stereotype option, and considering the consolidation of overlapping labels in subsequent phases.

In response to the MR's prompts regarding the **clarity of the guidelines**, annotators confirmed that the document was sufficiently detailed and generally unambiguous. They suggested targeted improvements, such as: (i)

expanding the section on stereotypes; (ii) reorganising the list of targets to improve usability; (iii) providing additional examples for borderline cases; and (iv) refining the explanations of certain dimensions for greater clarity. No major structural revisions were deemed necessary.

Beyond category-specific considerations, the Meeting also provided insight into the organisation of the second annotation block. Annotators agreed that the schema, despite its inherent complexity, was workable and did not require substantial modifications. They primarily requested minor adjustments—such as broadening the list of secondary actors and clarifying specific definitions—as well as a clearer scheduling framework for the annotation workflow. Annotators reported that annotating approximately 30–40 tweets per day—ideally supported by intermediate milestones (e.g. blocks of 300–400 tweets)—made the task manageable and helped mitigate fatigue.

The Meeting concluded with an open discussion. Annotators highlighted the emotional impact of certain tweets (particularly those concerning migration), the usefulness of alternating sets of ten tweets per protagonist, and the importance of regular check-ins to avoid the accumulation of unresolved ambiguities.

Overall, the Feedback Meeting confirmed the soundness and usability of the annotation schema while identifying specific points requiring refinement. These organisational and methodological considerations were incorporated into the planning of the second annotation step, and the feedback directly informed the updated guidelines adopted for the subsequent annotation block.

6.2.3.1 Second Revision of Schema and Guidelines

On the basis of the preliminary analysis and the Feedback Meeting, we carried out a second, targeted revision of the annotation schema and its guidelines. By systematically cross-checking the quantitative results (inter-annotator agreement and descriptive statistics) with the qualitative feedback provided by the annotators, we identified those dimensions that were perceived as unclear or redundant and those that, conversely, proved to be informative and robust. The outcome of this process was a new annotation schema—in which some marginal or less stable dimensions were removed—and a slightly simplified

and clearer version of the guidelines, where several definitions and examples were refined (see Appendix C).

The annotation interface was updated on the same platform used for the first block. The research team adapted the platform to the new set of labels, and annotators were asked to verify that all categories, labels and examples appeared correctly and were functioning before proceeding with the second block. All changes with respect to the previous version of the guidelines were explicitly highlighted in a document that was shared with the annotators—so as to facilitate the comparison with the old version—even if then they still continued having access to the updated guidelines directly on the platform.

From an organisational perspective, we also adjusted both the size and the scheduling of the second annotation block. The total number of tweets to be annotated in this phase was reduced to 1,000 (instead of the approximately 1,200 initially envisaged), in order to maintain a sustainable workload given the complexity of the task. Three internal deadlines were established. With the reduced total, the first two deadlines corresponded to 350 annotated tweets each, while the final deadline covered the remaining 300 tweets.

All information regarding the revised guidelines, the platform update, and the internal schedule was communicated to annotators via email, and the MR committed to sending reminder messages a few days before each deadline, in line with the collaborative and iterative principles that characterised the entire annotation process.

6.2.4 Second Block of Annotation

The second block, consisting of different 1,000 tweets, was annotated by the same three annotators. They were given approximately three months to complete this second block using the revised post-Feedback Meeting schema presented above, which constitutes the definitive annotation framework adopted for the subsequent analyses (see Appendix C).

The use of the same annotation team and an equally sized second block makes it possible to read the results of this phase in continuity with the preliminary one, while keeping in mind that the comparison is not based on the same tweets and must therefore remain cautious.

6.2.4.1 Corpus Statistics (Fin_Dataset)

This section reports descriptive statistics for the second block of annotation, consisting of 1,000 tweets annotated according to the revised schema defined after the Feedback Meeting. Tweets labelled as *non-classifiable* were excluded prior to aggregation; therefore, corpus statistics are computed on a total of 984 tweets (487 for **Schettino** and 497 for **Rackete**), aggregated through majority voting.

Whenever relevant, the distributions reported below are briefly read against those observed in the preliminary block, so as to highlight which tendencies remain stable across phases and which appear more sensitive to the schema revision.

As in the preliminary dataset, when annotators did not reach a majority decision the outcome was recorded as a *tie*. All proportions reported below correspond to the ratio between the number of tweets assigned to a given category and the total number of tweets within the relevant subset.

6.2.4.1.1 Target Table 6.26 reports the distribution of the **Target** dimension. In the revised schema the *other* option was removed, and annotators selected among *main*, *secondary*, and *figurative*. The results show a similar pattern compared to the Pre_Dataset as to the figurative target, which is substantially more frequent in ***Fin_Schettino***, whereas secondary targets are comparatively more common in ***Fin_Rackete***.

Set	Main	Secondary	Figurative	Tie
Fin_Overall	0.351 (345)	0.228 (224)	0.272 (268)	0.149 (147)
Schettino	0.326 (159)	0.084 (41)	0.454 (221)	0.136 (66)
Rackete	0.374 (186)	0.368 (183)	0.095 (47)	0.163 (81)

TABLE 6.26: Distribution of the **Target** dimension in the final dataset of 984 tweets (majority voting; first-level dimension).

The subset comparison confirms different discursive configurations: while Schettino tweets frequently involve figurative targeting, ***Fin_Rackete*** exhibits a more balanced distribution between main and secondary targets.

6.2.4.1.2 Responsibility Table 6.27 reports the first-level distribution of Responsibility.

Set	Yes	No	Idk	Tie
Fin_Overall	0.369 (363)	0.488 (480)	0.007 (7)	0.136 (134)
Schettino	0.526 (256)	0.298 (145)	0.004 (2)	0.172 (84)
Rackete	0.215 (107)	0.674 (335)	0.010 (5)	0.101 (50)

TABLE 6.27: Distribution of **Responsibility** in the final dataset (first-level dimension).

Responsibility is markedly more frequent in the **Schettino** subset, where over half of the tweets attribute responsibility to some actor. In contrast, the **Rackete** subset is characterised by a substantially higher proportion of tweets where responsibility is not explicitly attributed.

Table 6.28 reports the distribution of targets selected when the **Responsibility** dimension was annotated as present. Proportions are again computed over the subset of tweets where Responsibility = *yes*.

Targets of Responsibility	Fin_Overall	Schettino	Rackete
Main char.	0.981 (356)	0.996 (255)	0.944 (101)
Organisation	0.022 (8)	0.020 (5)	0.028 (3)
Institutions	0.003 (1)	0.000 (0)	0.009 (1)
Politics	0.030 (11)	0.004 (1)	0.093 (10)
Society	0.000 (0)	0.000 (0)	0.000 (0)
Victims	0.000 (0)	0.000 (0)	0.000 (0)
Other	0.017 (6)	0.020 (5)	0.009 (1)

TABLE 6.28: Targets selected for **Responsibility** (second-level; multi-label selection allowed).

Responsibility attribution overwhelmingly concerns the *main character*. This pattern is particularly strong in ***Fin_Schettino***, where responsibility is almost exclusively directed towards the protagonist, whereas ***Fin_Rackete*** displays slightly more diversification, including political actors.

6.2.4.1.3 Offensiveness Table 6.29 reports the distribution of Offensiveness.

Set	Yes	No	Idk	Tie
Fin_Overall	0.235 (231)	0.584 (575)	0.004 (4)	0.177 (174)
Schettino	0.170 (83)	0.612 (298)	0.004 (2)	0.214 (104)
Rackete	0.298 (148)	0.557 (277)	0.004 (2)	0.141 (70)

TABLE 6.29: Distribution of **Offensiveness** in the final dataset (first-level dimension).

Offensive language appears more frequently in ***Fin_Rackete***, whereas ***Fin_-Schettino*** contains a larger proportion of tweets classified as non-offensive and, similarly to the Pre_dataset, a larger portion of *ties*. .

Table 6.30 reports the distribution of targets selected when **Offensiveness** was annotated as present.

Targets of Offensiveness	Fin_Overall	Schettino	Rackete
Main char.	0.498 (115)	0.771 (64)	0.345 (51)
Organisation	0.017 (4)	0.000 (0)	0.027 (4)
Institutions	0.022 (5)	0.000 (0)	0.034 (5)
Politics	0.355 (82)	0.000 (0)	0.554 (82)
Society	0.199 (46)	0.120 (10)	0.243 (36)
Victims	0.061 (14)	0.000 (0)	0.095 (14)
Other	0.143 (33)	0.265 (22)	0.074 (11)

TABLE 6.30: Targets selected for **Offensiveness** (second-level; multi-label selection allowed).

Offensive discourse shows markedly different targeting patterns across the two subsets. In the Schettino corpus, offensiveness is primarily directed towards the protagonist, whereas in ***Fin_Rackete*** it more frequently targets political actors and broader societal groups.

6.2.4.1.4 Stance Table 6.31 reports the distribution of stance labels.

Set	Pro	Neutral	Against	Tie
Fin_Overall	0.238 (234)	0.154 (152)	0.455 (448)	0.152 (150)
Schettino	0.012 (6)	0.195 (95)	0.655 (319)	0.138 (67)
Rackete	0.459 (228)	0.115 (57)	0.260 (129)	0.167 (83)

TABLE 6.31: Distribution of **Stance** in the final dataset (first-level dimension).

The two subsets display clearly divergent stance profiles. ***Fin_Schettino*** is dominated by tweets expressing an *against* stance, whereas ***Fin_Rackete*** contains a substantially higher proportion of *pro* tweets.

6.2.4.1.5 Humour Table 6.32 reports the distribution of Humour.

Set	Yes	No	Idk	Tie
Fin_Overall	0.361 (355)	0.517 (509)	0.008 (8)	0.114 (112)
Schettino	0.559 (272)	0.294 (143)	0.010 (5)	0.138 (67)
Rackete	0.167 (83)	0.736 (366)	0.006 (3)	0.091 (45)

TABLE 6.32: Distribution of **Humour** in the final dataset (first-level dimension).

Humour is considerably more frequent in ***Fin_Schettino***, where more than half of the tweets contain humorous elements. By contrast, ***Fin_Rackete*** is largely characterised by non-humorous discourse.

Table 6.33 reports the distribution of humour targets.

Targets of Humour	Fin_Overall	Schettino	Rackete
Main char.	0.800 (284)	0.934 (254)	0.361 (30)
Organisation	0.031 (11)	0.033 (9)	0.024 (2)
Institutions	0.011 (4)	0.004 (1)	0.036 (3)
Politics	0.132 (47)	0.026 (7)	0.482 (40)
Society	0.073 (26)	0.026 (7)	0.229 (19)
Victims	0.008 (3)	0.000 (0)	0.036 (3)
Other	0.256 (91)	0.279 (76)	0.181 (15)

TABLE 6.33: Targets selected for **Humour** (second-level; multi-label selection allowed).

Humour is strongly centred on the main character in ***Fin_Schettino***, confirming the results of the preliminary annotation. The same confirmation is observable also for ***Fin_Rackete***, where humorous discourse is more frequently oriented towards political actors and broader societal references. In this respect, the second block does not alter the preliminary picture but makes it more robust, as the same event-specific asymmetry reappears with very similar proportions and target distributions.

6.2.4.1.6 Professional competence Table 6.34 reports the distribution of Professional competence.

Set	Positive	Negative	Not applicable	Tie
Fin_Overall	0.132 (130)	0.453 (446)	0.210 (207)	0.204 (201)
Schettino	0.004 (2)	0.667 (325)	0.203 (99)	0.125 (61)
Rackete	0.258 (128)	0.243 (121)	0.217 (108)	0.282 (140)

TABLE 6.34: Distribution of **Professional competence** in the final dataset (first-level dimension).

Fin_Schettino is strongly dominated by negative evaluations of professional competence, whereas ***Fin_Rackete*** shows a more balanced distribution between positive and negative evaluations.

6.2.4.1.7 Stereotype Table 6.35 reports the distribution of the Stereotype dimension.

Set	Yes	No	Idk	Tie
Fin_Overall	0.079 (78)	0.790 (777)	0.005 (5)	0.126 (124)
Schettino	0.060 (29)	0.817 (398)	0.000 (0)	0.123 (60)
Rackete	0.099 (49)	0.763 (379)	0.010 (5)	0.129 (64)

TABLE 6.35: Distribution of **Stereotype** in the final dataset (first-level dimension).

Across the entire dataset, stereotypes are relatively infrequent, appearing in fewer than one tenth of the tweets. ***Fin_Rackete*** contains a slightly higher proportion of stereotypical references compared to ***Fin_Schettino***.

Targets of Stereotype (Schettino)	**Proportion**
Men	0.103 (3)
Male captains	0.000 (0)
Women	0.345 (10)
East. European women	0.414 (12)
Cruising experience	0.000 (0)
Society	0.207 (6)
Other	0.103 (3)

TABLE 6.36: Targets of **Stereotype** in ***Fin_Schettino*** (second-level; denominator = 29).

Targets of Stereotype (Rackete)	**Proportion**
Women	0.408 (20)
Female captains	0.061 (3)
Men	0.020 (1)
Migrants	0.306 (15)
Africa/people from Africa	0.163 (8)
NGOs	0.000 (0)
Society	0.265 (13)
Other	0.286 (14)

TABLE 6.37: Targets of **Stereotype** in ***Fin_Rackete*** (second-level; denominator = 49).

The two subsets confirm the activation of different stereotype configurations. ***Fin_Schettino*** predominantly involves gender-related stereotypes concerning women and Eastern European victims, whereas ***Fin_Rackete*** shows a stronger presence of stereotypes concerning migrants and broader social categories.

Table 6.38 reports the distribution of stereotype spheres.

Spheres of Stereotype	Fin_Overall	Schettino	Rackete
Body (incl. sexuality)	0.372 (29)	0.690 (20)	0.184 (9)
Behaviour/attitude	0.551 (43)	0.517 (15)	0.571 (28)
Activity/role	0.128 (10)	0.103 (3)	0.143 (7)
Origin/nationality	0.333 (26)	0.241 (7)	0.388 (19)
Other	0.064 (5)	0.034 (1)	0.082 (4)

TABLE 6.38: Distribution of **Stereotype spheres** (second-level; multi-label selection allowed).

Across the corpus, stereotypes most frequently concern behavioural traits and perceived social attitudes. Origin-related stereotypes are also common, particularly in ***Fin_Rackete***, whereas stereotypes related to the sphere *body/sexuality* are more prevalent in ***Fin_Schettino***.

6.2.4.2 IAA (Fin_Dataset)

IAA was computed on the complete set of 1,000 tweets annotated by the three annotators. Agreement was measured using Fleiss' κ, which accounts for chance agreement among multiple raters. All values reported below refer exclusively to the three annotators and were calculated on the full dataset, including tweets labelled as *non-classifiable*.

These values should also be read in relation to those obtained for the preliminary block, not in order to establish a strict before-and-after benchmark, but to assess whether the revision preserved the same broad areas of stability and instability across dimensions.

6.2.4.2.1 First-level dimensions Table 6.39 reports Fleiss' κ values for all first-level dimensions in the final schema. Table 6.40 reports the same values computed separately for the Schettino and Rackete subsets.

First-level dimensions (Fin_Overall)	**Fleiss' κ**
Non-classifiable	-0.093
Target	0.308
Responsibility	0.060
Offensiveness	0.004
Stance	0.235
Humour	0.266
Professional competence	0.168
Stereotype	0.039

TABLE 6.39: Fleiss' κ values for first-level dimensions computed on the complete second-block of 1,000 tweets (three annotators; *non-classifiable* included).

Overall agreement varies considerably across dimensions. The highest values are observed for *Target* ($\kappa = 0.308$) and *Humour* ($\kappa = 0.266$), both reaching the *fair* agreement range. *Stance* ($\kappa = 0.235$) also falls within this range, indicating a moderate degree of convergence in annotators' interpretation of stance orientation.

Conversely, *Responsibility* ($\kappa = 0.060$), *Stereotype* ($\kappa = 0.039$), and especially *Offensiveness* ($\kappa = 0.004$) show substantially lower agreement levels. These dimensions require more complex pragmatic inference and contextual interpretation, which likely contributes to the observed variability. In this respect, the final block broadly confirms the same hierarchy already visible in the preliminary phase: dimensions grounded in more explicit cues remain comparatively more stable, whereas those requiring stronger pragmatic inference continue to generate greater dispersion.

First-level dimensions per Subset	Schettino	Rackete
Non-classifiable	-0.108	-0.080
Target	0.294	0.249
Responsibility	-0.111	0.125
Offensiveness	-0.084	0.091
Stance	-0.002	0.278
Humour	0.201	0.154
Professional competence	-0.042	0.209
Stereotype	0.001	0.068

TABLE 6.40: Fleiss' κ values for first-level dimensions computed separately for the Schettino and Rackete subsets.

The subset comparison reveals different stability patterns across the two events. Agreement for *Stance* is considerably higher in ***Fin_Rackete*** ($\kappa = 0.278$) than in ***Fin_Schettino*** ($\kappa \approx 0$), suggesting that stance towards Rackete is more consistently interpreted by annotators. Conversely, ***Fin_Schettino*** shows slightly higher agreement for humour and target identification.

6.2.4.2.2 Second-level dimensions Agreement for second-level dimensions was computed only on the subset of tweets where the corresponding first-level label was annotated as present.

Targets of Responsibility Agreement for Responsibility targets is generally low, with the notable exception of *Politics*, which reaches moderate agreement ($\kappa = 0.409$). Negative values for *Main character* reflect strong annotator disagreement in attributing responsibility to the protagonist, highlighting the interpretative complexity of this dimension.

Targets of Responsibility	**Fleiss' κ**
Main character	-0.238
Organisation	0.059
Institutions	0.104
Politics	0.409
Society	-0.001
Victims	1.000
Other	0.132

TABLE 6.41: Fleiss' κ values for second-level Responsibility targets.

Targets of Offensiveness Offensiveness targets show substantially higher agreement than the first-level Offensiveness detection. In particular, offensiveness directed at the *Main character* and *Politics* reaches moderate agreement, suggesting that annotators converge more reliably when identifying the target of offensive discourse than when determining whether a tweet is offensive.

Targets of Offensiveness	**Fleiss' κ**
Main character	0.429
Organisation	0.076
Institutions	0.237
Politics	0.438
Society	0.264
Victims	0.408
Other	0.357

TABLE 6.42: Fleiss' κ values for second-level Offensiveness targets.

Targets of Humour Humour targets display relatively stable agreement patterns. The highest agreement is observed for humour directed at *Politics* and *Victims*, while humour directed at institutions shows substantially lower agreement.

Targets of Humour	**Fleiss' κ**
Main character	0.267
Organisation	0.255
Institutions	0.048
Politics	0.448
Society	0.175
Victims	0.496
Other	0.218

TABLE 6.43: Fleiss' κ values for second-level Humour targets.

Targets of Stereotype Agreement for stereotype targets varies widely across categories within their respective subsets. Some targets reach fair to moderate agreement, particularly stereotypes related to victims' origin ($\kappa = 0.579$ in Rackete) or gender ($\kappa = 0.409$ for women in Schettino). Perfect agreement values ($\kappa = 1.0$) correspond to categories that are extremely rare or completely absent in the filtered subset, and should therefore be interpreted as a systematic lack of variance rather than genuine discriminative reliability.

Targets of Stereotype (Schettino)	**Fleiss' κ**
Men	0.223
Male captains	1.000
Women	0.409
East. European women	0.315
Cruising experience	1.000
Society	0.277
Other	0.105

TABLE 6.44: Fleiss' κ values for stereotype targets in ***Fin_Schettino*** (computed exclusively where Stereotype = *Yes*).

Targets of Stereotype (Rackete)	**Fleiss' κ**
Women	0.223
Female captains	0.229
Men	-0.007
Migrants	0.294
Africa/people from Africa	0.579
NGOs	-0.007
Society	0.056
Other	0.382

TABLE 6.45: Fleiss' κ values for stereotype targets in ***Fin_Rackete*** (computed exclusively where Stereotype = *Yes*).

Spheres of Stereotype Agreement across stereotype spheres remains consistently low, confirming that annotators find it considerably more difficult to converge on the semantic domain of a stereotype than on its target.

Stereotype sphere	**Fleiss' κ**
Body (incl. Sexuality)	0.132
Behaviour/Attitude	-0.131
Activity/Role	0.034
Origin/Nationality	0.047
Other	-0.026

TABLE 6.46: Fleiss' κ values for stereotype spheres.

6.2.5 Synthesis Across Annotation Phases

A cross-block overview between preliminary and final dataset converges on a consistent picture. The second block does not overturn the first one but tests the stability of the earlier findings under slightly stricter conditions, in particular a revised schema and a longer independent annotation phase.

First, the core discursive asymmetries between the Schettino and Rackete subsets remain structurally stable across phases. Patterns concerning target configuration (e.g. the persistent figurative framing of Schettino vs. the secondary targets in Rackete), stance polarity, humour distribution, and

competence evaluation do not undergo substantial redistribution after the schema revision. This continuity suggests that the observed configurations are not artefacts of a specific annotation version but reflect recurrent structural configurations within the corpus.

Second, comparing IAA across phases reveals a general decrease in Fleiss' κ values during the second block. For instance, exact agreement on *Target* shifted from $\kappa = 0.420$ in the preliminary phase to 0.308 in the final one, and *Humour* shifted from 0.438 to 0.266. Dimensions requiring deep pragmatic inference, such as *Responsibility* (0.158 to 0.060) and *Offensiveness* (0.196 to 0.004), experienced even more pronounced drops. Rather than a failure of the schema revision, this phenomenon could instead reflect the reality of a prolonged, independent annotation phase (approximately three months) where annotator drift may naturally occur. Furthermore, the removal of "escape" options (such as *other* in the Target dimension) forced annotators into harder, more strictly constrained categorical choices, inevitably increasing the dispersion of individual decisions.

Third, despite this decrease in exact tweet-by-tweet agreement, the cross-block analysis confirms that the iterative refinement of the schema did not compromise longitudinal comparability. The shared dimensions maintain highly coherent aggregate distributions (proportions) across the two phases. This indicates that methodological adjustments enhanced the granularity of the schema without altering the underlying analytical framework, as the macroscopic discursive trends survived the microscopic variability among human coders. This point is methodologically important for the whole case study. It suggests that the iterative revision of the schema did help clarify which dimensions were structurally more fragile and which ones were instead capturing recurrent properties of the discourse across phases.

From a broader methodological perspective, this discrepancy between stable macro-distributions and declining micro-agreement reinforces the view of annotation not as a purely mechanical coding procedure but as a situated interpretative practice. The persistence of lower agreement in socially and pragmatically complex dimensions does not signal methodological failure; rather, it reflects the inherent variability of meaning negotiation in politically and morally charged discourse. At the same time, the relative stability of

distributions across iterative revisions demonstrates that even when inter-subjective consensus fluctuates over time, the aggregated corpus statistics can still provide a reliable representation of the overarching narratives.

Overall, the quantitative evidence supports both the robustness and the reflexive adaptability of the annotation design, providing a stable empirical foundation for the post-hoc and qualitative analyses developed in the following sections. In other words, the comparison between the two blocks strengthens the claim that the main analytical patterns identified in the corpus are not reducible to a single annotation moment, but remain visible across successive phases of revision and re-annotation.

6.3 Post-Hoc Qualitative Analysis

This section presents two types of qualitative analysis.

The first analysis (see Section 6.3.1), more strictly linguistic in nature, is a content analysis conducted through a close-reading of the annotated texts.

The second one (see Section 6.3.2) is a meta-methodological analysis, consisting of an examination of the annotators' meta-analysis. Consequently, the dedicated subsection introduces a post-annotation questionnaire, designed to capture the opinions of the individuals who annotated the data throughout the task and to provide a space for their critical reflections about the annotation experience. This serves to complete the process in which the annotation team was involved through several structured and semi-structured interactions. In this sense, the section remains coherent with the broader FATA orientation of the thesis, insofar as annotators are treated not merely as label providers, but also as situated participants whose reflections are methodologically relevant.

6.3.1 Content Analysis

As explained at the beginning of this chapter, the selection of the two events for the construction of our corpus stemmed from the idea of comparing two figures of opposite genders performing the same profession—one traditionally associated with men in the collective imagination—whose respective circumstances had a significant, albeit differing, media impact. The focal point

concerned the expected difference in the ways public discourse (specifically within social media) is articulated around the two captains.

As the case study progressed in an increasingly clear direction, the diversity of the discursive spheres in which the two debates unfolded and the divergence between the characters' actions and their consequences on human lives (without delving into legal and political discourse[5]) provided a valuable opportunity to explore and highlight the manner in which these two figures were narrated and described in certain sectors of social media.

The difference in treatment between the two characters, already confirmed by the annotation data, offered interesting—yet not always predictable—insights when the data were examined more closely.

This qualitative analysis therefore begins with a selection of seven tweets—all extracted from the Rackete subset for obvious temporal reasons—in which we observe a cross-over reference to Schettino.

In the example (11), the comparison between the two figures introduces a clear moral distinction, where the invocation of Schettino highlights the state of public discourse and the perceived sloth of the Italian people.

(11) È che in Italia siamo abituati a Schettino, una come Carola ci mette davanti alla nostra pochezza umana #SeaWacht #Capitana #CAROLA_RACKETE[6]

We note the use of the sole first name for Rackete and the surname for Schettino. In this instance, the use of *Carola* seems to suggest emotional proximity on the part of the author, indicating a decisively positive stance. The use of the given name alone is a relevant nomination strategy, which recurs in many texts and therefore will be explored further in Section 6.3.1.2.

In the example (12), Schettino's case is recalled due to an actual factual cross-over. Indeed, Captain De Falco—a figure who gained media prominence during the Schettino case for the (later viral) phrase "Salga a bordo, cazzo!" (eng. 'Get back on board, damn it!')—expressed himself in 2019 in favour of Rackete (and international law) following the *Capitana's* arrest.

[5]Yet effectively doing so, provided that we do consider the respect for human rights and lives a core legal and political concern.

[6]Our translation: It is just that in Italy we are used to Schettino; someone like Carola puts us in front of our own human smallness #SeaWacht #Capitana #CAROLA_RACKETE.

(12) Mi aspetto al più presto che qualcuno dei fedeli elettori della Lega dirà che Schettino è un eroe e De Falco un delinquente,solo perchè è stato l'unico a spiegare come sarebbe andata a finire questa storia. #CarolaRacketeLIBERA[7]

In this tweet, the criticism towards the Lega Nord and its voters highlights the perceived state of polarisation in public discourse, using Schettino (unequivocally regarded as a negative example of a commander) as a paradoxical emblem of political discourse.

In the remaining five tweets, the protagonists are nominated by the usage of nomina agentis indicating their roles, as in the following examples:

(13) Vi racconto una storia.
Ci sono 2 navi, in una la comandante #CarolaRackete mette al sicuro delle vite, nell'altra il comandante Schettino uccide e scappa. La prima viene arrestata per direttissima, il secondo fa conferenze all'università. [8]

(14) Il comandate Schettino provocò la morte di 32 persone per un gesto goliardico e criminale. Restò libero per 5 anni. Il capitano Carola Rackete salva 42 persone e viene arrestata, come la peggiore dei delinquenti. Benvenuti nell'Italia di oggi Ladies and Gentlemen![9]

(15) L'Italia è quel curioso paese dove un capitano di una nave che salva vite umane viene arrestato mentre un altro capitano che abbandona la propria nave senza prima accertarsi che tutti siano stati messi in salvo,

[7]Our translation: I expect that very soon some of the loyal Lega voters will say that Schettino is a hero and De Falco a criminal, only because he was the only one to explain how this story would end. #CarolaRacketeLIBERA.

[8]Our translation: I will tell you a story.
There are 2 ships: in one, the (female) commander #CarolaRackete saves lives; in the other, the (male) commander Schettino kills and runs away. The former is summarily arrested, the latter gives lectures at university.

[9]Our translation: The (male) commander Schettino caused the death of 32 people through a goliardic and criminal act. He remained free for 5 years. Captain Carola Rackete saves 42 people and is arrested like the worst of criminals. Welcome to today's Italy, Ladies and Gentlemen!.

viene mandato a dare lezioni all'università. #2Luglio #CarolaRackete[10]

It is interesting to note that in this last case, Schettino's name is not explicitly mentioned; instead, the nomen *capitano* is used to refer to him indirectly yet unequivocally. Here, as in the previous two tweets, the paradoxical implications of the two actions are highlighted through a moral comparison between "a captain who saves lives and a captain who abandons ship".

(16) Tra la capitana #CarolaRackete e chi sta portando a sbattere il nostro #paese alla #Schettino: non ho dubbi! Giocare con le paure, col tempo, è come seminare vento raccogliere tempesta. SeaWatchItaly seawatch_intl #cesolounacapitana[11]

In this final instance, Schettino is instead used figuratively—once again as a symbol of decadence—with an expression indicating a country *à la* Schettino. The comparison does not appear to be between the *capitana* and him, but rather between Rackete and the actors of the prevailing political discourse (presumably, Matteo Salvini and other politicians who shared the same stance towards the circumstance).

Among the examples here presented, only (16) introduces Rackete with the feminine form *capitana*. Nonetheless, as we already noticed in previous examples this does not necessarily implies that the use of masculine forms carries negative stance or stereotypes. In Section 6.3.1.1 we investigate this aspect, among others.

6.3.1.1 Nomina Agentis

Building on the preliminary work presented in Chapter 5, part of our content analysis focuses on two specific nomina agentis used in Italian to denote the role of a **person in command of a vessel**: *capitana/capitano* [captain.F/captain.M] and *comandante* [commander].

[10]Our translation: Italy is that curious country where a captain of a ship who saves human lives is arrested, while another captain who abandons his ship without first ensuring that everyone has been brought to safety is sent to give lectures at university. #2July #CarolaRackete.

[11]Our translation: Between the (female) captain #CarolaRackete and those who are crashing our #country in the style of #Schettino: I have no doubts! Playing with fears, over time, is like sowing the wind and reaping the whirlwind. SeaWatchItaly seawatch_intl #thereisonlyonecaptain.

Linguistically, these two forms do not behave in the same way in Italian. *Capitana/o*, as a *mobile gender* noun, overtly encodes a formal masculine/feminine opposition. By contrast, *comandante* is a *common-gender* noun and therefore keeps the same lexical form across references to male and female referents; in this case, formal assignment can be recovered only through agreement targets. This difference is relevant for the present analysis, because it means that gender becomes visible in different ways: directly through derivational morphology in the former case, and through co-textual agreement cues in the latter.

Regarding their pragmatic use, a cautious distinction may also be proposed. In the *Treccani* definitions, *comandante*[12] is more tightly associated with the exercise of command itself and, in the maritime domain, with its institutional and operational dimension, whereas *capitana/o*[13] appears as a broader designation of the person or professional role entrusted with that command. In the context of navigation, the two terms may certainly overlap, and in our corpus they are often used as near-synonyms.

This investigation aims to determine whether the use of masculine and feminine forms of these nouns in relation to Carola Rackete activates specific semantic fields or discursive tensions, including metalinguistic reflection and reported speech.

To facilitate this analysis, we merged the preliminary and final datasets, subsequently extracting two distinct sub-corpora: one comprising all occurrences of the nomen agentis built on the nomen ***capitan-*** and another for the nomen ***comandante***. These sub-corpora were then filtered based on their association with the **Schettino** or **Rackete** subsets. For each sub-corpus, we classified the formal assignment and counted how many times each form (masculine and feminine, plural and singular) appears.

The composition of the formal assignment in the sub-corpora is detailed in Table 6.47 and Table 6.48.

[12] `https://www.treccani.it/vocabolario/comandante/`.
[13] `https://www.treccani.it/enciclopedia/capitano/`.

Form	***Schettino***	***Rackete***	**Overall**
Masculine			
capitano	*25*	*57*	74
capitan	*17*	*7*	25
capitani	*2*	*5*	*7*
Subtot.	***44***	***69***	**106**
Feminine			
capitana	-	*81*	81
capitane	-	*1*	1
Subtot.	**-**	***82***	**82**
Total occurrences	***44***	***151***	**188**

TABLE 6.47: Distribution of *Capitan-* per subset based on formal assignment. N = 166: total of tweets involved (n = 38 for Schettino; n = 128 for Rackete).

comandante	***Schettino***	***Rackete***	**Overall**
Masculine	*51*	*8*	**59**
Feminine	-	*9*	**9**
Unassigned	*15*	*3*	**18**
Total occurrences	***66***	***20***	**86**

TABLE 6.48: Distribution of *Comandante* per subset based on formal assignment. N = 83: total of tweets involved (n = 64 for Schettino; n = 19 for Rackete).

In both tables, the total amount of occurrences (*capitan-* = 188; *comandante* = 86; total = 274) is more than the total amount of tweets (tweets with *capitan-* = 166; tweets with *comandante* = 83; total = 249), as in some cases more nomina agentis are used in the same tweet.

Moreover, a little amount of tweets (n = 1 in the Schettino subset; n = 2 in the Rackete one; total n = 3) presents both nomina agentis simultaneously. Therefore, a small overlap of tweets between the two sub-corpora was noted.

A total of 246 tweets over the entire corpus (2,000 tweets) contain at least one nomen agentis, representing the 12.3% of total tweets annotated by the annotation team. Of these, 101 tweets are from the **Schettino** subset and 145 from the **Rackete** one, showing a larger use of nomina in the tweets regarding the Rackete case. This is true if we consider the occurrences of the nomen ***capitan-*** (151 occurrences in Rackete-related tweets involved and only 37 occurrences triggered by Schettino), but an opposite trend emerges with respect to the use of *comandante* (66 occurrences in the Schettino subset versus 20 occurrences in the Rackete one).

As seen in Tables 6.47 and 6.48, the first step considered *formal assignment* only. In the case of ***capitan-***, formal assignment was overtly encoded by the nominal morphology itself, i.e., by the suffixal opposition between masculine and feminine forms. In the case of ***comandante***, by contrast, formal assignment could only be recovered when agreement targets were present (such articles, adjectives, pronouns). Where no such target was available, the occurrence was therefore classified as (formally) *Unassigned*.

From this classification, the initial observation we can do is that no feminine form is detected in tweets from the **Schettino** subset, whereas the masculine forms (***capitan/o/i***) appear in texts regarding both figures. The plural form ***comandanti*** is also absent throughout the whole corpus.

While the initial classification was conducted solely on the basis of formal assignment, in order to quantify the different forms attested in the corpus, the second step consisted in classifying all occurrences also in terms of *referential specificity*, i.e., according to whether the nomen was used with generic reference or referred to a specific identifiable individual. This second step is grounded in the analytical framework developed in Chapter 5 and, more broadly, in the literature discussed there on gender assignment in Italian and the relation between formal marking, agreement, semantic incongruity, and reference in Italian. Regarding referential specificity, what is at stake in this second step is not the *referential gender* of the individual as such, but the relation between formal assignment and the recoverability of a specific referent in discourse.

The classification scheme is the following: *Unassigned Generic* (no formal assignment, generic reference); *Assigned Generic* (formal assignment, generic reference); *Unassigned Specific* (no formal assignment, but specific identifiable

reference); and *Assigned Specific* (formal assignment together with a specific identifiable reference), as summarised in Table 6.49.

	Generic reference	*Specific reference*
No formal assignment	**Unassigned Generic** (e.g., bare *comandante* without agreement target, used with non-specific reference)	**Unassigned Specific** (e.g., bare *comandante* without agreement target, referring to a specific identifiable individual)
Formal assignment present	**Assigned Generic** (e.g., *capitano* or *capitani* used with non-specific reference or for the category in general)	**Assigned Specific** (e.g., *capitana* or *capitano* referring to a specific identifiable individual)

TABLE 6.49: Taxonomy used for the classification of the nomina agentis (classification of the forms by formal assignment and referential specificity).

Following this scheme, Tables 6.50 and 6.51 report the distribution of the two nomen types across the four categories just introduced. While formal assignment was established on the basis of overt formal cues (suffixes in the case of *capitan-*, agreement targets in the case of *comandante*), referential specificity was assessed through anaphoric, cataphoric, or exophoric cues, i.e. relying on contextual pragmatic inferences, using our world knowledge, or co-textual cues.

capitan-	***Schettino***	***Rackete***	**Overall**
Unassigned Generic	-	-	**-**
Assigned Generic	*2*	*11*	**13**
Unassigned Specific	-	-	**-**
Assigned Specific	*35*	*140*	**175**

TABLE 6.50: Distribution of occurrences of ***Capitan-*** by formal assignment and referential specificity per subset.

comandante	*Schettino*	*Rackete*	**Overall**
Unassigned Generic	*1*	*1*	**2**
Assigned Generic	*2*	*1*	**3**
Unassigned Specific	*14*	*2*	**16**
Assigned Specific	*49*	*16*	**65**

TABLE 6.51: Distribution of occurrences of ***Comandante*** by formal assignment and referential specificity per subset.

6.3.1.1.1 Unassigned Generic The category *Unassigned Generic* is marginal in quantitative terms, as it includes only 2 occurrences overall, both of them involving the common gender form *comandante* without an overt agreement target and without reference to a specific identifiable individual. This distribution is fully consistent with the forms involved: unlike *capitan-*, which always carries an overt formal assignment, bare *comandante* may remain formally unassigned while also being used with generic reference.

In the example (17), *comandante di nave* appears as part of a Carnival costume description; the role is evoked as a recognisable social type rather than as a reference to an actual commander:

(17) Alla terza persona vestita da comandante di nave e due vestite da dalmata realizzo che quelli sono vestiti da schettino ed è carnevale.[14]

In example (18), the genericity is explicitly licensed by the universal formulation *ogni comandante di naviglio deve fare*, which frames the role in deontic and professional terms:

(18) Secondo il Consiglio direttivo della Camera penale la capitana Rackete "ha fatto solo ciò che ogni comandante di naviglio deve fare: ha prestato soccorso a dei naufraghi allo stremo delle forze e in balia del destino oltre che delle onde".[15]

[14]Our translation: By the time I see the third person dressed as a ship's commander and two others dressed as Dalmatians, I realise they are dressed up as Schettino and that it is Carnival.

[15]Our translation: According to the Executive Board of the Criminal Lawyers' Chamber, Captain Rackete "did only what every ship's commander must do: she assisted shipwrecked people who were utterly exhausted and at the mercy of fate as well as of the waves".

This last example is particularly interesting because the use with generic reference appears inside reported speech attributed to an institutional source (*Secondo il Consiglio direttivo della Camera penale...*). In other words, the generic reading is not only semantic—as we interpret the described characteristics as part of the role of an ideal ship's commander, regardless of their referential gender—but also discursively embedded in a quoted legal-normative formulation. For this reason, and in light of the asymmetries discussed throughout this thesis, we classify it as an *Unassigned Generic* form, even though one might speculate that, had the same formulation involved the type *captain*, the realised form could have been either *capitano* or *capitana*.

More broadly, these examples suggest that *Unassigned Generic* forms are rare in the corpus and tend to emerge either when the role is invoked as a social type, or when it is framed through generalising and normative discourse rather than through direct reference to one of the two protagonists. This scarcity is consistent with the comparatively lower frequency of the nomen agentis type *commander* in the corpus as a whole.

At the same time, the purpose of introducing this taxonomy and analysing the tweets through this lens is not primarily to seek straightforward discursive confirmation of how these nomina agentis are used. Rather, the taxonomy is intended to make visible how the distribution of particular inflectional patterns and naming practices may itself provide further evidence of the semantic and pragmatic difficulty of interpretation that characterises these forms, as well as of the asymmetries discussed throughout this chapter. From this perspective, even a quantitatively marginal category such as *Unassigned Generic* remains analytically relevant, insofar as it helps clarify the conditions under which formal and referential assignment may remain suspended. By contrast, the subsequent categories—especially *Assigned Generic* and, above all in the Rackete subset, *Assigned Specific*—are expected to offer richer evidence of how professional role labels could become sites of evaluative, ideological, and interactional negotiation.

6.3.1.1.2 Assigned Generic The category *Assigned Generic* includes cases in which the nomen agentis is formally assigned, but refers not to a specific identifiable individual, but to the role in general. Quantitatively, the category

is limited but not negligible: across the two *nomina agentis*, it includes 16 occurrences overall, 4 in the Schettino subset and 12 in the Rackete subset.

(19) Dopo il capitan #Schettino il capitan #Buffon...Che capitani che abbiamo in Italia!!!!'[16]

(20) Foto di schettino su tgcom.la mia amica Pamela esordisce " badate che capitani di nave che mettono, nemmeno quelli della lego" oioi moio[17]

In the **Schettino** sub-corpus, *Assigned Generic* forms—supported by the fact that they appear in their plural form—mainly project the individual case onto a broader type, allowing the role label to function as a vehicle for ridicule, blame, or generalised incompetence. Similarly, in **Rackete** the following examples are interesting from a discursive point of view:

(21) Delirio di onnipotenza manovrata da chi spudoratamente specula su #migrinti #freeCarolaRackete Ogni giorno un capitano si alza si alza al mattino e sa che dovrà entrare forzatamente nei porti italiani![18]

(22) #CarolaRackete è libera: bene! Ci saranno altre capitane ed altri capitani a salvare vite in mare: bene! Ci saranno più navi: bene! Ci saranno più partenze: meno bene. Si saranno meno scrupoli per far soldi: male! Ci saranno più trafficanti: molto male! #CarolaRacketeLIBERA[19]

Here, *Assigned Generic* forms are associated with idealised or normative typifications of the figure of the captain: the role is projected collectively or even proverbially[20]. In both cases, the generic form appears in a frame where Rackete is seen as part of a broader class of legitimate actors engaged in rescue.

[16]Our translation: After Captain #Schettino, Captain #Buffon... What captains we have in Italy!!!!

[17]Our translation: A photo of Schettino on Tgcom. My friend Pamela opens with: "Just look at the ship's captains they put there, not even the ones from Lego " ouch ouch I'm dying.

[18]Our translation: Delusions of omnipotence manipulated by those who shamelessly speculate on #migrints #freeCarolaRackete Every day a captain gets up in the morning and knows that they will have to force their way into Italian ports!

[19]Our translation: #CarolaRackete is free: good! There will be other captains.F and other captains.M saving lives at sea: good! There will be more ships: good! There will be more departures: less good. There will be fewer scruples about making money: bad! There will be more traffickers: very bad! #CarolaRacketeFREE

[20]In ironic analogy with the popular motivational fable's incipit: "Every morning a lion gets up and knows it must run faster than the slowest gazelle, or it will starve".

Particularly noteworthy is the example (22), where generic reference coexists with explicit gender marking through the coordinated split form "altre capitane ed altri capitani". Here, by contrast, masculine and feminine forms coexist in an inclusive formulation through a conservative visibility strategy (see Rosola et al., 2023). The juxtaposition does not produce tension between competing labels, but rather projects a mixed collective horizon of legitimate sea rescue actors. This is one of the clearest examples in the corpus in which gender marking is made visible within a generic and positively valued frame.

(23) Io avrei una domanda: ma invece di fare sta cosa del reggiseno per sostenere Carola Rackete, non sarebbe meglio e più opportuno fare il brevetto da capitano?[emoji][21]

(24) Ammiro il coraggio di #CarolaRackete. Pochissimi avrebbero rischiato in prima persona come ha fatto lei per assolvere ai suoi doveri di capitano di una nave e per il bene delle persone a bordo.
Tutta la mia stima.
#SeaWatch3
#fateliscendere
#RestiamoUmani
#Capitana
#Carola[22]

The second pair points in the same direction, while making the professional and deontic dimension more explicit. In (23), "fare il brevetto da capitano" invokes the role as a certified professional category; in (24), "i suoi doveri di capitano di una nave" foregrounds what a captain is expected to do. In both cases, the generic form contributes to construing the role as a site of legitimacy, competence, and normative expectation, and appear to be coherent with a positive evaluation of Rackete's professional competence which we have seen as quite frequent in our annotation data.

[21] Our translation: I have a question: instead of doing this bra thing to support Carola Rackete, wouldn't it be better and more appropriate to get a captain's licence? [emoji].

[22] Our translation: I admire #CarolaRackete's courage. Very few people would have taken such a personal risk as she did in order to fulfil her duties as the captain of a ship and for the good of the people on board. She has all my respect. #SeaWatch3 #letthemdisembark #StayHuman #Captain #Carola.

At the same time, what begins to emerge as particularly significant in this category is the near absence of formally feminine generic forms. Apart from the coordinated split form "altre capitane ed altri capitani", feminine morphology is virtually absent from *Assigned Generic* forms in the corpus. This does not in itself give us explanations, nor does it allows strong claims about speakers' underlying representations. It does, however, remain consistent with a broader pattern already observed in the data: overextended uses of feminine forms—i.e. forms used generically for mixed groups—are absent in general, and moreover generic reference to the role is itself only exceptionally associated with feminine forms.

From this perspective, the asymmetries at stake are not exhausted by lexical or semantic meaning alone. They also concern the pragmatic availability of particular forms for generic projection. Put differently, the issue is not simply which form denotes the role, but which *Assigned Generic* forms are discursively available for normative or professional reference, and which remain restricted or dependent on explicit agreements.

Therefore, the distribution of *Assigned Generic* forms can offer further evidence of the aforementioned semantic and pragmatic difficulty of interpretation, as well as of the persistence of asymmetrical patterns. In this sense, *Assigned Generic* already proves more revealing than *Unassigned Generic*.

6.3.1.1.3 Unassigned Specific The category *Unassigned Specific* includes those cases in which the nomen agentis does not carry an overt formal assignment, but nevertheless refers to a specific identifiable individual through contextual, pragmatic, or co-textual cues. In other words, the referent is recoverable even though the noun remains formally unassigned.

As expected, this category is attested only with the common gender form *comandante*. In quantitative terms, it includes 16 occurrences overall, with a clear concentration in the Schettino subset (14 occurrences, against only 2 in the Rackete one).

(25) "Cazzo di scoglio. Una striscia tutta nei capelli mi è finita. Sembro Mr. Fantastic." CIT. Comandante #Schettino subito dopo l'incidente'[23]

[23]Our translation: "Fucking rock. A streak has ended up all through my hair. I look like Mr Fantastic." CIT. Commander #Schettino immediately after the accident.

In this example, specificity is established through the hashtag #Schettino and through the explicit quotation frame (*CIT.*), which presents the utterance as reported speech attributed to him. The role label *comandante* is therefore referentially specific, even though no agreement target is present to help us mark gender formally, unless we consider 'Schettino' as an agreement target. However, we would say that is our knowledge of the commander's gender—which, by the way, in general, is often inferred—to allow the assignement of a semantic gender to the noun.

(26) @F_Schettino FORZA COMANDANTE, SONO CON TE !!!!!'[24]

Here specificity is even more directly licensed by the addressivity of the tweet. The mention `@F_Schettino` and the vocative use of *comandante* make it clear that the term refers to one individual only, namely Schettino. At the same time, the absence of any overt gender marker leaves the form formally unassigned. This example is particularly noteworthy because the vocative frame produces a strong pragmatic anchoring of the referent while keeping the lexical form itself formally not assigned.

(27) NO VABBÈ,'MO ADDIRITTURA IL VESTITO DI CARNEVALE DA COMANDANTE SCHETTINO,Non bastava quello da Michele Misseri?Ahahah che schifo l'Italia C:'[25]

This example is similar to (17), presented in Paragraph 6.3.1.1.1, where the carnival costume is evoked. However, differently the specific reference is built through a formula that resembles a compound denomination (*comandante Schettino*), where 'Schettino' functions as referential cue. The phrase evokes the figure as a recognisable Carnival costume, yet the role still points to a specific individual rather than to the profession in general. It is therefore referentially specific, while remaining formally unassigned.

The two Rackete-related examples show a similar semantic mechanism, but in a much more limited way within the corpus:

[24]Our translation: @F_Schettino STAY STRONG COMMANDER, I'M WITH YOU !!!!!

[25]Our translation: NO WAY, NOW THERE'S EVEN THE COMMANDER SCHETTINO CARNIVAL COSTUME, Wasn't the Michele Misseri one enough? Ahahah, how disgusting Italy C:

(28) Rischia l'arresto Carola Rackete, la 31enne tedesca comandante della #SeaWatch, che ieri ha forzato il blocco entrando in acque italiane con la nave – a bordo della quale vivono da due settimane 42 migranti, ormai allo stremo delle forze – che ora si trova a 3miglia da Lampedusa.[26]

(29) La Procura di Agrigento ha iscritto Carola #Rackete, comandante della #SeaWatch3, nel registro degli indagati per i reati di favoreggiamento dell'immigrazione clandestina e rifiuto di obbedienza a nave militare.[27]

In both cases, *comandante* is not formally assigned, yet the appositional structure (*Carola Rackete, comandante della #SeaWatch*) makes the referent fully specific. Unlike the Schettino-related examples, however, these tweets do not involve direct address, quotation, or strong reported-speech framing. Rather, they belong to a more report-like style, in which the bare common gender noun functions as apposition.

Taken together, these examples suggest that *Unassigned Specific* forms are not merely cases of missing morpho-syntactic information, but an important site where specificity is delegated to the pragmatics of the discourse structure. Hashtags, mentions, appositions, quotation frames, and exophoric world knowledge all contribute to making the referent identifiable even in the absence of overt gender marking. This is particularly clear in the Schettino subset, where the figure of *comandante* often appears as a highly recognisable media persona, recoverable through contextual cues alone. In methodological terms, this category is therefore especially significant, because it shows that referential assignment may be robust even when formal assignment is absent, and that the interpretation of role nouns depends heavily on pragmatic inference.

[26] Our translation: Carola Rackete, the 31-year-old German commander of the #SeaWatch, risks arrest after yesterday forcing the blockade by entering Italian waters with the ship – on board which 42 migrants have been living for two weeks, now exhausted – which is now located 3 miles from Lampedusa.

[27] Our translation: The Public Prosecutor's Office of Agrigento has placed Carola #Rackete, commander of the #SeaWatch3, under investigation for the offences of aiding illegal immigration and refusal to obey a military ship.

6.3.1.1.4 Assigned Specific Given the relevant presence of secondary targets throughout the dataset, first of all we proceeded to a granular examination of the *Assigned Specific* forms, which showed that the occurrences do not always refer to the two protagonists. Although the category remains quantitatively dominated by references to Schettino and Rackete, a non-negligible set of occurrences is redirected towards other figures, either through analogy, irony, or topical overlap with contemporaneous events.

As for the form *capitan-*, in the **Schettino** subset, 9 occurrences have as their referent—either explicitly or clearly inferable from the co-text—another figure. Two of them refer to fictional or commercial *personae* (*Capitan Findus* and *Pasta del Capitano*), whilst the majority of the others refer to the captain of the *Costa Allegra*, a cruise ship owned by the same company which suffered a fire off the coast of the Seychelles only six weeks after the Giglio disaster. In these cases, the title is still referentially specific, but the referent shifts away from Schettino, often in order to establish a contrast or ironic parallel.

Within the Rackete subset, the nomen *capitan/o* is used 15 times to refer to other individuals. More specifically, it appears 11 times in reference to Minister Matteo Salvini, frequently employed with rhetorical or ironic intent (e.g. "*capitano senza coraggio*" [captain without courage], "*capitan coniglio*" [captain rabbit], or "*Capitan Fracassa Fellone*"[Captain Fracasse Scoundrel]). Alongside this exclusively ironic attribution of the nomen to Salvini, the epithet *capitone* [*large eel*]—in phonetic and graphic consonance with *capitano*—occurs throughout the corpus (six times in the singular and once in the plural form), as in the following example:

(30) La Capitana #CarolaRackete annuncia querela contro il capitone.
Adesso vediamo quanto ci mette a tornare in lacrime, rifugiandosi dietro la sua immunità del cazzo.
#60milionimenouno #portiaperti #piùcapitanimenocapitoni[28]

Here the playful deformation *capitone* works as a mocking anti-title: Salvini is inserted into the semantic space of *capitano*, but only to be immediately degraded through wordplay. The rhetorical operation is revealing, because it

[28]Our translation: Captain.F #CarolaRackete announces a lawsuit against the *capitone*. Now let's see how long it takes him to return in tears, hiding behind his fucking immunity. #60millionminusone #openports #morecaptainless*capitoni*

shows that the title had by then become salient enough in the public debate to be re-functionalised polemically against a different actor.

(31) UNA VERA CAPITANA VESTITA DA SOCCORRITORE e un finto capitano travestito da politico
#SeaWatch
#CarolaRackete
#capitone[29]

This example is especially useful because it stages the contrast between Rackete and Salvini through a pair of opposed role labels: *vera capitana* versus *finto capitano*. The title is thus reassigned across two figures in order to oppose authentic rescue work to political performance. What is interesting in this case is the fair use of the feminine form *capitana* (which seems to feed the positive stance even more), but immediately followed by "soccorritore", which is surprising to us, as we consider the use of the feminine equivalent *soccorritrice* not problematic at all (mostly, in the context of the nomen that we are exploring).

Regarding the nomen ***comandante***, 15 occurrences—and only in the Schettino subset—refer to alternative figures; here as well the majority of the instances pertain to the commander of the *Costa Allegra*. In these tweets, the overlap between two near-contemporary *Costa* incidents enables a dense ironic interplay between the two commanders, often by suggesting analogy, bad advice, or family resemblance between them.

(32) SMS di Schettino al Comandante della Costa Allegra: "1 pari"....'[30]

(33) Schettino ha consigliato al comandante della #Costa allegra di inclinare la nave su un fianco per spegnere le fiamme.. #renameCosta CostaSub'[31]

Both tweets rely on the recognisability of Schettino as a failed commander in order to produce humour through false advice and sporting metaphor.

[29]Our translation: A real captain.F dressed as a rescuer / and a fake captain dressed up as a politician #SeaWatch #CarolaRackete #capitone

[30]Our translation: SMS from Schettino to the Commander of the Costa Allegra: "1–all"....'

[31]Our translation: Schettino advised the commander of the #Costa allegra to tilt the ship on one side to extinguish the flames.. #renameCosta CostaSub'

The *comandante* of the *Costa Allegra* is therefore specific, but his mention is discursively subordinated to the Schettino frame. In other words, the alternative referent is made interpretable precisely through inter-discursive proximity to the original scandal.

Nonetheless, the majority of occurrences of nomina agentis are effectively referred directly to the two protagonists. Even if *comandante* remains proportionally more frequent for Schettino than for Rackete (51 protagonist-oriented occurrences versus 20), when the two sub-corpora are considered together Rackete is still nominated by means of a professional title more often than Schettino. This seems to tell us that the role itself becomes especially salient in the public discourse surrounding Rackete.

Given this premise, from now on we examine more closely the use of nomina agentis within the Rackete subset by comparing feminine and masculine forms while considering the annotation of stance (always based on majority vote). Prior to this, however, it is necessary to determine whether a significant difference exists in the functional use of these nomina. By investigating the distribution of masculine and feminine forms and their respective stance profiles, we aim to uncover any divergent patterns of use. Furthermore, whilst the prevalence of *capitan-* in reference to Rackete is evident, the subsequent section will provide a broader interrogation of other nomination strategies. It is indeed observed that the subject is frequently addressed not only by her role but also by her given name or through titles such as *signora* [Mrs] or *signorina* [Miss].

Within the Rackete subset, when only occurrences referring unequivocally to the protagonist are considered, feminine forms are clearly more frequent than masculine ones. More specifically, *capitana* accounts for 81 occurrences (in 78 tweets), whereas masculine *capitano/capitan* referring to Rackete accounts for 44 occurrences (in 33 tweets). As for the common gender *comandante*, 9 occurrences are formally interpretable as feminine, while 6 occurrences are formally interpretable as masculine. Table 6.52 summarises these distributions.

In the terms adopted in Chapter 5, the formally masculine forms here are *incongruous* uses as they are referred to a woman.

***Assigned Specific* forms referring to Rackete**	**Occurrences**	**Tweets**
capitana	81	78
capitano/capitan	44	33
comandante (feminine agreement)	9	9
comandante (masculine agreement)	6	6
Feminine total	**90**	**87**
Masculine total	**50**	**39**

TABLE 6.52: Amount of **occurrences** of nomina agentis referring unequivocally to Rackete, by formal gender.

When these tweet subsets are matched with the majority-vote stance annotation, both feminine and masculine forms turn out to occur more often in *pro* tweets than in *contro* ones. In particular, tweets containing feminine forms referring to Rackete are predominantly supportive (58 *pro* out of 87 tweets overall, if *capitana* and feminine *comandante* are considered together), but masculine forms also appear predominantly in supportive contexts (24 *pro* out of 39 tweets overall, if masculine *capitano/capitan* and masculine *comandante* are considered together). Table 6.53 details the distribution.

***Assigned Specific* forms referring to Rackete**	**Pro**	**Against**	**Neutral**	**Tie**
capitana	52	18	5	3
capitano/capitan	22	4	4	3
comandante (feminine agreement)	6	2	0	1
comandante (masculine agreement)	2	2	1	1
Feminine total tweets	**58**	**20**	**5**	**4**
Masculine total tweets	**24**	**6**	**5**	**4**

TABLE 6.53: Majority-vote **stance** in tweets containing nomina agentis referring unequivocally to Rackete, by formal gender.

Therefore, what the present data show is that the incongruous assignment (i.e., masculine nomination) cannot be automatically equated with hostile stance. On the contrary, the data suggest that the opposition between masculine and feminine forms is functionally more complex. The congruous

feminine assignment is clearly the preferred morphosyntactic option in explicit reference to Rackete, but the masculine is still used, and far from being confined to antagonistic or delegitimising environments. Both congruous and incongruous forms may occur in supportive, hostile, neutral, and ambiguous contexts, albeit with different frequencies.

A first point worth emphasising is that some masculine forms are embedded in tweets whose supportive stance is very clear:

(34) Io sto con il capitano #CarolaRackete perché prima vengono gli esseri umani, sempre, e dopo vengono i confini e le regole scritte da chi non è mai stato torturato in un lager libico.
Tutta l'italia solidale e per bene è con la #SeaWatch3.[32]

Here the masculine *capitano* is directly attached to Rackete in a strongly solidaristic and humanitarian frame. There is no signal of irony or distance; on the contrary, the title functions as a dignifying professional label. This kind of example is crucial because it shows that the masculine form may still be used as an honorific or institutionally legitimate role noun in a supportive discourse.

(35) Il Capitano #CarolaRackete ha dimostrato a ogni europeo cosa voglia dire avere coraggio e cuore, ergendosi in questo tempo alla rovescia che condanna il giusto ed esalta il malvagio come un faro nella notte. #SeaWatch3 #freeCarola[33]

In this tweet, too, the masculine *Capitano* appears within a clearly laudatory stance. The role noun is associated with a moral and symbolic elevation of Rackete, rather than with a diminishment of her. Such cases suggest that the masculine may still retain a value of institutional gravitas or generic prestige, even when applied to a female referent.

[32]Our translation: I stand with captain.M #CarolaRackete because human beings always come first, and only afterwards come borders and rules written by people who have never been tortured in a Libyan lager. All of Italy that is supportive and good stands with #SeaWatch3.

[33]Our translation: Captain.M #CarolaRackete has shown every European what it means to have courage and heart, standing in this upside-down age that condemns the just and exalts the wicked like a beacon light in the night.

At the same time, the feminine form is not only restricted to supportive usage. It also occurs in clearly hostile discourse:

(36) Non sopporto le mistificazioni dei Delrio, Faraone, Fratoianni, ecc.. La capitana #Rackete non è stata arrestata per aver salvato vite umane, ma per aver violato leggi e codici mettendo a repentaglio vite umane, perseguendo l'imperativo categorico di portare migranti in Italia.[34]

This example is particularly useful because it shows that the feminine form *capitana* may appear in openly antagonistic contexts without thereby indexing positive stance. The mere use of the feminine therefore does not in itself imply feminist alignment, supportive attitude, or stereotype-free discourse. It may simply function as the chosen role label within a hostile frame.

(37) Visto che il @pdnetwork ha raccolto tanti soldi per la capitana #KarolaRackete potrebbe almeno vestirla? O la lasciano andare in tribunale seminuda? La forma é sostanza, diceva Benedetto Croce[35]

In (37) the feminine title coexists with an explicitly gendered and appearance-based attack. This is particularly relevant for the present study, because it confirms that even the formally feminine form may be embedded in sexist or objectifying discourse. In other words, the grammatical visibility of femininity does not neutralise the possibility of stereotypical framing.

An important point of this analysis concerns the degree to which the role noun seems to be selected *in the running text* rather than merely inherited through a pre-existing hashtag. Although this distinction can only be treated as approximate, it is still informative. In the case of feminine forms referring to Rackete, 67 out of the 87 relevant tweets contain the nomen in the body of the tweet, whereas 20 tweets appear to rely on a hashtagged form only. By contrast, in the masculine group, all 39 relevant tweets contain the nomen in the running text; the hashtagged masculine forms retrieved in the broader subset

[34]Our translation: I cannot stand the misrepresentations of Delrio, Faraone, Fratoianni, etc.. Captain.F #Rackete was not arrested for saving human lives, but for violating laws and codes, endangering human lives, and pursuing the categorical imperative of bringing migrants into Italy.

[35]Our translation: Given that the @pdnetwork has raised so much money for captain.F #KarolaRackete, they could at least dress her. Or are they letting her go to court half naked? Form is substance, as Benedetto Croce said.

are either generic, metalinguistic, or refer to other figures rather than constituting clear protagonist-directed uses. Table 6.54 summarises this distinction at tweet level.

***Assigned Specific* forms referring to Rackete**	**In running text**	**In hashtag only**
Feminine forms total tweets	67	20
Masculine forms total tweets	39	0

TABLE 6.54: Approximate distinction between running-text use and hashtag-only use of nomina agentis referring unequivocally to Rackete.

This distinction is not secondary. If a form is present only as a hashtag, one may hypothesise that its use is at least partly facilitated by the circulation of a ready-made public label. By contrast, when the nomen appears in the running text, it is more plausibly the result of a local lexical choice made by the author while composing the tweet. From this point of view, the data suggest that the feminine form enjoys both a higher overall circulation and a stronger hashtagged visibility, whereas the masculine forms referring to Rackete appear to be selected almost entirely in the body of the tweet. This strengthens the interpretation that masculine nomination is not merely an artefact of a pre-packaged tag, but often a deliberate discursive choice.

A further qualitative pattern concerns the metalinguistic salience of these forms. As already shown by the overlapping examples discussed below, the corpus contains several cases in which the masculine and feminine are explicitly juxtaposed, corrected, or contrasted. This suggests that the choice of role noun had itself become an object of public reflection. However, the quantitative tables above help refine that impression: the coexistence of forms is not reducible to overt debate alone, since even outside explicitly metalinguistic tweets both masculine and feminine forms circulate in supportive and hostile discourse.

For these reasons, the data seem to support a more nuanced interpretation. The feminine is clearly the dominant explicit form for Rackete and is more strongly available as a public label, including in hashtagged form. The masculine, however, remains well represented and cannot be treated as a simple marker of antagonism. Rather, it appears to oscillate between at least three functions, that is (a) a conservative or institutionally prestigious title, (b) a

genericised professional noun applied to a specific referent, and, in some cases, (c) a metalinguistic or contrastive resource in gender-aware discourse. This is precisely why a purely formal count of masculine versus feminine forms would be insufficient on its own: only once these uses are crossed with stance and interpreted qualitatively the functional complexity of the nomination patterns become fully visible.

The coexistence of supportive and hostile contexts in both formal options also helps explain why some of the most revealing cases are those in which masculine and feminine forms appear within the same tweet. It is precisely in these local juxtapositions that the metalinguistic visibility of the choice becomes most explicit.

The juxtaposition of masculine and feminine forms is particularly noteworthy, as demonstrated in the following examples.

(38) Il gip di Agrigento non ha convalidato l'arresto di #CarolaRackete, il comandante (e non la "capitana") della #SeaWatch3 e non ha disposto nei suoi confronti alcuna misura cautelare.
Lo stato italiano sa sempre come farsi rispettare, non c'è che dire!
Applausi[36]

In this first case[37], the tweet is explicitly metalinguistic: the masculine form *comandante* is not merely used, but contrasted with the feminine *capitana* in parenthetical form, embedding a negative stance. The opposition is overtly framed as a correction, and therefore already signals that the choice of professional title had become a matter of debate in public discourse.

(39) Carola Rackete ha agito nel rispetto delle leggi internazionali dei diritti umani gerarchicamente superiori a quelli interne e a politici da strapazzo.
La giustizia rivolti il coltello con la punta verso i "capitani" e non

[36]Our translation: The investigating judge in Agrigento did not validate the arrest of #CarolaRackete, the.F commander (and not the "captain.F") of the #SeaWatch3, and did not impose any precautionary measure against her. The Italian state always knows how to command respect, that's for sure!
Applause.

[37]Incidentally, this is one of the three overlapping tweets featuring both *capitan-* and *comandante*

contro questa capitana eccezionale.[38]

This tweet stages a contrast between a generic masculine plural (*capitani*) and a singular feminine form (*questa capitana eccezionale*). The generic plural appears to refer broadly to those male political actors or role-holders who should be targeted instead, while the feminine singular individualises Rackete positively. The contrast is therefore formal and semantic, as we can assume that the referent is Salvini (or in general politicians associated to him), but the form "i capitani" is still to be considered a generic use of masculine morphological gender.

The following tweet is one of the most explicit examples in the corpus in terms of reflexive gender awareness:

(40) Ma Q U A N T O sarebbe diversa la narrazione di questi eventi, se al posto di una capitana ci fosse stato un capitano? Quali sarebbero stati i toni usati? #SeaWatch3 #CarolaRacketeLIBERA[39]

The juxtaposition between *capitana* and *capitano* is used as both a meta-discursive and a metalinguistic element to formulate a counterfactual question about media representation itself, making the issue of gendered narration overtly thematic. In this sense, the tweet directly addresses the broader question that motivated our case study.

(41) Hanno arrestato la #Capitana. Ha forzato il blocco. Ha la colpa di aver salvato 40 persone. Questo mondo va alla rovescia. E ritengo ci siano valori più importanti dei voti in politica Stima per il capitano #CarolaRackete #SeaWatch3 #SeaWatch #Lampedusa[40]

On the other side, in this example the feminine form is used to identify Rackete directly, whereas the final expression *Stima per il capitano* remains

[38]Our translation: Carola Rackete acted in compliance with international human rights law, which is hierarchically superior to domestic law and to second-rate politicians. May justice turn the knife's point towards the "captains.M" and not against this exceptional captain.F.

[39]Our translation: But H O W different would the narrative of these events be if, instead of a captain.F, there had been a captain.M? What tone would have been used? #SeaWatch3 #CarolaRacketeFREE

[40]Our translation: They have arrested the #Captain.F. She forced the blockade. She is guilty of having saved 40 people. This world is upside down. And I believe there are values more important than votes in politics. Respect for the captain.M #CarolaRackete #SeaWatch3 #SeaWatch #Lampedusa

more ambiguous and may be read either as a genericised masculine title or as a local reformulation of the same referent, the latter being the more plausible interpretation in context. Precisely because of this slight instability, similarly to the example (31) the tweet is especially interesting for the present analysis: it shows how masculine and feminine labels may coexist even when the overall stance is clearly supportive.

(42) Agli italiani che in queste ore uniscono maschilismo e razzismo, farei notare che la capitana Carola Rackete oltre a dare prova di umanità, sa perfettamente parcheggiare una nave.
#SeaWatch3 #Lampedusa #capitano #CarolaRackete #Iostoconcarola #stayhuman #26giugno[41]

Finally, (42) juxtaposes the feminine *capitana* in the body of the text with the hashtag #capitano. The co-occurrence is particularly striking because it appears in a tweet that explicitly denounces *maschilismo*. As such, it condenses several of the corpus' central themes at once: gender marking, meta-discursive stereotype-related awareness, and the instability of professional nomination in a polarised public debate, where the metalinguistic competence is not always taken for granted.

As we have seen above, in the corpus we found a small but clearly identifiable set of cases in which the nomen agentis becomes itself the object of discourse, rather than functioning merely as a referential label. These cases are concentrated mainly in the Rackete subset and typically take the form of explicit correction (*il comandante* e non la "capitana"), contrastive reflection between feminine and masculine forms (*se al posto di una capitana ci fosse stato un capitano*), or overt comments on the semantic and symbolic value of the title itself, as in tweets where *capitano* is presented as a designation presupposing honour or legitimacy.

In other words, the professional title does not simply identify the referent, but becomes a site of metalinguistic and ideological negotiation. By contrast, in the **Schettino** subset, the nomen is much more often used referentially—even if in strong evaluative or ironical contexts—and never becomes the object

[41] Our translation: To the Italians who at this moment are combining sexism and racism, I would point out that Captain.F Carola Rackete, besides showing humanity, knows perfectly well how to park a ship. #SeaWatch3 #Lampedusa #captain.M #CarolaRackete #IstandwithCarola #stayhuman #26June

of an explicit debate on the form itself. This asymmetry is particularly relevant for the present case study, since it suggests that the linguistic visibility of gender in the nomination of Rackete may transform the title into a discursive issue in its own right.

6.3.1.2 *Signorina*, *Signora*, *Carola*: Other Nomination Strategies

Alongside professional titles, the Rackete subset also contains other nomination strategies, in particular the use of *signora*, *signorina*, and the given name *Carola*. These forms are especially relevant because they shift the focus away from professional role and towards age, social status, familiarity, or gendered personhood. In other words, whereas *capitana*, *capitano*, and *comandante* position the referent primarily within a professional frame, *signora* and *signorina* tend to reframe her through forms of civility, hierarchy, or paternalistic diminution.

A closer look at the subset retrieved through the filter *signor** also shows that not all occurrences are equally relevant to the nomination of Rackete. Some tweets include figurative or non-referential uses, or refer to other political actors instead. For this reason, the discussion below focuses only on those cases in which *signora* or *signorina* function as actual nomination strategies directed at Rackete. Within this more restricted set, the distinction between the two forms is particularly revealing.

(43) La Signora #CarolaRackete affronterà la giustizia italiana basta parlarne che sia giusto o sbagliato ha infranto le Leggi .. avete rotto cambiate il disco le vostre petizioni valgano un fico secco[42]

In this first case, *Signora* appears to operate as a formally polite label, but the overall stance of the tweet is clearly hostile. The nomination strategy therefore does not mitigate criticism; rather, it frames the target within a discourse of legal blame and moral reproach. The form of address is formally respectful, yet pragmatically distant.

[42]Our translation: Mrs #CarolaRackete will face Italian justice enough talking about it whether it is right or wrong she broke the law .. you lot are tiresome change the record your petitions are worth absolutely nothing.

(44) L'ABC dell'educazione.
"Carola fatti stuprare dai negri" solo se sei un parente o un amico d'infanzia. Per tutti gli altri "Signora Rackete fatti stuprare dai negri". Ripartiamo da qui.[43]

This example is especially delicate, since the tweet reproduces an extremely violent formulation while explicitly commenting on forms of address. In spite of its metalinguistic dimension, the text remains strongly problematic at the level of discourse, and this is consistent with its annotation as *against* in the majority vote. What makes it relevant here is that the opposition between *Carola* and *Signora Rackete* is itself made explicit: first-name usage is associated with an illegitimate and aggressive familiarity, whereas *Signora Rackete* is framed as the minimum threshold of public respect. The tweet is therefore not neutral, but rather a case in which metalinguistic reflection is embedded within a highly violent utterance.

(45) La signora Carola Rackete deve sapere che siamo in Italia e in Italia le leggi si rispettano. Se magari trasportava un capo clan della 'Ndrangheta oppure aveva rubato 49 milioni agli italiani, beh allora un accordo si trovava. Invece così, in galera! #SeaWacht[44]

(46) #Salvini ha apostrofato la Signora #CarolaRackete con l'epiteto, fra gli altri, di #figliadipapà. Che dire. Se tutti i #figlidipapà avessero #solidarietà, sarebbe già #Rivoluzione.
#scritturebrevi[45]

The examples (45) and (46) are particularly useful because, although both contain the formal label *signora*, their supportive stance is established through different discursive mechanisms. In the first tweet, the apparent insistence on legality is clearly ironic: the comparison with a mafia boss and with the

[43]Our translation: The ABC of good manners. "Carola go and get raped by niggers" only if you are a relative or a childhood friend. For everyone else "Mrs Rackete go and get raped by niggers". Let us start again from here.

[44]Our translation: Mrs Carola Rackete must understand that we are in Italy and in Italy the law is obeyed. If perhaps she had been transporting an 'Ndrangheta clan boss or had stolen 49 million from Italians, well, then an arrangement would have been found. But like this, prison! #SeaWatch

[45]Our translation: #Salvini addressed Mrs #CarolaRackete with, among other things, the epithet #daddysgirl. What can one say. If all the #daddyskids had #solidarity, it would already be #Revolution. #shortwritings

reference to the 49 million euros (clear reference to the party Lega and Salvini) functions as an implicit accusation against unequal political and judicial treatment, so that the final "Invece così, in galera!" is not directed against Rackete in a straightforward way, but rather against the absurdity of the situation being denounced. In the second tweet, by contrast, the supportive stance is explicit and non-ironic: *Signora #CarolaRackete* appears in a tweet that defends her against Salvini's disparaging wording and resemanticises *#figliadipapà* positively through the association with solidarity. Taken together, these examples show that *signora* does not carry a stable negative orientation in the corpus. Rather, it functions as a formal nomination strategy that may occur in both openly supportive and irony-mediated supportive discourse.

(47) La questione Carola Rackete, non è chiusa. Ne sentiremo e vedremo delle belle. Abbiamo fede e vedremo (spero) la signorina nel posto dove merita di stare.[46]

(48) Sono giorni che sento e leggo di questa signorina #Rackete È plurilaureata, parla 4 lingue... Mo esce che parla solo la lingua madre e che mastica qualche lingua così per farsi capire, non è nemmeno capitano ecc ecc.. Mo fra poche ore sie viene a sapere che non è quello.....[47]

(49) Siete malati, tutti. La #SeaWacht3 OSA ENTRARE IN TERRITORIO ITALIANO NONOSTANTE IL DIVIETO E VOI ESULTATE? #trenta dorme? #salvini spedisci i clandestini in #olanda e da fuoco a quella bagnarola. E la signorina #CarolaRackete in GALERA. SUBITO.[48]

The use of *signorina* is much more homogeneous. In the relevant cases retrieved in the subset, it is consistently embedded in hostile, dismissive, or delegitimising discourse and seems to work as a diminutive or belittling form. The nomination strategy shifts the frame from professional competence to a

[46]Our translation: The Carola Rackete affair, is not over. We will hear and see some good ones. We have faith and we will see (I hope) the miss in the place where she deserves to be.

[47]Our translation: For days I have been hearing and reading about this Miss #Rackete She has multiple degrees, she speaks 4 languages... Now it turns out that she only speaks her mother tongue and that she knows a bit of some language, just enough to make herself understood, she is not even a captain, etc. etc.. Now in a few hours it will come out that she is not that either.....

[48]Our translation: You are all sick, all of you. The #SeaWacht3 DARES TO ENTER ITALIAN TERRITORY DESPITE THE BAN AND YOU CHEER? Is #trenta asleep? #salvini send the illegals to #holland and set fire to that tub. And Miss #CarolaRackete to JAIL. NOW.

gendered and age-marked image of the referent, often implying immaturity, impropriety, or lack of authority. In this sense, *signorina* appears much closer to a delegitimising strategy than to a neutral alternative of nomination.

(50) Comunque con quel "la signorina andrà un galera " il caro Salvini ha cercato di trasformare l'immagine di una donna forte e coraggiosa in quella di una ragazzina viziata,dimostrando solo che #CarolaRackete ha 2 palle che lui se le sogna. #freeCarola #SeaWach3[49]

This last example is particularly revealing because it makes the ideological function of *signorina* fully explicit. The writer comments on Salvini's wording and interprets it as an attempt to recast Rackete from "a strong and courageous woman" into "a spoiled little girl". The tweet is therefore not only supportive of Rackete, but also overtly metalinguistic: it identifies the nomination strategy itself as a discursive tool of gendered delegitimisation. This makes *signorina* especially relevant for the present case study, since it shows how nomination may operate not only referentially, but also as a means of reshaping public perception of authority, maturity, and credibility.

The contrast between *signora* and *signorina* is therefore not merely one of degree of formality. In the subset, *signora* remains comparatively flexible and may be embedded in both supportive and hostile discourse, whereas *signorina* is much more clearly associated with dismissal, paternalism, or explicitly sexist framing. This asymmetry is particularly important because it shows that the shift away from professional titles is not neutral: once the nomination leaves the sphere of role and enters the sphere of gendered civility, the available forms do not behave in the same way.

More broadly, these examples confirm that nomination strategies can themselves become objects of public reflection. Some tweets do not merely use a label, but comment on the choice of label as socially meaningful. In this sense, the distinction between *signora*, *signorina*, and first-name usage participates in the same broader meta-discursive space already observed for *capitana*, *capitano*, and *comandante*: a space in which naming is not only referential,

[49] Our translation: Anyway, with that "the young lady will go to prison " dear Salvini tried to turn the image of a strong and courageous woman into that of a spoiled little girl, only proving that #CarolaRackete has balls he can only dream of. #freeCarola #SeaWach3

but ideological, evaluative, and tied to public struggles over legitimacy and representation.

Finally, we isolated the texts where at least once the given name *Carola* occurs.

This further step is especially relevant because the use of the given name alone may signal a nomination strategy that differs from both professional titles and more formal combinations such as *Carola Rackete*. In studies on gender and political naming, first-name reference has repeatedly been treated as a potentially informal or deprofessionalising strategy. In particular, Uscinski and Goren (2011)'s analysis of media coverage during the 2008 Democratic primary found that Hillary Clinton was referred to by first name more often than Barack Obama, and argued that this asymmetry could not be reduced simply to campaign branding, since male candidates who branded themselves through their first name were not treated in the same way by journalists. Likewise, Marjanovic, Stańczak, and Augenstein (2022) show, on a much broader corpus of online political discussion, that female politicians are much more likely than men to be named by their given name, whereas male politicians are more often referred to by surname, a pattern interpreted as less professional and less respectful treatment of women in politics.

In our material, however, it is again necessary to distinguish between two different configurations. The first includes tweets in which the given name appears in hashtagged form, i.e. as *#Carola*. The second includes tweets in which *Carola* appears in the running text without a preceding hashtag.

In the reduced preliminary and final subsets considered jointly, the given name occurs in 48 tweets overall. Of these, 13 contain the hashtagged form *#Carola*, while 35 contain *Carola* without a preceding hashtag. Within this latter group, 24 tweets use the given name without the surname, whereas 11 contain the full form *Carola Rackete*. The contrast is analytically important, because the non-hashtagged given-name-only uses constitute the strongest evidence that the protagonist is being discursively re-framed through familiarity rather than professional title.

The hashtagged cases are not irrelevant. On the contrary, they show that the given name had already become available as a recognisable public tag. This is visible in tweets such as the following:

(51) ++ Niente arresto per #Carola. Il gip non convalida. ++

> Non è sono stati contestati i reati. Finalmente la realtà dei fatti. #freeCarolaRackete[50]

Here #*Carola* functions as a compact public label within a clearly supportive tweet. The hashtag contributes to the circulation of the referent as an emblematic figure and is less obviously a case of interpersonal familiarity than a form of political indexing. A similar mechanism can be observed in formulations such as #*Carola libera*, where the given name becomes part of a slogan-like formulation. In this respect, the hashtagged uses are indeed significant, but they should not automatically be equated with the more spontaneous use of the bare given name in running text.

The non-hashtagged occurrences are more revealing in this sense. Consider, for instance, the following tweet:

(52) È che in Italia siamo abituati a Schettino, una come Carola ci mette davanti alla nostra pochezza umana #SeaWacht #Capitana #CAROLA_RACKETE [51]

Here *Carola* appears unaccompanied by surname or title in the body of the tweet, and the effect is one of clear proximity and personalisation. The contrast with *Schettino*, who is referred to by surname, is especially important: the asymmetry between the two protagonists is not merely morphological, but nominative. Rackete is brought closer through the given name, whereas Schettino remains publicly distanced through surname usage.

(53) Sbarcati. Finalmente. Con segni di violenza sulla pelle e sull'anima, ma su terra ferma.
Un abbraccio alla mia capitana Carola.
Caro, du hast das Richtige getan und wir stehen hinter dir. Danke.[52]

In this example, the given name is integrated into a strongly affective and supportive frame. The phrase *la mia capitana Carola* combines professional role and

[50] Our translation: ++ No arrest for #Carola. The investigating judge does not validate it. ++ The charges have not been upheld. At last, the reality of the facts. #freeCarolaRackete

[51] Our translation: It is just that in Italy we are used to Schettino, someone like Carola confronts us with our human smallness #SeaWacht #Captain.F #CAROLA_RACKETE

[52] Our translation: Disembarked. At last. With signs of violence on their skin and on their soul, but on dry land.
A hug to my captain.F Carola. *Caro, du hast das Richtige getan und wir stehen hinter dir. Danke.*

personal familiarity, producing a nomination strategy that is at once admiring and intimate. This is particularly noteworthy because it shows that the given name does not necessarily displace the professional title altogether; rather, the two may co-occur and jointly construct a figure who is both competent and emotionally close.

(54) Sono contenta che ci sono ancora persone che sono dalla parte dell'umanità e, come volevasi dimostrare, della giustizia. Io sto Carola. #CarolaRackete[53]

A formula such as *Io sto Carola* goes even further in this direction. The use of the first name alone is here plainly affiliative and slogan-like, but without the formal mediation of the hashtag. This makes it a particularly strong case of direct first-name nomination.

(55) Carola attaccata su estetica e igiene personale da gente che non l'ha mai vista di persona. E comunque meglio la pelle sporca che la coscienza.[54]

This tweet is also important because the first-name use co-occurs with a defence against attacks on appearance and hygiene. The nomination strategy therefore participates in a discourse that is simultaneously solidaristic and gendered: Rackete is defended as *Carola*, while the attack she is subjected to concerns bodily and aesthetic dimensions that are themselves highly gendered in public discourse.

Not all bare first-name uses are supportive, however. The corpus also includes hostile or problematic cases such as:

(56) Carola perchè queste bellissime e gloriose missioni di salvataggio non le hai fatte in Germania quando ce n'era bisogno? MHA #ArrestateCarolaRackete[55]

[53]Our translation: I am glad that there are still people who are on the side of humanity and, as was to be expected, of justice. I stand with Carola. #CarolaRackete

[54]Our translation: Carola attacked over her looks and personal hygiene by people who have never seen her in person. And in any case, better dirty skin than conscience.

[55]Our translation: Carola, why did you not carry out these beautiful and glorious rescue missions in Germany, when there was a need for them? MHA #ArrestCarolaRackete

In this case, the given name is used in a direct accusatory frame. The effect is not solidarity but confrontational address. This is important because it shows that first-name nomination is not in itself sufficient to determine stance. Nevertheless, even here the use of the bare given name contributes to a less professional framing than one based on title or surname.

At the same time, a small group of cases uses the full form *Carola Rackete*. These are often more informational, report-like, or less strongly affective than the bare given-name cases, for instance in sentences such as "*Carola Rackete è libera*" or "*Beh, che dire: Carola Rackete infonde fiducia nell'umanità*". In these occurrences, the referent is still identified through the given name, but the co-presence of the surname partially restores public formality. This middle ground is also consistent with Marjanovic, Stańczak, and Augenstein (2022)'s observation that women in political discourse are frequently referred to by full name rather than surname alone.

A final point concerns comparison with Schettino. Even allowing for the possibility that *Carola* may have had some mnemonic salience as a relatively less common name in Italy, the nominative asymmetry remains difficult to ignore. In the material considered here, the given name becomes a recurrent and publicly available way of indexing Rackete, both in hashtagged and non-hashtagged form. By contrast, no comparable first-name pattern emerges for Schettino: although isolated occurrences of *Francesco Schettino* do exist elsewhere in the corpus, *Francesco* does not become a productive or recognisable nomination strategy in the same way that *Carola* does. This suggests that the prominence of the given name in the Rackete case is not merely accidental, and is better understood as part of a broader asymmetry in the public naming of the two protagonists. This confirms that the given name is an important part of the broader nomination patterns through which gendered public discourse is articulated.

It is also worth making the relation between first-name nomination and annotated stance more explicit. In the subset, the hashtagged form *#Carola* appears in 13 tweets, of which 8 are annotated as *pro*, 2 as *against*, 2 as *tie*, and 1 as *neutral*. The non-hashtagged occurrences of *Carola* are more numerous (35 tweets overall) and are likewise predominantly supportive, with 24 *pro* cases, 6 *against*, 3 *neutral*, and 2 *tie*. This does not mean that first-name nomination is in itself sufficient to predict stance. On the contrary, the corpus clearly

shows that the same strategy may support solidarity, confrontational address, irony, or criticism. Nevertheless, the predominance of supportive contexts is significant, especially in the non-hashtagged cases, because it suggests that the given name is frequently used to construct proximity, alignment, or affective identification with Rackete. At the same time, this supportive tendency should not obscure the fact that first-name usage still contributes to a less formal public framing than one based on title or surname alone.

Taken together, the three nomination strategies examined in this section—professional title, civility form, and given name—point to three different ways of gendering Rackete in public discourse. *Capitana* keeps the referent within the sphere of professional role, even when the form itself becomes an object of ideological dispute. *Signorina*, by contrast, shifts the frame towards infantilisation, paternalism, or overt delegitimisation, and is the clearest case of a nomination strategy whose evaluative load is strongly gendered. The given name *Carola* occupies an intermediate but no less important position: it often constructs solidarity, emotional proximity, or symbolic identification, yet it also contributes to moving the protagonist away from a strictly professional register and into a more personalised one. In this sense, these forms should not be treated as interchangeable labels. Rather, they activate different discursive framings of authority, legitimacy, familiarity, and public visibility. Considered together, they reinforce what the literature on political naming has shown in other contexts: that the way women are named in public discourse is not a neutral stylistic detail, but part of the broader symbolic negotiation of their status and credibility.

6.3.1.3 A Qualitative Note on the Stereotype-labelled Cases

Although stereotype-labelled cases do not constitute the dominant pattern of the corpus, a brief qualitative inspection of them remains useful. Rather than representing the main explanatory layer of the dataset, these tweets can be read as particularly revealing sites in which more explicit forms of gendered, sexualised, or socially stereotyped evaluation surface. Considered after the analysis of nomination strategies, they help clarify whether stereotype annotation captures a distinct discursive layer or overlaps with mechanisms already observed qualitatively.

The two annotation phases are not perfectly identical in their internal taxonomy, since the final schema distinguishes targets and domains more analytically than the preliminary one. Nonetheless, the two datasets converge on a common point: stereotype-labelled cases tend to cluster around a limited number of recurrent patterns. In particular, the **Rackete** subsets more often activates stereotypes concerning women, appearance, behaviour, and, in some cases, migrants and victims; the **Schettino** subsets, by contrast, often reaches stereotype annotation not only through the protagonist himself, but through secondary female figures, sexual gossip, or broader cultural scripts.

A first cluster is clearly visible in the Rackete-related texts and concerns the gendering of professional competence. In a number of tweets, stereotype annotation is triggered not because the protagonist is attacked in generic terms, but because her role is reframed through a feminised script of incompetence, trivialisation, or displacement from the professional sphere.

(57) #CarolaRackete non aveva nessuna cattiva intenzione verso i finanzieri. Stava solo parcheggiando[56]

This tweet is particularly interesting because the stereotype lies in the apparently light or ironic wording. The expression *stava solo parcheggiando* does not simply describe the manoeuvre, but activates a familiar and gendered script in which driving or parking competence is implicitly feminised and trivialised. The fact that the tweet was annotated as stereotype-bearing, despite its relatively compact formulation, suggests that some stereotypical patterns in the corpus are not overt insults but compressed allusions to very well-known clichés.

(58) Agli italiani che in queste ore uniscono maschilismo e razzismo, farei notare che la capitana Carola Rackete oltre a dare prova di umanità, sa perfettamente parcheggiare una nave.[57]

In this case, the same semantic field is explicitly recontextualised. The tweet is supportive, yet it still presupposes the circulation of a sexist script strong

[56] Our translation: #CarolaRackete had no bad intentions towards the finance police. She was only parking.

[57] Our translation: To the Italians who, at this moment, are combining sexism and racism, I would point out that captain.F Carola Rackete, besides showing humanity, knows perfectly well how to park a ship.

enough to require explicit rebuttal. This is a useful reminder that stereotype annotation in the corpus does not correspond only to hostile tweets: some *pro* texts are labelled as stereotype-bearing precisely because they quote, reverse, or expose a stereotypical frame already present in public discourse.

A second Rackete-related cluster concerns appearance, body, and sexualisation. These are the cases in which the stereotype becomes more explicit and often overlaps with nomination, especially when the protagonist is pulled away from the professional frame and reinserted into a discourse of attractiveness, bodily adequacy, or heterosexual desirability.

(59) Carola dammi retta, tagliati i capelli, comprati un bel vestitino e trovati un fidanzato .. che non sei bellissima e fai anche un po' paura.[58]

Here the stereotype is overt. The protagonist is directly repositioned within a normative script of femininity: hair, dress, heterosexual couplehood, and bodily attractiveness are all mobilised in order to delegitimise her public presence. The point is not simply that Rackete is insulted, but that her authority is displaced into a discourse about how a woman should look and behave.

(60) Per #Libero il fondamentale 'dettaglio sfuggito a molti' è che #CarolaRackete si è presentata in Procura senza reggiseno.[59]

This tweet is again supportive, but it is stereotype-bearing because it makes visible a sexualised and gendered framing already circulating in the media. As in the previous example, the issue is not professional action but bodily presentation. What emerges from these cases is that, when stereotype becomes explicit in the **Rackete** subsets, it often does so through attempts to re-anchor the protagonist in a regime of appearance, sexuality, or domesticised femininity.

A third and partly distinct cluster in the Rackete data concerns broader social stereotypes, especially around migrants and victims. In these cases, the protagonist is not necessarily the sole or even the main target of the stereotype,

[58]Our translation: Carola, listen to me, cut your hair, buy yourself a nice little dress and find yourself a boyfriend .. because you are not very beautiful and you are also a bit scary.

[59]Our translation: For #Libero, the crucial 'detail that many missed' is that #CarolaRackete showed up at the Prosecutor's Office without a bra.

but the tweet is nevertheless attached to the Rackete event and contributes to its discursive environment.

(61) Ma l'Italia vi sta tanto in culo e pensate che l'unica salvezza per la civiltà siano gli africani, perché non vi trasferite in Africa? Magari facendo un po' di volontariato, che dal salottino di casa siamo tutti umanitari.
#salvini #SalviniNonMollare #migranti #CarolaRackete[60]

This example is useful because it shows that the stereotype label may also capture the ideological field around Rackete rather than her person alone. Here the stereotypical construction concerns Africans and humanitarian discourse, yet it appears within the same polarised communicative space in which Rackete is discussed. In this sense, the content analysis confirms that the event is embedded in a broader discursive constellation where gendered attacks coexist with racialised and anti-migrant stereotypes.

The **Schettino** subsets present a different profile. Here stereotype-labelled cases often do not centre straightforwardly on the male protagonist as bearer of a gender stereotype. Rather, they frequently involve secondary women, sexual scripts, or wider cultural imaginaries. A recurrent nucleus concerns the figure repeatedly referred to as *la moldava*, whose presence becomes a vehicle for sexualisation and blame displacement.

(62) Ancora si parla di #schettino e la #moldava, suu si sà che le storie vacanziere sono destinate a finire![61]

(63) Schettino non ha fatto affondare il concordia... E stata la moldava.. #lebugie[62]

(64) #schettino e alla fine era colpa di una donna.[63]

[60]Our translation: But is Italy so far up your arse and do you think that the only salvation for civilisation is Africans, why don't you move to Africa? Maybe by doing a bit of volunteering, because from the little sitting room at home we are all humanitarians. #salvini #SalviniNonMollare #migranti #CarolaRackete

[61]Our translation: People are still talking about #schettino and the #Moldovan woman, come on, it is well known that holiday flings are bound to end!

[62]Our translation: Schettino did not sink the Concordia... It was the Moldovan woman.. #lies

[63]Our translation: #schettino and in the end it was a woman's fault.

These examples show that stereotype annotation in the Schettino subset often arises through narratives in which a woman—which incidentally was a simple passenger of the cruise ship—is sexualised, nationalised, or turned into a convenient explanatory figure. The stereotype is therefore less often attached directly to Schettino as a man, and more often mediated through the women associated with him or through gendered clichés about seduction, distraction, and blame.

Another recurrent Schettino-related pattern concerns a more explicit sexual innuendo and vulgarisation:

(65) Film time, Honey... Pubblicità time -.- vaffanculo a Schettino e alla sua notte di sesso con la tipa bionda #concordia[64]

(66) Altro ke cocaina! Schettino era strafatto di Viagra per la notte in cabina colla moldava: vedeva tutto blu, cosi' questo e' il mistero del perche' nn manovrava piu'.'[65]

(67) Schettino trombava la moldava.. Ora tutto ha più senso ...[66]

These tweets reduce the event to a sexualised and sensationalist script, in which responsibility is reimagined through vulgar gossip and hypermasculine innuendo. Here the stereotype label captures a type of discursive degradation that is different from the Rackete cases: whereas in the latter the protagonist is frequently repositioned through appearance or femininity, in the Schettino subset the stereotypical force often lies in the transformation of the shipwreck into a story of male sexual excess and female distraction.

A smaller but still relevant group in the Schettino material points towards broader cultural stereotyping:

(68) rivedendo i video al tg5, schettino è l'emblema dell'italiano...prende le decisioni dopo un bel pò e pure da far schifo[67]

[64]Our translation: Film time, Honey... Advert time -.- fuck Schettino and his night of sex with the blonde girl #concordia

[65]Our translation: Not cocaine at all! Schettino was high on Viagra for the night in the cabin with the Moldovan woman: he was seeing everything blue, so this is the mystery of why he was no longer steering.

[66]Our translation: Schettino was fucking the Moldovan woman.. Now everything makes more sense ...

[67]Our translation: watching the videos on TG5 again, Schettino is the emblem of the Italian man... he makes decisions after quite a while and, on top of that, in a disgusting way

(69) l'taliano medio si sfama dei mostri della tv: dopo zio misseri, ecco schettino.[68]

In these examples, the protagonist becomes a figure through which a wider stereotype about "the Italian" is articulated. This is particularly interesting because it differs from the more local gendered dynamics seen for Rackete. Schettino is less often framed through explicit masculine stereotypes than through national typification, ridicule, or media monstrosity.

Altogether, these stereotype-labelled cases are useful precisely because they are not the dominant pattern of the corpus. Their relative scarcity suggests that the asymmetry between Rackete and Schettino is not driven primarily by a dense concentration of overt stereotype labels throughout the data. Rather, much of the gendered difference identified in this case study seems to emerge through more diffuse linguistic mechanisms and discursive framings, such as subtler nomination strategies, professional titles, and metalinguistic debate.

6.3.2 The Post-Annotation Questionnaire

The content analysis presented in the previous subsection supports one of the broader claims of the present chapter: the differential treatment of the two protagonists is visible not merely in what is said about them, but in how they are named, categorised, compared, and made socially legible within public discourse. Instead, this subsection shifts the focus from the annotated texts to the annotation process itself, in order to examine how the annotators experienced, interpreted, and reflected upon these same complexities while working on the corpus.

After the annotation process was completed, the three annotators were invited to fill in a post-annotation questionnaire. The adoption of a post-annotation questionnaire aligns with methodological approaches that conceptualise annotator disagreement as an expected and informative feature of subjective annotation tasks rather than as noise to be eliminated. Building on this perspective (see Sandri et al., 2023), the analysis of annotators' reflections and self-reported experiences can further shed light on sources of disagreement,

[68] Our translation: the average Italian feeds on the monsters on TV: after Uncle Misseri, here comes Schettino.

interpretative strategies, and the perceived clarity of the annotation schema. Accordingly, the questionnaire was not designed to resolve disagreement, but to contextualise it by making explicit the annotators' familiarity with the topics, their prior experience, and the difficulties they perceived during the task. More broadly, the use of such a questionnaire is consistent with the overall orientation of this thesis, which treats annotation as an interpretative and socially situated practice rather than as a purely technical operation.

6.3.2.1 Description of the questionnaire

The questionnaire was administered via email through a Google Form (see Appendix D) and aimed at collecting reflective feedback on the annotation process from the three annotators. It consisted of three main sections, combining socio-demographic information with questions concerning familiarity with the annotation schema and reflections on the annotation process.

6.3.2.1.1 Introductory note and disclaimer. At the beginning of the form, annotators were informed that the questionnaire represented the final step of the annotation process in which they had been involved. They were thanked for their time, availability, and active participation. The questionnaire was presented as a tool to collect socio-demographic information and reflective insights on the annotation experience and the annotation schema. It was specified that the responses would be used to contextualise the analyses and to enhance the transparency and reflexivity of the research process. All data were to be treated confidentially and used exclusively for research purposes, in compliance with Regulation (EU) 2016/679 (GDPR).

6.3.2.1.2 Section A - Personal information. This section collected basic socio-demographic and background information. Annotators were asked to provide the following information:

- age range (18–24; 25–34; 35–44; 45+);
- gender identity, selecting the option that best represented them among several predefined categories, with the possibility to choose "Prefer not to say" or to specify an alternative option;

- main educational or training background, choosing among Linguistics, Computational Linguistics, Computer Science/AI/Data Science, Social or Human Sciences, or an open "Other" option.

6.3.2.1.3 Section B - Familiarity with the annotation schema. The second section focused on annotators' familiarity with the annotation schema and with the topics addressed in the corpus. It included the following items:

- a self-assessment of perceived difficulty for each annotation category and label, rated on a five-point Likert scale (1 = very easy; 5 = very difficult); the list of categories and labels was reported again to facilitate recall, as several weeks had elapsed since the end of the annotation task;
- prior experience with linguistic annotation, with one choice among: no prior experience; limited experience (1–2 projects); moderate experience (3–5 projects); extensive experience (more than five projects);
- involvement in other annotation activities in parallel with this project (yes, one; yes, more than 1; no);
- an open follow-up question (if applicable) asking whether and how parallel annotation activities influenced the annotation process in this project;
- a self-assessment of initial familiarity with a set of relevant topics, rated on a five-point scale (1 = no familiarity; 5 = high familiarity), including:
  - gender stereotypes;
  - migration-related discourse;
  - intersectionality;
  - the Costa Concordia shipwreck and Francesco Schettino;
  - the Sea-Watch-3 case and Carola Rackete;
- a final question asking whether annotators believed that their personal sensitivity towards these topics had influenced their annotation decisions, with four possible answers ranging from "Not at all" to "Very much".

6.3.2.1.4 Section C - Annotation process. The final section addressed the annotation process more directly and included the following items:

- a second self-assessment of perceived difficulty for each annotation category, again rated on a five-point Likert scale (1 = very easy; 5 = very difficult);
- an open question asking whether annotators recalled experiencing uncertainty with specific labels, which labels these were, and which strategies they believed they had adopted to deal with such uncertainty;
- a multiple-choice question on the perceived main sources of disagreement among annotators, in light of the collective discussion and the subsequent revision of the annotation schema, with the following options:
  - textual ambiguity;
  - lack of information or context (e.g. incomplete texts);
  - subjectivity (e.g. personal interpretations or biases);
  - limited clarity of the annotation guidelines;
  - other (with specification);
- an optional open follow-up question allowing annotators to briefly motivate their choice;
- a final optional open field for additional comments on the annotation experience, for instance to suggest possible improvements to the annotation schema.

6.3.2.2 Annotators' responses

All three annotators completed the questionnaire (results are displayed in the following tables. All three reported **limited** prior experience in linguistic annotation (*1–2 projects*). Two annotators reported having carried out one parallel annotation activity during the same period; both explicitly stated that it did not influence the present task (one of them motivating this by noting that the topics were substantially different).

Annotator	Age range	Gender identity	Background	Prior annotation	Parallel tasks	Influence
A1	18–24	Cis woman	Linguistics	Limited	No	—
A2	25–34	Cis woman	CS / AI / DS	Limited	Yes, one	No
A3	18–24	Cis man	Linguistics; CL	Limited	Yes, one	No (topics distant)

TABLE 6.55: Annotators' background and prior experience.

With respect to initial topic familiarity (scale 1–5), the three annotators reported values ranging from 3 to 5 for gender stereotypes, from 3 to 4 for migration-related discourse, and a constant value of 3 for intersectionality. Familiarity with the two case-related backgrounds varied more substantially: for Schettino/Costa Concordia the values ranged from 2 to 5, and for Rackete/Sea-Watch-3 from 1 to 3. When asked whether personal sensitivity towards the topics influenced annotation decisions, two annotators answered *Abbastanza* ('enough') and one answered *Molto* ('a lot').

Annotator	Gender stereotypes	Migration discourse	Inter-sectionality	Schettino case	Rackete case
A1	5	4	3	4	1
A2	4	4	3	5	3
A3	3	3	3	2	1

TABLE 6.56: Annotators' initial familiarity with the topics (1–5).

Annotator	Influence reported
A1	Moderately
A2	Strongly
A3	Moderately

TABLE 6.57: Perceived influence of personal sensitivity on annotation decisions.

The questionnaire also included seven numerical items (scale 1–5) eliciting perceived difficulty for the annotation categories ***Target***, ***Responsibility***, ***Offensiveness***, ***Stance***, ***Humour***, ***Competence***, and ***Stereotype***. The reported

values were as follows: Annotator 1 (2, 1, 2, 2, 3, 1, 1), Annotator 2 (3, 2, 1, 2, 1, 2, 4), and Annotator 3 (3, 2, 2, 2, 3, 3, 4).

Annotator	Target	Responsibility	Offensiveness	Stance	Humour	Competence	Stereotype
A1	2	1	2	2	3	1	1
A2	3	2	1	2	1	2	4
A3	3	2	2	2	3	3	4

TABLE 6.58: Perceived difficulty by annotation category (1–5).

All three annotators also provided an open response about uncertainty during annotation. Annotator 1 reported initial uncertainty about identifying the target and described difficulty in distinguishing whether some tweets were serious or humorous, especially when contextual information was missing. Annotator 2 mentioned uncertainty related to the ***Stereotype*** label, at times perceived as "velato" ("veiled, covert") or "forzato" ("forced") , as potentially influenced by sensitivity to the topic. Annotator 3 reported early uncertainty about how to interpret ***Responsibility***, noting that "l'attribuzione di responsabilità può infatti apparire inizialmente come qualcosa di negativo, ma in realtà è possibile riconoscere la responsabilità di un'azione o di un evento senza necessariamente darle una connotazione negativa"[69].

When asked about possible sources of disagreement, the options selected were: (i) lack of information/context (Annotator 1); (ii) textual ambiguity and subjectivity (Annotator 2); and (iii) textual ambiguity and limited clarity of the guidelines (Annotator 3). Two annotators provided an additional motivation. Annotator 1 insisted on context, noting that it would have been useful to have access to preceding tweets in a thread. Annotator 3 attributed disagreements to textual ambiguity and, in the early phase, to guidelines perceived as unclear or incomplete, later resolved through discussion and revision.

[69]Our translation: attributing responsibility can indeed initially appear as something negative, but it is actually possible to recognise the responsibility of an action or an event without necessarily giving it a negativ connotation

Annotator	**Sources selected**
A1	Lack of information / context
A2	Textual ambiguity; Subjectivity
A3	Textual ambiguity; Limited guideline clarity

TABLE 6.59: Perceived sources of annotator disagreement.

Lastly, two annotators left a final comment.

Annotator 1 described the experience as positive and the task as progressively more fluent:

> La mia esperienza è stata positiva, e non avrei apportato cambi significativi allo schema di annotazione. I tweet erano tanti, e a volte ripetitivi, ma l'annotazione diventava sempre più fluida e non ho avuto grandi problemi a terminare il tutto nonostante gli altri impegni.[70]

Annotator 3 highlighted increasing effectiveness over time, reflected in higher annotation speed and appreciation for the subdivision of the tweets:

> Penso che, a un certo punto della task, siamo arrivati ad avere uno schema di annotazione e un background di conoscenze acquisito durante il lavoro sufficienti per svolgere il compito in modo efficace. Questo è evidente anche dall'aumento progressivo della velocità di annotazione, man mano che prendevamo confidenza sia con lo schema sia con l'argomento in generale. Inoltre, ho particolarmente apprezzato la suddivisione dei tweet, che ha reso la task meno monotona.[71]

[70] Our translation: My experience was positive, and I would not have made significant changes to the annotation scheme. There were many tweets, and at times they were repetitive, but the annotation became increasingly smoother and I did not have major difficulties in completing everything despite my other commitments.

[71] Our translation: I think that, at a certain point in the task, we reached a stage where we had an annotation scheme and a background of knowledge acquired during the work that were sufficient to carry out the task effectively. This is also evident from the progressive increase in annotation speed, as we became more familiar both with the scheme and with the topic in general. Moreover, I particularly appreciated the subdivision of the tweets, which made the task less monotonous.

6.3.2.2.1 Interpretative remarks These responses support the decision to complement agreement measures with reflexive instruments: even when annotators share similar prior experience levels (all reporting *1–2 projects*), they highlight different *loci* of uncertainty (e.g. missing context for humour and stance; borderline or implicit stereotypes; conceptual interpretation of responsibility). At the same time, they generally recognised the usefulness of iterative meetings and feedbacking, not only for the clarity of the annotation task but for the practical sustainability as well. This reinforces the view that, in socially sensitive tasks—moreover if complex as the present one—it is true that disagreement is shaped by the interaction between textual underspecification, interpretative labour, and the evolving calibration of guidelines. It is also clear that a dialogical and participative approach can increase annotators' motivation and clarity and—at the same time—it can keep researchers more aware of the impact the annotation schema in its design.

6.4 Final Discussion

This chapter investigated how two protagonists occupying the same professional role—Francesco Schettino and Carola Rackete—were constructed in Italian Twitter discourse, while at the same time testing a multidimensional annotation process designed to capture stereotypes together with related evaluative and pragmatic dimensions. Taken together, the results show that the asymmetry between the two cases cannot be reduced to stance alone. Rather, the **Schettino** subsets are more frequently characterised by humour, ridicule, figurative reuse, and strongly individualised blame, whereas the **Rackete** subsets are more often marked by political polarisation, secondary targets, and conflicts over legitimacy and public responsibility.

From a linguistic point of view, the qualitative analyses suggest that these differences emerge not only through explicit stereotype labels, but also through broader discursive mechanisms. In particular, the Rackete case proved especially revealing with respect to nomination strategies, including the variable use of professional titles, civility forms such as *signora* and *signorina*, and first-name reference. These forms do not behave as neutral alternatives, but rather activate different framings of authority and familiarity.

By contrast, the Schettino case more frequently turns the protagonist into a figurative and reusable public type, often through humour and analogy.

Methodologically, the chapter also supports a broader point. Besides the resulting corpus, its contribution lies in the transparent documentation of an iterative annotation process in which schema design, pilot testing, structured discussion with annotators, preliminary analysis, revision, and post-annotation reflection all formed part of the research itself. In this sense, the study is clearly aligned with the orientation of FATA, although only partially: rather than treating annotation as a downstream technical step, it moves reflexive design and interaction with annotators into the core of the pipeline. What emerges from this process is that the annotation of socially sensitive discourse cannot be treated as a purely technical operation, but must be approached as an interpretative and situated practice. The manually annotated dataset produced here—2,000 tweets in total, equally divided between the two subcorpora—is relatively limited in size compared with larger NLP resources, but it was not designed primarily to maximise scale. Its main purpose was to test the analytical usefulness and practical limits of a fine-grained schema applied to a clearly delimited case study, while preserving the conditions under which judgements were produced and revised as situated information.

For this reason, the present work should be understood as both a linguistic and a methodological proposal. The chapter suggests that stereotypes are more effectively studied when embedded in a wider annotation architecture that also captures stance, humour, responsibility, offensiveness, and competence-related evaluation. At the same time, it shows that not all dimensions are equally stable. Categories such as *Target*, *Stance*, and *Humour* proved comparatively more robust, whereas *Responsibility*, *Offensiveness*, and especially *Stereotype* remained more interpretatively demanding. This means that the schema may be reusable in other contexts, but only cautiously and with renewed pilot testing, contextual adaptation, and careful calibration of the less stable dimensions.

A related point concerns scalability. Some components of the process documented here—such as iterative schema design, pilot annotation, revised guidelines, and the combination of quantitative and qualitative validation—are clearly transferable to other projects. Other aspects, however, are less easily scalable, especially those that depend on frequent interaction with annotators

and close discussion of borderline cases. This limitation should not be reduced to a merely practical inconvenience. For phenomena such as stereotypes or pragmatic role nomination, scale alone does not necessarily produce better data: once annotation is detached from the conditions of interpretation, it becomes easier to lose precisely those nuances that make the object of study analytically meaningful. In this respect, the chapter speaks to a broader concern already raised in NLP and dataset scholarship, namely that context and label variation are not secondary to data collection, but part of what makes a dataset scientifically usable. The present study therefore does not propose a ready-made large-scale protocol, but rather a model of controlled extensibility: one that can be adapted to further phenomena, provided that interpretative complexity is not flattened in the name of efficiency or apparent robustness alone.

Several limitations must nonetheless be acknowledged. First, the annotation process relied on a small group of annotators, which limits the possibility of statistically generalisable conclusions. Agreement and disagreement must therefore be interpreted primarily as indicators of task complexity and interpretative variability. Second, the MR was strongly involved throughout the process, having designed the schema, contributed to the pilot phase, annotated a subset of the data, and coordinated discussion with annotators. This does not invalidate the process, but it requires explicit acknowledgement of the researcher positionality. Third, the corpus is restricted to two specific events, embedded in different historical and political contexts, and limited to a single social media platform. The findings should therefore not be understood as representative of online discourse in general, but as a situated analysis of a theoretically motivated case study.

These limitations define the scope of the chapter rather than undermine its contribution. Future research may build on this work in several directions: by extending the annotation process to larger datasets, by testing the schema on other socially sensitive phenomena, and by further exploring which dimensions are more amenable to computational modelling and which instead continue to require stronger qualitative or human-in-the-loop support.

Overall, Chapter 6 shows that the differential treatment of Rackete and Schettino is visible not only in explicit evaluation, but also in how the two protagonists are named, framed, and made socially intelligible within public

discourse. At the same time, it shows that the construction of an annotation schema for such phenomena is itself an exploratory and reflexive process. In this sense, the chapter offers both a situated empirical contribution and a methodological argument for treating annotation as a genuinely interpretative practice.

Chapter 7

Case Study 2: Reclaiming Slurs in the LGBTQ+ Italian Context: the FAVOLOSA project

> "[...] su di me preferisco il frocio semplice. [...] ha tutta una tradizione, una storia [...] se non lo utilizzassi è come se si perdesse."
>
> *Focus group participant*

This case study is part of the *FAVOLOSA—Fair Automatisation and Visibility Of the LGBT+ Community On Slur (Re)Appropriation* project, which, in turn, represents the empirical part of the broader interdisciplinary body of work within which the FATA paradigm was developed (as described in Chapter 4, Section 4.5).

The FATA paradigm proposes an alternative NLP pipeline that integrates multidisciplinary perspectives, especially intervening in the early stage of the pipeline, e.g. through the implementation of techniques typically used in Sociolinguistics—such as focus groups and surveys—before the phases of data collection and annotation (Ferrando et al., 2025). Particularly suitable for studies that focus on social linguistic phenomena, the involvement of the linguistic community (which can be intended as a community of practice, following the definition given in Section 2.2.1) since the early stage of the research aims to correct for potential researchers' biases and supports the development of contextually grounded linguistic categories (Grieve et al., 2025). The engagement of individuals and communities is moreover conceived as fundamental for a more collaborative definition of the subject of the study,

in order to ensure that data collection is both inclusive and ethically sound.

The case study presented in this chapter is an extension of the survey-based study described in Marra et al., 2026, here presented in a more comprehensive and detailed manner. This case represents a first attempt to operationalise the FATA paradigm. It deals with the investigation of speakers' perception of reclamatory uses of slurs in the LGBTQ+ Italian context, and it is conceived as an initial step to gain the necessary knowledge to inform NLP studies on reclamation and to provide a crucial foundation for developing detection engines that are sensitive to this phenomenon. It is not only concerned with whether reclamatory uses can be identified, but also with how their extent and boundaries may be described in a way that is analytically useful for annotation and subsequent computational tasks.

The most relevant technique used in our study is the sociolinguistic survey—consisting of a questionnaire with open and closed-ended questions—which was disseminated online at the beginning of 2025 and through which we collected participants' opinions and perceptions (and contextually their socio—demographic data) on slur reclamation. More specifically, the survey was also designed to involve both perspectives from LGBT+ and non-LGBT+ people, so as to explore how the phenomenon is recognised, perceived, interpreted, legitimated, reproduced, and delimited across different social positions.

In particular, we present the methodological steps through which we approached this case study: the lesson learnt from previous work, the preliminary focus group, the design and implementation of the sociolinguistic survey, the results obtained, the analyses carried out, and the conclusions drawn from them.

7.1 Previous Work: the ReCLAIM Project

As anticipated in Section 3.2.3, Cuccarini et al. (2024) present the *ReCLAIM Project*, which, to the best of our knowledge, constitutes the first computational investigation of reclamatory slur use in Italian. We discuss it here in greater detail because it offers a useful entry point into the classificatory challenges posed by slur usage in the Italian LGBTQ+ context, and because it also serves as an important point of departure for the present study

By leveraging a subset of 1,742 Italian tweets extracted from Nozza et al. (2023)'s HODI dataset—originally composed of 6,000 tweets collected through 21 keywords related to LGBT+ discourse—Cuccarini et al. (2024) used the HurtLex multilingual lexicon (Bassignana, Basile, and Patti, 2018) to identify specific derogatory terms. They ultimately selected 17 homotransphobic slurs—such as *frocio*, *checca*, *ricchione*, and *finocchio*[1]—and intentionally excluded terms perceived as more neutral like *gay* or *omosessuale* to focus strictly on expressions capable of undergoing semantic reappropriation. The research highlights a critical tension in NLP: the necessity of distinguishing between abusive language and reclaimed, identity-affirming discourse to prevent the systemic censorship of marginalised voices.

Through a manual annotation process involving three expert researchers, the study identified different reclamatory contexts, from informal friendly uses to more explicitly political or artistic forms of reclamation. At the same time, the task proved markedly subjective, as shown by the moderate IAA (Fleiss' $\kappa = 0.57$). Only 168 instances were ultimately labelled as *reappropriative*, a result that also points to the difficulty of dealing with Italian-specific ambiguities, including idiomatic expressions and culturally situated echoic uses. A particularly relevant case is represented by the word *frocia*,[2] which accounted for a substantial share of disagreement and appears to have been especially difficult to annotate because, unlike the other target words, it is described by the authors as originating in an already reclamatory context. Drawing on Nossem (2019), they relate it either to a calque of the English *queer* or to a specifically Italian, territorialised post-queer concept (Cuccarini et al., 2024). The annotation difficulty becomes more intelligible if one also considers the broader lexicological discussion introduced in Section 3.2.2: *frocia* is not simply treated as a feminine variant of *frocio*, but as a form whose semantic and pragmatic profile has partly developed through reclamatory use. In this

[1]These are all derogatory terms traditionally used to refer to gay men. They can be treated, with some approximation, as slurs whose neutral counterparts are *gay* or *omosessuale* (eng. 'homosexual'), although they differ mainly diachronically and diatopically across Italy. For a broader historical and linguistic discussion of these terms, see Pepponi (2024). We will return to them later in the thesis, especially in Section 7.2.2, when describing the input sentences selected for the questionnaire.

[2]The feminine form of *frocio*. It may refer to a feminine addressee or referent, or appear as a feminine agreement form associated with a feminine noun. A rough English equivalent would be *fag*, but also *dyke*, to the extent that, as we will see, it is attested as reclaimed by women as well. However, any translation can only be approximate.

sense, it may function not only as a label, but also as an identity flag or a stance-bearing form, and its interpretation may shift depending on whether it is used substantivally or adjectivally. These results show that, even within a corpus originally developed for homotransphobia detection, the identification of reclamatory uses requires a much finer understanding of context and speaker positioning.

Moreover, in the subsequent computational phase of the study, the authors tested *Qwen*, a large language model, in a zero-shot setting, that is, without task-specific training examples. The experiments showed that performance depends heavily on the precision of the prompt. The most detailed prompt reached an accuracy of 0.82—that is, it produced the correct label in 82% of the cases—and a weighted F1-score of 0.79, a metric that takes into account both the system's ability to identify the relevant cases and its tendency to misclassify them across categories. By contrast, more general prompts were less effective in handling the conceptual ambiguity of "semantic reappropriation".

These preliminary findings reveal that existing homotransphobic datasets often lack the socio-demographic context—such as the speaker's in-group membership—required for accurate detection. In this respect, the ReCLAIM Project provides the immediate point of departure for our own case study, which extends this line of inquiry upstream by incorporating participatory sociolinguistic methods within the FATA paradigm. At the same time, the focus-group stage described below was also intended to check, in a preliminary way, for some of the issues that the ReCLAIM Project leaves partially open, such as the perceived legitimacy of reclamatory uses, the acceptability of different terms, and the possible cultural specificity of reclamation in Italian.

7.2 Methodological Description

By activating metalinguistic awareness and foregrounding community perspectives, the sociolinguistic survey was meant to be a tool able to guide subsequent annotation and corpus creation stages. It also reinforces the importance of positionality, making researchers more conscious of their interpretive biases (Yip, 2024). The structure of the questionnaire that we disseminated partially mirrors a typical annotation schema used in NLP, allowing the survey to serve as a prototype for future reclamation-sensitive language annotation

and modelling. More specifically, it was conceived as a way of making explicit some of the interpretive knowledge that remains difficult to capture in standard binary annotation settings, especially when the phenomenon under investigation is socially negotiated and context-dependent.

The presentation of the survey here has a two-fold objective. On the one hand, as mentioned above we implemented it as a technique that fits into our paradigm and whose results can inform our computational study and future NLP applications. On the other hand, our survey is an explorative tool whose analysis can help us define methodological strengths and limitations of the survey itself, with the aim of proposing it as a viable way to further investigate the phenomenon, not necessarily within the NLP field.

7.2.1 The Preliminary Focus group

Prior to deployment, the survey was refined through a pilot focus group. This preliminary stage helped tailor the examples and structure of the questions to ensure cultural and contextual relevance, and to address potential ambiguities. Inclusivity and sensitivity were prioritised throughout—including in the design of the subsequent questionnaire, which allowed for nuanced identity self-descriptions and the option to abstain from certain questions.

Our focus group was conducted by a facilitator and two observers with four LGBTQ+ participants and helped us identify salient themes and select tweet examples for further evaluation. It was organised as a semi-structured group interview. An informal setting was created: participants sat around a table together with the facilitator (the MR), while the two observers were positioned on a sofa in the same room and took notes. The meeting was audio-recorded, and the participants were previously informed about the aim and the academic context of the focus group and the role of the observers. Before starting, participants were asked to fill in a brief socio-demographic questionnaire and then sign an informed consent form. At the beginning of the session, participants were also invited to introduce themselves by stating their names and pronouns, which were then written on name tags pinned to their clothing. Then, the facilitator introduced the purpose of the meeting and recalled the freedom to intervene in the discussion and the need to pay attention to the speech turns and to respect other participants' opinions and feelings. The interview lasted 1 hour and 30 minutes.

The focus group was organised into two main phases, each based on different stimuli designed to elicit participants' reflections on slur reclamation.

The first phase began with a printed image (see Figure 7.1) of a still frame taken from the Netflix animated comedy-drama series *Strappare lungo i bordi* (eng. "Tear Along the Dotted Line") by Italian cartoonist Zerocalcare, representing this writing: "Amare le femmine è da froci"[3]. The image was presented and asked about, with a follow-up explanation of its context, after which participants were invited to offer free comments. The facilitator then asked them about the emotional impact of the image and how they interpreted both the visual stimulus and the sentence it contained. This opening stimulus was followed by a transition question aimed at broadening the discussion and encouraging group involvement through the sharing of personal experiences, before moving to a key question that shifted the debate from individual experience to a broader reflection on the contemporary uses and meanings of reclaimed slurs.

FIGURE 7.1: Still frame taken from the Netflix animated comedy-drama series *Strappare lungo i bordi* by Zerocalcare, used as a visual stimulus during the focus group.

In the second phase, participants were asked to draw slips of paper from mugs placed on the table, each labelled with a specific slur (e.g. *frocio*, *checca*, etc.). Each slip contained a sentence, selected from the **ReCLAIM reappropriation subset**, featuring the corresponding slur. Participants took turns commenting on the sentence and were asked to try to contextualise it, to discuss whether the use of the slur could be interpreted as reclamatory or not, and to say whether it was a term they personally used.

[3]Our translation: Loving women is for faggots.

The discussion concluded with a final question intended to summarise the main points that had emerged and to prompt meta-reflections on the debate itself. In this closing phase, the observers could also intervene, focusing not on the content of participants' responses but on aspects related to the flow of the discussion and group interaction.

Finally, participants were asked whether they could suggest possible sources for collecting data on linguistic reclamation usage; after that, the session ended with acknowledgements and thanks.

Although the focus group was preliminary in scope, it proved useful in bringing out a set of themes that would later inform the questionnaire more explicitly. In particular, the discussion repeatedly returned to questions of legitimacy—namely who may use reclaimed slurs, with whom, and in which contexts—as well as to the acceptability of different labels, the coexistence of different social and identity-based perspectives, and the possibility of understanding reclamation as an ongoing rather than completed process. The discussion also suggested that the Italian case should be approached with particular caution from a cultural point of view. Some participants, for example, drew a distinction between *queer* and *frocio/frocia*, describing the former as more publicly presentable or broadly circulating, and the latter as more restricted to community-internal, activist, or otherwise highly contextual uses. As one participant observed,

> (*queer*) si può usare anche in maniera presentabile, no? mentre *frocio/frocia* no [...] quando ti riappropri del termine *frocio/frocia* lo puoi usare all'interno della comunità o della tua ristretta comunità e la tua cerchia ristretta mentre *queer* no, vai là fuori e dici *queer*[4]

A second participant similarly remarked:

> Le sento diversissime *queer* e *frocio* [...] *queer* io lo utilizzo molto quando descrivo delle situazioni, dei posti [...] cioè lo riesco a utilizzare proprio in modo molto dolce come parola. Cioè per me

[4]Our translation: (*queer*) can also be used in a presentable way, can't it? whereas *frocio/frocia* cannot [...] when you reclaim the term *frocio/frocia*, you can use it within the community or within your close-knit community and inner circle, whereas with *queer* you go out there and say *queer*.

> è un'accezione molto includente. È che è un termine un po' di transizione secondo me. [5]

The focus group also suggested that reclamation may develop unevenly across terms and contexts, and that some usages are still perceived as emergent rather than fully stabilised. In this respect, one participant explicitly framed the process as partial and still evolving, noting that:

> Penso sia a centri sociali sia a contesti di attivismo: ci si è appropriati della parola *frocia*, però al femminile. Anche quando si parla di maschi gay.[6]

At the same time, these reflections were not homogeneous, which further confirmed the need to avoid treating reclamation as a fixed or uniformly shared category.

Therefore, the focus group suggests that participants recognised slur reclamation as an existing phenomenon within the LGBTQ+ community, while also indicating that its boundaries are negotiated rather than fixed. Their reflections highlighted the extent to which context, speaker identity, communicative intent, and relational proximity shape the interpretation of potentially reclaimed uses, thereby calling into question overly rigid binary distinctions. These preliminary observations are not presented here as generalisable findings. However, they served to refine the subsequent questionnaire and to foreground, from the outset, the cultural specificity, the ingroup/outgroup dynamics, the internal variability, and the positioning complexity of reclamatory practices in Italian.

7.2.2 The Input Sentences

Before designing our survey, we selected ten input sentences from the **RECLAIM dataset**—five labelled as reclamatory (*R*) and five as non-reclamatory (*Not-R*)—which were used as controlled stimuli for the questionnaire. Some of them—or closely related formulations built around the same lexical items—

[5] Our translation: I feel that *queer* and *frocio* are completely different [...] I use *queer* a lot when I describe situations or places [...] I can really use it in a very sweet way as a word. For me, it has a very inclusive meaning. I think it is also somewhat of a transitional term.

[6] Our translation: I think both in social centres and in activist contexts there has been an appropriation of the word *frocia*, but in the feminine form. Even when referring to gay men.

had already proved useful in the pilot focus group, especially as they triggered debate where legitimacy, community membership, contextual appropriateness, and cultural specificity emerged as particularly salient.

From an analytical point of view, these examples are also useful because they are not homogeneous: alongside relatively clear abusive or self-reclamatory uses, they include meta-discursive statements and reported speech. In this respect, the deflationary perspective discussed by Bianchi (2015) is especially revealing, since reported speech would straightforwardly count as unacceptable: that perspective is associated with a silentist stance and tends to treat even mention and reported uses of slurs as infractions, to the point of regarding utterances such as "*Frocio* è una parola offensiva" (eng. "*Frocio* is an offensive word") or "Nessuno dovrebbe usare la parola *frocio*" (eng. "No one should use the word *frocio*") as problematic. For the purposes of the present study, however, we do not treat reported speech, in itself, as a sufficient criterion either for classifying an utterance as offensive or for determining whether it qualifies as reclamatory. Both judgements require further contextual and pragmatic cues.

Some input sentences also include cases in which the target item appears in adjectival position, as the word *frocia*, which, as discussed above, is particularly sensitive in the Italian context. More specifically, the cases involving *frocia* are relevant because they must be interpreted in relation to both morphological resources and context-dependent pragmatic uses. This is one of the reasons why the binary annotation *R/Not-R*, inherited from Cuccarini et al. (2024), should here be understood as the result of an implicit interpretive effort of annotators, in which both context and textual configuration play a role. A more situated annotation should aim to make this interpretive work more explicit, shifting our focus toward a more pragmatic perspective. In other words, the discussion that follows does not treat the *R/Not-R* label as self-sufficient, but reads each sentence through the interaction between lexical choice, morphosyntactic configuration, speaker positioning, and discursive framing.

The input sentences presented below are accompanied by a brief qualitative commentary. This commentary is itself the outcome of an interpretative effort and, as such, cannot be detached from the positionality of the researchers. In each sentence, one slur is highlighted in bold. As will become clear later,

these are the items that participants were asked to evaluate in the second part of the questionnaire. They are left untranslated, as are other slurs referring to the LGBTQ+ community, since any adequate translation would need to account for the socio-cultural specificity of each form, which lies beyond the scope of the present thesis.

(1) #1 [*Not-R*] per riproclamarlo E BASTA. Se lo usano altri è solo continua oppressione. Per di più Harry non è un vostro amico, e non vi dovete permettere di dire certe cose. Se una mia amica mi chiama **frocia**/troia non me la prendo perché hanno il permesso di farlo ma lui non lo conoscete.[7]

This sentence is especially interesting because it is largely meta-discursive. Its reclamatory interpretation is not straightforward as it does not directly perform a reclaimed use. It invokes a possible community use while explicitly arguing about who may use the term and under which relational conditions. The only linguistic cue is the use of a self-reference, which can help us understand the writer as a member of the LGBTQ+ community. However, this input anticipates one of the main themes that already emerged in the focus group and, later, in the analysis of the questionnaire, regarding legitimacy as a matter of group belonging and permission.

(2) #2 [*R*] Le mie coinquiline mi hanno chiesto di parlare col padrone di casa il quale non sospetta minimamente che io sia **ricchione** sarà divertente spiegargli la rainbow flag in camera[8]

We consider this sentence one of the more straightforward reclamatory examples in the list, since the slur is used in the first person and is embedded in a narrative of self-disclosure rather than in an attack on someone else. The lightly playful framing of the utterance, together with the mention of the rainbow flag, further supports a non-abusive reading.

[7] Our translation: To reclaim it and THAT'S IT. If others use it, it's just continuous oppression. Moreover Harry is not your friend.M, and you shouldn't dare say certain things. If a friend.F of mine calls me a **frocia**/slut, I don't get offended because they have permission to do so, but you don't even know him.

[8] Our translation: My roommates asked me to talk to the landlord, who has no idea that I'm a **ricchione** it's going to be fun explaining the rainbow flag in my room.

(3) #3 [*Not-R*] Un po' come quando tra gay si chiamano "finocchio", "**frocio**" o "ricchione". Lo dicessi io sarei un omofobo[9]

This example is again meta-discursive, and there is little reason to read it as reclamatory. Similarly to input N. 1, the utterance includes a form of reported speech while explicitly foregrounding the issue of who is, or is not, entitled to use certain terms.

The sentence does not simply describe differential entitlement in the abstract: it also contains linguistic cues that suggest the writer is positioning themself outside the gay community. In particular, the use of the third person plural encourages the interpretation that those who may legitimately use the terms are being construed as a group to which the speaker does not belong. This impression is reinforced by the subsequent use of the first person where we can read a form of aggrieved self-positioning (*Lo dicessi io sarei un omofobo*). In this respect, the utterance also recalls a broader discursive pattern often found in complaints about allegedly excessive restrictions on speech, where the speaker frames themself as unfairly constrained. For all these reasons, the writer does not appear to perform a reclaimed use. More generally, this kind of case shows that the deflationary perspective is right to remind us that reported speech may still convey offensiveness. However, as discussed above, this alone is not enough: reported speech cannot be treated as a sufficient criterion in itself, since the interpretation of the utterance also depends on other contextual and pragmatic cues.

(4) #4 [*Not-R*] Bro ma ti rendi conto che così pari solo un succhiacazzi per quell'altro **frocio** [EMOTICON CHE RIDE][10]

Here the derogatory intent is much more evident. The target terms are directed at others, the mocking stance is reinforced by the laughing emoji, and there is no sign of self-reference, community use (in Bianchi, 2015's terms), or meta-discursive distancing, such as cues of reported speech. Compared with other examples in the list, this one therefore works almost as a contrastive baseline:

[9] Our translation: Kind of like when gay people call each other "*finocchio*" "**frocio**" or "*ricchione*". If I said it, I'd be a homophobe.

[10] Our translation: Bro, do you realize you just sound like a cock-sucker for that other **frocio**? [LAUGHING EMOJI]

it is precisely the kind of overtly abusive use against which reclaimed uses must be distinguished.

(5) #5 [*R*] Fra siamo letteralmente gay, posso usare le parole **frocio** finocchio e ricchione, nessuno l3 usa più come insulti, dovresti aggiornarti un attimo che siamo 2022[11]

This meta-discursive sentence is explicit both about speaker positioning and about the ideological defence of reclamation. It presents ingroup membership as sufficient grounds for legitimate use, and it frames those terms as no longer functioning as insults. At the same time, the categorical force of the statement ("nessuno l3 usa più come insulti") is analytically interesting, because it contrasts with the more cautious and internally differentiated picture that emerged from the focus group, where acceptability was repeatedly treated as context-dependent and uneven across terms.

(6) #6 [*R*] Sto tornando a casa ora dal pride + serata **frocia** [EMOTICON TESCHIO][12]

This example is particularly relevant because *frocia* occurs here in an adjectival position, modifying an event label rather than directly referring to an individual. This pattern of usage is noteworthy in light of both the ReCLAIM disagreement analysis and the focus group discussion, where expressions such as "serata frocia" were treated as highly context-dependent and community-encoded (and even defined as "sweet"). In relation to the discussion prompted by Pepponi (2024), adjectival realisations of *frocia* can be interpreted as stance-bearing or scene-framing devices. From this perspective, an expression like "serata frocia" may index an event as socially and affectively situated rather than simply describing it. The explicit reference to Pride strongly supports a reading in terms of reclamation. That is why this sentence also illustrates how such uses of *frocia* become difficult to classify without a strong pragmatic effort.

[11] Our translation: Bro, we are literally gay, I can use the words **frocio** *finocchio* and *ricchione*, no one uses them as insults anymore, you should update yourself that's 2022.

[12] Our translation: I'm heading home now from Pride + **frocia** night [SKULL EMOJI]

(7) #7 [*R*] Allora posso dirti per la parte non cishet e io ho trovato un lavoro, anzi me ne sono stati offerti anche Coi capelli viola corti e visibilmente **frocia**. Come al solito hanno fatto speculazioni alle mie spalle ma il lavoro me lo hanno dato. Però conta che vivo a milano e+[13]

This is another case in which *frocia* functions adjectivally, but here as part of a self-descriptive reference. The sentence is reclamatory insofar as the speaker appears to adopt the term in first person, yet it does so in a context that simultaneously alludes to social judgement and speculation by others. Moreover, this seems to be a case where *frocia* is used for a feminine referent; however, as also emerged in the focus group discussion, one possible reading is that the feminine term may be self-referred by a (gay) man, thus involving a double reclamatory movement, lexical and morphological at once, especially if the speaker is positioning themself against normative expectations of gender and sexuality.[14] However, from a discursive point of view, the coexistence of self-identification and anticipated stigma in the sentence is especially important, because it shows that reclaimed uses do not necessarily occur in settings of self-affirmation alone, but may remain entangled with hostile social perceptions.

(8) #8 [*Not-R*] Sei solo un invertito che non vuole ammettere che certi comportamenti sono schifosi. Vuoi fare il **ricchione**? Bene, mettiti il guanto.[15]

This sentence is overtly abusive and leaves little room for a reclaimed reading. The slurs are used in direct address to a specific person and are accompanied by the term "invertito" and by an explicit moral condemnation ("certi comportamenti sono schifosi"). The utterance is structured as a hostile provocation,

[13]Our translation: So, I can tell you for the non-cishet part, and I found a job, actually I even got multiple offers. With short purple hair and visibly **frocia**. As usual people speculated behind my back, but I still got the job. But keep in mind, I live in Milan and+

[14]At the lexical level, this may plausibly be read as a reclamatory use of the slur. The status of the feminine morphological marking is less straightforward, and we do not develop that issue further here, as it is too broad to be addressed adequately within the scope of the present thesis. It is nevertheless worth flagging, since it may open an interesting line of inquiry for future research.

[15]Our translation: You're just an invert who doesn't want to admit that certain behaviors are disgusting. You want to act as a **ricchione**? Fine, wear the glove.

and the rhetorical question "Vuoi fare il *ricchione*?" intensifies this effect further, since the use of "fare" frames the slur not only as an insult, but also as a mockery of a supposedly adopted or performed way of being. In analytical terms, it represents an unambiguous *Not-R* case.

(9) #9 [*Not-R*] RT Mi basta ascoltare 10 secondi DRAGHI per realizzare quanto Conte è una nullità (e **checca** rancorosa).[16]

Although formally this text seems to be framed as a retweet ("RT"), the evaluative direction of the utterance remains derogatory. This interpretation depends on the absence of any distancing marker: the retweet frame does not neutralise the offensive stance, since the slur is reproduced as part of the speaker's own evaluative alignment rather than as an object of critical quotation. The slur is used metaphorically to feminise and belittle a political figure, and there is no indication of self-reference or other community-internal reclamation cues. This example is useful because it shows again that not all non-literal or mediated uses are ambiguous in the same way: reported or relayed discourse may still preserve a clearly offensive stance. More broadly, it also reminds us that the presence of quotation, circulation, or textual embedding does not in itself neutralise harmful meaning.

(10) #10 [*R*] Ama io parlo come mi pare essere **checca** è uno spasso giuro[17]

This sentence is comparatively short, but several cues favour a reclaimed interpretation: the intimate address form "Ama", the first-person stance, and the explicitly positive evaluation ("è uno spasso"). At the same time, as with other first-person examples, its interpretation still presupposes some inference about speaker positioning and intended audience. This is precisely the kind of compact, context-poor social media utterance for which automatic systems may struggle unless they are trained to handle pragmatic cues beyond the lexical presence of the slur.

[16]Our translation: RT: Just listening to 10 seconds of DRAGHI is enough for me to realize how useless Conte is (and a bitter **checca**).

[17]Our translation: Babe I talk however I want being a **checca** is a blast I swear.

7.2.3 The Sociolinguistic Survey

The sociolinguistic survey aimed to capture both attitudes and personal experiences related to the reclamatory use of slurs, with a particular focus on the LGBTQ+ community and its broader social environment. We conducted the survey using a self-administered web-based questionnaire. The survey was disseminated by members of the research team primarily through their individual personal and professional networks (via email, instant messaging applications, and word of mouth), and to a lesser extent through academic mailing lists and activist WhatsApp and Telegram groups. The dissemination phase lasted approximately two weeks, between January and February 2025, and reached 282 individuals.

7.2.3.1 The Google Form Questionnaire

The questionnaire was presented in the form of a Google Form titled *Linguaggio e Riappropriazione* (see Appendix E). Immediately below the title, the form description displayed an introductory text providing ethical information and details about the study. This embedded description informed prospective participants that the survey was fully anonymous and had received ethical approval from the Bioethics Committee of the University of Turin. The study was conducted within the framework of the project *FAVOLOSA—Fair Automatisation and Visibility Of the LGBT+ Community On Slur (Re)Appropriation*, and participants were informed that data collection, processing, and storage complied with the European Union General Data Protection Regulation (EU GDPR; Regulation (EU) 2016/679).

The same introductory description included a content warning indicating that the questionnaire contained sentences extracted from social media contexts, which might convey offensive or potentially distressing language. Participants were informed that there were no right or wrong answers and that the sentences were intentionally presented without their original context. Respondents were encouraged to base their responses exclusively on their own perceptions and personal experiences. Contact information of the MR was provided, and an estimated completion time of the questionnaire of approximately 25 minutes was indicated.

After reading this introductory description, participants were presented with the first section of the form, titled *Consenso* (*"Consent"*). This mandatory section operationalised informed consent by requiring respondents to select one of two mutually exclusive options. The first option stated that the participant agreed to take part in the study and confirmed that they were over 18 years of age and that Italian was their native language (or one of their native languages). Selecting this option was necessary in order to proceed with the questionnaire and constituted explicit informed consent, as well as confirmation of the inclusion criteria. The second option indicated that the individual did not wish or could not participate in the survey. Selecting this option automatically submitted the form and immediately terminated participation, thereby ensuring that no data were collected from individuals who did not provide consent or did not meet the eligibility requirements.

The questionnaire was structured into 4 parts, respectively composed of:

- sentence-level linguistic attribution tasks—based on 10 input sentences—investigating perceived offensiveness, hatefulness, and communicative context;
- word-level assessments on the same 10 input sentences, focusing on perceived reclamatory intent and personal habit of using specific slurs;
- questions on exposure to, legitimacy of, and personal engagement with reclamatory language, and then general feedback and other examples of reclamation;
- socio-demographic profiling, including self-identification variables such as gender identity, sexual orientation, and self-reported closeness to the LGBT+ community.

In the questionnaire, the ten sentences were evenly distributed between *R* and *Not-R* and presented in random order to prevent pattern recognition. The sentences appeared in both Part 1 and Part 2; however, the target slur was highlighted in bold only in Part 2, where reclamation was explicitly addressed.

7.2.3.1.1 Part 1. Sentence-level Linguistic Attribution Task In the first part, for each input sentence 4 mandatory closed-ended questions are proposed, including 2 multiple-choice questions (1 with the option to add more information) and 2 scalar evaluations (from 1 to 5).

This section aims to understand the communicative context and the perceived intention behind each input sentence. The questions explore how readers interpret the whole sentence and the speaker's attitude.

Item 1A (Context Attribution): ***"In what communicative context might this sentence appear?"*** This question asks respondents to imagine where or in what social situation the sentence might be used. The options (*friendly, family, work, activist* etc.) were meant to help us understand how the perceived meaning of a sentence could shift depending on the relationship between the speaker and the audience. The underlying idea was that the same words may sound more or less offensive depending on the context.

Item 1B (Membership Attribution): ***"Does the writer of the sentence belong to the LGBT+ community?"*** This question explores how the perceived speaker's identity influences the interpretation of their words. Our hypothesis was that many terms that are historically offensive may take on different meanings when used by someone belonging to the marginalised group being referenced.

Item 1C (Offensiveness): ***"Does the writer have offensive intent? (Rate from 1 to 5)"*** This question aims at capturing how much intentional offensiveness the respondent perceives. It is not about objective truth but about the reader's interpretation of why the sentence was written. The Likert scale from 1 (*not offensive at all*) to 5 (*definitely offensive*) was conceived to allow the study to measure subtle differences in perceived aggressiveness or hostility.

Item 1D (Hatefulness): ***"Does the sentence convey hate? (Rate from 1 to 5)"*** This question—using a Likert scale as well—helps understanding whether the sentence could be considered Hate Speech and to which degree. It is distinguished from the previous item as in our questionnaire we discern offensiveness from hatefulness, following the idea that a sentence may be rude without expressing real hatred, or it may be hate-driven without sounding explicitly insulting. However, we need to clarify here that, as we are not sure that respondents were familiar or shared with us the same conceptualisation of hate speech, we could not consider their intention as a detection of HS.

7.2.3.1.2 Part 2. Word-level Reclamation Assessment Task At the beginning of this section, the following text is presented to give respondents a simplified definition of 'reclamation':

> Ti chiediamo di rispondere alle prossime domande considerando che con **riappropriazione** (e **riappropriativo**) facciamo riferimento a un fenomeno linguistico per il quale *le persone rivendicano l'utilizzo di termini dispregiativi per esprimere la propria identità e/o l'appartenenza ad una certa comunità.*[18]

For each input sentence, 2 mandatory closed-ended questions and 1 optional open-ended question related to the first question are proposed.

This section aims at investigating the perception and opinions of respondents regarding specific cases of linguistic reclamation.

Item 2A (Reclamation Assessment): ***"In this case, has the writer reclaimed the highlighted word?"*** This question asks whether the highlighted term is being used in a reclaimed, non-offensive sense. The options wanted to capture the respondent's interpretation through a *yes*, *no*, or *not sure* answer. Moreover, this question was supposed to give us idea of whether and to which extent the 10 input sentences were perceived as reclamatory from the participants to the study as we did when selecting them from Cuccarini et al. (2024).

Item 2B (Reclamation Explanation): ***"Please briefly explain why:"*** This open-ended question invites respondents to briefly justify their previous answer. It was meant to help us understand which clues (e.g. tone, context, the speaker's identity, the emotional content...) may lead to perceiving a term as reclaimed or not.

Item 2C (Self-reported Similar Use): ***"Do you happen or have you ever happened to use the highlighted word in a similar way?"*** This question investigates the respondent's personal experience on the specific slurs, i.e. whether they themselves have used derogatory terms in a reclamatory manner.

[18] We ask you to answer the following questions considering that **reclamation** (and **reclamatory**) refers to a linguistic phenomenon in which *people reclaim the use of derogatory terms to express their identity and/or belonging to a certain community.*

In our mind, it could help start distinguishing between theoretical understanding and actual linguistic behaviour.

7.2.3.1.3 Part 3. Respondents' General Opinions and Insights This section investigates general opinions and insights about reclamatory language, moving beyond the evaluation of specific sentences.

Item 3A (Perceived Exposure): ***"In your daily experience, do you happen to hear reclamatory language uses similar to the examples above?"*** This question aimed at measuring how present or familiar reclamatory language is in the respondent's everyday life. It helped compare subjective experience with broader sociolinguistic trends.

Item 3B (Legitimacy): ***"You consider the use of such words legitimate in a reclamatory way by:"*** This question asks who is considered entitled to use reclaimed expressions. The options include *LGBT+ people, activists, friends, family, public figures, artists, etc.*, as well as an *Other*___ field for more nuanced views. It was thought to shed light on social norms and boundaries regarding who can reclaim stigmatised language.

Item 3C (Use): ***"Do you happen to use words or expressions with a reclamatory intent?"*** This question explores the general respondent's own linguistic practices, not just their opinions. It helped assess whether reclamation is something they personally engage in.

Item 3D (Additional Reclaimed Words): ***"Besides those mentioned, do you know other words used in a reclamatory way?"*** This optional question invites respondents to list additional examples of reclaimed words. It aimed at enriching the data and broadens the range of terms recognised as reclaimed in contemporary usage.

Item 3E (Final Comments): ***"If you wish, briefly comment on your answers:"*** This optional open-ended question allows participants to add reflections, explanations, or clarifications. It was thought to provide valuable qualitative insights to complement the structured data collected earlier.

7.2.3.1.4 Part 4. Socio-Demographic Profiling The final part of the questionnaire collects socio-demographic information and self-identification variables, including age group, gender identity, sexual and/or romantic orientation, educational background, geographical context, and self-reported closeness to the LGBT+ community.

Importantly, this section was deliberately positioned at the end of the questionnaire in order to minimise potential priming and order effects. Survey methodology literature suggests that placing sensitive or identity-related questions at the conclusion of a survey helps reduce the risk of biasing earlier responses and improves participant comfort. By allowing respondents to first engage with the linguistic judgement and opinion tasks without foregrounding their own demographic categorisation, we aimed to preserve greater neutrality in Parts 1–3.

Age was collected through predefined brackets commonly used in demographic research. Geographical background was operationalised through both regional location (Italian regions or abroad) and municipality size, following ISTAT population-based categories. A distinction between geographical background and actual geographical location of residence was made (before and after being 18 years old), resulting in two series of questions (region and municipality size) in order to pave the way to possible longitudinal trends in every person biography.

Gender identity and sexual/romantic orientation was collected through multiple-choice items including both an '*I prefer not to answer* option and an open-ended *'Other: ___'* field. This design choice was motivated by principles of inclusive demographic practice, allowing respondents to self-identify using labels that best reflect their lived experience rather than forcing alignment with predefined categories only. However, these variables were treated analytically as distinct from the final item—measuring self-reported closeness to the LGBT+ community—in order to distinguish personal identity with community participation or affective proximity.

Closeness to the LGBT+ community was assessed through a five-point Likert scale ranging from *weak* to *strong* connection. This item was intended to capture respondents' subjective sense of affiliation or engagement independently of their identity-based self-definition. While this measure provides a useful exploratory indicator, its limitations—particularly with respect to

social desirability bias and conceptual granularity—are addressed in Section 7.3.2.

Overall, the socio-demographic section was designed to balance openness and analytical utility, while maintaining a clear separation between identity variables and attitudinal or experiential measures related to the LGBT+ community.

7.2.4 Data Organisation, Results and Discussion

A total of 279 responses were collected, representing a diverse sample across age groups, gender identities, sexual orientations, and regional backgrounds. At the same time, as will become clear from the socio-demographic profile presented below, the sample should not be understood as statistically representative of the Italian population. Rather, it constitutes an exploratory dataset that is particularly useful for investigating patterns of perception and self-reported practice in relation to reclamatory language. In line with the overall design of the study, the presentation of the results is therefore guided not by an exhaustive descriptive aim, but by the attempt to foreground those dimensions that are most relevant to the research questions and to the interpretive issues highlighted in the previous sections.

As the questionnaire was closed, we created a Google Sheets file in which all responses were collected, cleaned, and organised for analysis. Respondents were then clustered into groups on the basis of their self-reported gender identity and sexual orientation. This grouping procedure was adopted for analytical purposes, in order to compare broad social positions that are especially relevant to the phenomenon under investigation. Results—organised by questionnaire part—are therefore presented here in an order that differs from the original structure of the questionnaire. This reorganisation reflects an analytical rather than chronological rationale: it first establishes the respondent groups used throughout the analysis, and then moves from the assessment of reclamation itself to the interpretation of contextual and metalinguistic dimensions. The results are therefore discussed in the following order:

- Part 4. Socio-demographic Profiling (Paragraph 7.2.3.1.4), as our first step was to cluster respondents into social groups based on self-reported

identity (*ingroup*, *outgroup*, *other group*). In this way, for each item, we observe from time to time the results for each group;

- Part 2. Word-level Reclamation Assessment Task (Paragraph 7.2.3.1.2), as it contains interpretations about reclamation for each input sentence, and allows us to consider participants' assessment of the presence of reclamation;
- Part 1. Sentence-level Linguistic Attribution Task (Paragraph 7.2.3.1.1), because once participants' judgements about reclamation have been presented, it becomes possible to examine how they attributed contextual and annotation-like features to the same inputs, such as communicative setting, speaker membership, offensiveness, and hateful intent;
- Part 3. Respondents' General Opinions and Insights (Paragraph 7.2.3.1.3), which provides a broader and more reflective overview of the phenomenon. This final part is discussed after the more sentence-bound tasks because it brings together meta-linguistic and meta-methodological considerations that help interpret the survey as a whole and, at the same time, provide useful material for future research.

Given the number of items and the richness of both closed and open-ended responses, it would be neither practical nor analytically useful to comment in equal detail on every single result. Instead, we focus on those issues that are most relevant to the aims of the study: the **perception of reclamation**, the **role of group belonging**, the **importance of context and speaker positioning**, and the extent to which **participants' responses confirm, complicate, or challenge the assumptions** inherited from previous work. When relevant, statistics are combined with selected qualitative comments, in order to preserve both the comparative and the interpretive value of the dataset.

For each part we describe the results for each item and at the end of the subsection a brief discussion summarises the main tendencies that emerge, relates them to the literature and to the preliminary focus-group insights, and highlights any methodological cautions that should be kept in mind when interpreting the findings.

7.2.4.1 Part 4. Socio-Demographic Profiling (Results)

7.2.4.1.1 Belonging to the LGBT+ community We defined ***ingroups*** as *LGBT+* individuals, including all participants who, in at least one of the two dimensions considered (gender identity and sexual orientation), selected a category we classified as belonging to the LGBT+ community. Overall, this group comprised 132 participants (47.3%) out of 279. Among them, 35 identified as cisgender men (26.5%), 73 as cisgender women (55.3%), and 24 as individuals with other gender identities or who did not specify a gender identity (18.1%). Within the ingroup, 33 participants self-identified as gay (25%) and 22 as lesbian (16.7%), while the remaining 77 reported other sexual orientations (58.3%), most of which fell under the umbrella term "bisexual" In addition, 15 participants explicitly defined their gender identity outside the man/woman binary.

The ***outgroups*** consisted of participants who identified themselves as both cisgender and heterosexual and were therefore classified as *non-LGBT+* individuals. This group included 136 respondents (48.8% of the total sample), of whom 101 were women (74.3%) and 35 were men (25.7%).

A residual category labelled ***other group*** included 11 participants (3.9% of the total sample) who either did not self-identify along either dimension, identified only on one of the two dimensions (identity or orientation), or provided responses such as *"libera"* ("free"), *"innamorata"* ("in love"), or *"persona"* ("person"). As these responses in our perspective did not allow for a clear classification, these participants were not assigned to either the ingroup or the outgroup. This residual category was retained in the dataset for completeness, but, given both its small size and internal heterogeneity, it is not treated as analytically comparable to the two main groups in the discussion of group-based patterns.

7.2.4.1.2 Age With respect to socio-demographic characteristics, the largest age group in the sample was 25–34 years (45.5%), followed by participants aged 35–44 (22.6%). The remaining respondents were relatively evenly distributed across other age ranges: 7.5% were 18–24 years old, 9% were between 45 and 54, and 9.3% were between 55 and 64. The smallest proportion of participants (6.1%) was aged 65 or older.

7.2.4.1.3 Region of residence before the age of 18 With regard to the region in which participants lived for most of the time before the age of 18, the majority reported having grown up in Northern Italy (55.6%). Southern Italy and the islands accounted for a substantial proportion of the sample (35.1%), whereas a smaller share of respondents reported having lived in Central Italy during their formative years (9.0%). Only one participant (0.4%) reported having lived abroad for most of the time before turning 18. At the regional level, Piedmont was by far the most represented region (35.1%), followed by Campania (10.4%), Lombardy (9.3%), Sicily (7.5%), and Puglia (7.2%). All remaining regions were represented by smaller proportions of participants.

7.2.4.1.4 Municipality size before the age of 18 With regard to the size of the municipality in which participants lived for most of the time before the age of 18, the largest proportion reported having grown up in large municipalities with more than 100,000 inhabitants (35.5%). Medium-sized municipalities were also well represented, with 26.9% of respondents reporting residence in municipalities with between 20,000 and 100,000 inhabitants, and 24.4% in municipalities with populations between 2,000 and 20,000. Smaller municipalities were less frequently reported: 12.9% of participants grew up in municipalities with between 500 and 2,000 inhabitants, and only 0.4% reported having lived in municipalities with fewer than 500 residents during their formative years.

7.2.4.1.5 Current region of residence With respect to current region of residence, most participants reported living in Northern Italy (64.5%). Southern Italy and the islands accounted for 21.1% of the sample, while 9.3% of respondents reported residing in Central Italy. A non-negligible proportion of participants reported living abroad at the time of data collection (5.0%). At the regional level, nearly half of the sample reported residing in Piedmont (47.0%), followed by Campania (6.5%), Lombardy (6.1%), and Puglia (5.7%). All other regions were represented by smaller proportions of respondents.

7.2.4.1.6 Current municipality size With regard to the size of the municipality in which participants currently live in the most stable manner, the majority reported residing in large municipalities with more than 100,000 inhabitants (56.0%). Medium-sized municipalities with between 20,000 and

100,000 residents accounted for 23.5% of the sample, while 14.1% reported living in municipalities with populations between 2,000 and 20,000 inhabitants. Smaller municipalities were less frequently represented: 6.1% of participants reported living in municipalities with between 500 and 2,000 inhabitants, and only 0.4% reported residing in municipalities with fewer than 500 residents.

Due to a technical error in the survey design, this question was not mandatory, resulting in two missing responses (N = 277). Both missing cases corresponded to participants who grew up in Campania in municipalities with between 500 and 2,000 inhabitants and who currently reside in Campania.

7.2.4.1.7 Educational level With regard to educational background, the sample was characterised by a relatively high level of formal education. The largest proportion of participants reported holding a master's degree or equivalent (35.8%), followed by those with a high school diploma (18.6%) and those holding a bachelor's degree or equivalent (17.9%). In addition, a substantial share of respondents reported having completed postgraduate education, including a doctoral degree (14.3%) or a postgraduate master programme (12.9%). Only one participant (0.4%) reported having completed education up to lower secondary school.

7.2.4.1.8 Closeness to the LGBT+ community Self-reported closeness to the LGBT+ community was assessed using a 5-point Likert scale ranging from 1 (*not close at all*) to 5 (*very close*). Overall, responses were skewed toward the higher end of the scale, indicating a generally high level of perceived closeness within the sample. Specifically, the majority of participants selected the highest value on the scale (43.4%), followed by responses of 4 (28.3%) and 3 (21.9%). Lower ratings were comparatively infrequent, with only 3.6% of respondents selecting 2 and 2.9% selecting 1. Descriptively, the mean level of closeness was high (M = 4.06).

7.2.4.1.9 Part 4. Socio-Demographic Profiling (Discussion) Taken together, these socio-demographic data are important not only descriptively, but also methodologically.

On one side, they justify the analytical decision to foreground group belonging in the presentation of the results, since the sample is almost evenly

divided between LGBT+ and non-LGBT+ respondents, thereby making group-based comparisons meaningful within the limits of an exploratory study.

On the other side, the profile of the sample also calls for caution. Participants are disproportionately concentrated in younger and middle adult age groups, in Northern Italy—and especially in Piedmont—in urban contexts, and at relatively high levels of education. Moreover, self-reported closeness to the LGBT+ community is high overall, which suggests that even part of the non-LGBT+ subgroup may be relatively familiar with the social and linguistic issues addressed in the questionnaire.

These characteristics do not invalidate the survey, but they do shape the scope of the conclusions that can be drawn from it. More specifically, the dataset is particularly well suited for exploring how reclamatory language is interpreted within a sample that is relatively proximate—socially, geographically, or culturally—to LGBT+ contexts, whereas it is less suitable for claims about the broader Italian population. In this sense, the value of the socio-demographic profiling lies in making explicit the conditions under which the subsequent results should be read. It also suggests that any recurring differences between ingroups and outgroups should not be simplistically interpreted as the effect of absolute distance versus closeness, since the survey captures a more stratified picture in which identity, familiarity, age, and social positioning may intersect in complex ways.

7.2.4.2 Part 2. Word-level Reclamation Assessment Task (Results)

After looking at the demographic composition of our participants, we then turned to the second part of the questionnaire in order to assess whether participants' responses aligned with the preliminary binary classification of the highlighted word as reclaimed or non-reclaimed. More specifically, this part of the analysis allows us to verify to what extent the distinction inherited from the ReCLAIM Project is also recognised by our respondents, and whether this recognition remains stable across the different social groups identified above. For this initial step, we rely primarily on majority voting in the overall sample as a heuristic criterion for identifying which inputs were more likely to be perceived as reclamatory in this dataset. This criterion should not be understood as establishing a definitive status for each item; rather, it provides

a first empirical threshold that can guide the more qualitative discussion that follows.

We decided to look at the majority voting in the general population and—in the following analysis—the inputs[19] which were confirmed as non reclamatory were not taken forward for the same degree of detailed qualitative discussion in the subsequent steps. This choice was made for reasons of analytical economy and does not imply that non-reclamatory items are methodologically irrelevant. On the contrary, they remain useful as contrastive cases, especially when they help clarify which contextual or pragmatic cues participants appear to associate with non-reclaimed uses.

7.2.4.2.1 Item 2A: Reclamation Assessment Table 7.1 shows the distribution of the reclamatory assessment for each input (described in Section 7.2.2), highlighting the inputs that were overall considered reclaimed by calculating majority voting.

Overall, these results confirm the preliminary annotation of the ReCLAIM dataset in nine out of the ten cases. The only reversal concerns input #1, which had originally been classified as *Not-R* but was assessed as reclamatory by a clear majority of respondents in the present survey (70.3%). On this basis, six out of the ten inputs can be treated here as reclamatory according to participants' predominant assessments (#1, #2, #5, #6, #7, and #10), whereas the remaining four (#3, #4, #8, and #9) are predominantly non-reclamatory.

Among the items interpreted as reclamatory, the highest degree of consensus is found for input #2 (91.4% Yes), followed by input #7 (83.2% Yes). The other four reclamatory inputs show lower, but still substantial, levels of agreement: 77.4% for #10, 74.6% for #6, 72.4% for #5, and 70.3% for #1. By contrast, three of the four non-reclamatory items received particularly strong negative majorities, namely #8 (91.0% No), #3 (86.4% No), and #9 (82.1% No), while #4 also remained predominantly non-reclamatory, although with a lower level of

[19]For sake of simplicity, from now on we will generally refer to *input*, thus sometimes implicitly defining—indistinguishably—the whole sentence as reclamatory or not, even if the question proposed to participants asked about the single highlighted word. However, we take into account that the presence of a reclaimed word does not necessarily make the entire sentence reclamatory. This distinction will be then briefly discussed later, as it highlights from one side the limitations of our item but at the same time the complexity of the phenomenon. Moreover, we take into consideration the possibility that participants did not perceived such an implicit distinction, or that—in any case—they did not necessarily applied this distinction in their interpretation.

Input	Group	Yes	No	Idk
#1	*Ingroup (LGBT+)*	78.8% (104)	4.5% (6)	16.7% (22)
	Outgroup (non-LGBT+)	64.0% (87)	17.6% (24)	18.4% (25)
	Other	45.5% (5)	18.2% (2)	36.4% (4)
	Overall	**70.3% (196)**	11.5% (32)	18.3% (51)
#2	*Ingroup (LGBT+)*	96.2% (127)	0.8% (1)	3.0% (4)
	Outgroup (non-LGBT+)	88.2% (120)	3.7% (5)	8.1% (11)
	Other	72.7% (8)	9.1% (1)	18.2% (2)
	Overall	**91.4% (255)**	2.5% (7)	6.1% (17)
#3	*Ingroup (LGBT+)*	4.5% (6)	88.6% (117)	6.8% (9)
	Outgroup (non-LGBT+)	8.1% (11)	85.3% (116)	6.6% (9)
	Other	9.1% (1)	72.7% (8)	18.2% (2)
	Overall	6.5% (18)	**86.4% (241)**	7.2% (20)
#4	*Ingroup (LGBT+)*	6.1% (8)	77.3% (102)	16.7% (22)
	Outgroup (non-LGBT+)	8.8% (12)	73.5% (100)	17.6% (24)
	Other	9.1% (1)	36.4% (4)	54.5% (6)
	Overall	7.5% (21)	**73.8% (206)**	18.6% (52)
#5	*Ingroup (LGBT+)*	68.9% (91)	11.4% (15)	19.7% (26)
	Outgroup (non-LGBT+)	76.5% (104)	10.3% (14)	13.2% (18)
	Other	63.6% (7)	18.2% (2)	18.2% (2)
	Overall	**72.4% (202)**	11.1% (31)	16.5% (46)
#6	*Ingroup (LGBT+)*	78.8% (104)	6.1% (8)	15.2% (20)
	Outgroup (non-LGBT+)	72.1% (98)	10.3% (14)	17.6% (24)
	Other	54.5% (6)	9.1% (1)	36.4% (4)
	Overall	**74.6% (208)**	8.2% (23)	17.2% (48)
#7	*Ingroup (LGBT+)*	87.9% (116)	3.8% (5)	8.3% (11)
	Outgroup (non-LGBT+)	78.7% (107)	7.4% (10)	14.0% (19)
	Other	81.8% (9)	0.0% (0)	18.2% (2)
	Overall	**83.2% (232)**	5.4% (15)	11.5% (32)
#8	*Ingroup (LGBT+)*	3.0% (4)	96.2% (127)	0.8% (1)
	Outgroup (non-LGBT+)	6.6% (9)	87.5% (119)	5.9% (8)
	Other	0.0% (0)	72.7% (8)	27.3% (3)
	Overall	4.7% (13)	**91.0% (254)**	4.3% (12)
#9	*Ingroup (LGBT+)*	3.0% (4)	85.6% (113)	11.4% (15)
	Outgroup (non-LGBT+)	8.8% (12)	80.9% (110)	10.3% (14)
	Other	0.0% (0)	54.5% (6)	45.5% (5)
	Overall	5.7% (16)	**82.1% (229)**	12.2% (34)
#10	*Ingroup (LGBT+)*	78.8% (104)	6.8% (9)	14.4% (19)
	Outgroup (non-LGBT+)	77.2% (105)	6.6% (9)	16.2% (22)
	Other	63.6% (7)	9.1% (1)	27.3% (3)
	Overall	**77.4% (216)**	6.8% (19)	15.8% (44)

TABLE 7.1: Responses to Item 2A by Input and Group. For each Input, the Majority Vote for the General Population (Overall) is Highlighted in Bold.

convergence (73.8% No) and the highest proportion of *Idk* responses in the table (18.6%).

A further point concerns the distribution of disagreement across groups. In general, the non-reclamatory items appear to generate more stable patterns between ingroup and outgroup respondents, whereas some of the reclamatory items show wider divergences. Input #1 is the clearest case in this respect: although it is interpreted as reclamatory overall, the gap between LGBT+ and non-LGBT+ participants is nearly 15 percentage points (78.8% vs. 64.0%), making it both the weakest reclamatory majority and the most polarised item between the two main groups. This suggests that the sentence may activate interpretive dimensions—such as legitimacy and speaker positioning—that are more readily recognised by ingroup respondents. Moreover, its initial *Not-R* label is therefore understandable if one foregrounds the argumentative and reported dimension of the utterance, whereas the later majority-voting shift towards a reclaimed interpretation is plausible if one gives greater weight to the explicit normalisation of in-group use in friendly contexts.

The pattern is less marked, but still visible, in some of the other reclamatory items. Input #7, for instance, also shows a noticeable difference between ingroup and outgroup assessments (87.9% vs. 78.7%), while inputs #6 and #10 display smaller gaps and remain relatively stable across the two groups. Input #5 is somewhat different again, since both groups classify it as reclamatory by majority vote, but the outgroup does so slightly more often than the ingroup (76.5% vs. 68.9%). This is analytically interesting because it suggests that explicit declarations of legitimacy or normalisation do not necessarily produce greater caution among non-LGBT+ respondents; in some cases, they may instead encourage a more immediate classification of the item as reclaimed.

Another relevant indicator is the proportion of *Idk* responses. These are comparatively low in the clearest cases, such as #2, #3, and #8, but become more substantial in inputs such as #1, #4, #5, #6, and #10. This is methodologically important, because it suggests that even when a majority interpretation emerges, a non-negligible share of respondents still experience uncertainty. In other words, the data do not simply sort the items into two neat groups; they also indicate varying degrees of interpretive stability, which is precisely what makes these materials useful for a reclamation-oriented analysis.

7.2.4.2.2 Item 2B: Reclamation Explanation This item as an open question allows us to intercept the meta-reflexive level. Although the comment was not mandatory, we collected nearly 200 responses for each input. More precisely, the number of non-empty comments ranges from 157 to 183 across the ten inputs. Given both the quantity and the qualitative richness of this material, the analysis presented here does not aim at a fully exhaustive coding of every single response. Rather, it proceeds by identifying recurrent interpretive themes in the comments, with particular attention to the aspects that are methodologically most informative for the present study.

Overall, the comments confirm that participants rarely rely on the lexical item alone when judging whether a use is reclamatory. Across the focal inputs, respondents repeatedly referred to a relatively stable set of interpretive dimensions, namely **speaker positioning**, **ingroup versus outgroup legitimacy**, **relational proximity**, **communicative intent**, and the **distinction between directly enacting a reclaimed use and merely commenting on it**. In this respect, Item 2B is especially relevant because it makes explicit the metalinguistic reasoning that remains only implicit in the closed responses to Item 2A. In this sense, the comments also make more explicit the same interpretive dimensions that had already emerged in our preliminary qualitative reading of the input sentences in Section 7.2.2, where the *R*/*Not-R* labels were treated as non-self-sufficient and as dependent on contextual and discursive cues.

Input #1 is the most revealing case in this respect. As noted above, it is the only item that reverses the preliminary ReCLAIM annotation, while at the same time showing the lowest reclamatory consensus and the largest gap between ingroup and outgroup respondents. The open comments help explain why. Across both groups, one major cluster of responses interprets the sentence as **reclamatory because it normalises the use of the highlighted term within a restricted relational circle**: participants repeatedly refer to friendship, permission, trust, and the absence of offensive intent. In this reading, the sentence is taken to indicate that the speaker accepts—and to some extent controls—the term's use within a familiar ingroup context. A second cluster, however, insists that this is not yet reclamation proper, but **rather a statement about who is allowed to use the term**. In these comments, respondents stress that the speaker is not directly reclaiming the slur for self-description, but delimiting its legitimate use. This is consistent with our earlier

discussion of #1, where the sentence was treated less as a straightforward reclaimed use than as a meta-discursive statement about legitimacy and community-bound circulation. A third, related cluster foregrounds residual ambiguity: several comments explicitly state that **the term may still retain a negative value, especially when used by outsiders**, and that the sentence therefore reflects a partial, or still unstable, form of reclamatory practice rather than a fully settled one.

The difference between ingroup and outgroup comments is particularly interesting here. Ingroup respondents more often appear willing to treat **relational permission and controlled ingroup circulation as sufficient evidence of reclamation**, even when the use is mediated or discussed rather than directly enacted. Outgroup respondents, by contrast, more frequently **separate accepting a term from friends from actually reclaiming it**, and are correspondingly more likely to describe the sentence as hypothetical, conditional, or merely explanatory. In other words, the divergence already visible in the quantitative result for #1 seems to reflect two slightly different interpretive thresholds: one more ready to include negotiated, context-bound uses within reclamation, and one more demanding with respect to direct self-ascription.

Inputs #3 and #5 are also useful because they bring out the role of metadiscourse more explicitly. In #3, the comments are strikingly convergent across groups: the dominant reading is that the sentence describes or explains reclamatory usage without performing it. Respondents repeatedly note that the speaker explicitly marks a boundary between "gay people" and themselves ("If I say it, I'd be a homophobe"), and this is taken as evidence that the utterance is about legitimacy rather than an instance of reclamation. This is methodologically important because it shows that participants are able to distinguish between metalinguistic awareness of reclamation and reclamation as a situated practice.

Input #5, by contrast, is generally read as reclamatory, but its comments reveal a more complex picture than the majority vote alone might suggest. Many respondents justify the positive assessment by referring to the explicit self-positioning of the speaker ("we are literally gay") and to the overt defence of the right to use these terms. At the same time, a non-negligible number of comments—especially among ingroup respondents—problematise the sentence's categorical claim that no one uses such words as insults any more. These

responses do not necessarily reject the reclamatory interpretation altogether, but they question its universalising tone, pointing out that the derogatory force of the terms is not evenly erased across contexts, speakers, or audiences. In this sense, #5 is revealing because it shows that even an apparently explicit declaration of reclamation may be accepted only with reservations when respondents perceive it as overstating the social stabilisation of the process.

The comments on #6 and #7 are especially relevant for the present thesis because they concern uses of *frocia* that are more syntactically and pragmatically complex than direct nominal self-labelling. In #6 ("serata frocia"), many comments interpret the phrase as reclamatory because *frocia* is used to characterise an event or scene in a positive, playful, or community-internal way, often in relation to the explicit Pride reference. Several respondents note that the word seems to describe the "nature of the evening" rather than insult a person. At the same time, this item also attracts a substantial number of uncertain comments, especially because participants are not always sure whether the speaker belongs to the community and because the skull emoji is interpreted in divergent ways: for some it reinforces irony or informality, whereas for others it introduces ambiguity or even a possible negative stance. Input #7 produces a partly similar but more stable pattern. Here *frocia* again appears in adjectival position, but within an overtly self-referential account ("with short purple hair and visibly *frocia*"). Many comments, in both groups, treat this as reclamatory precisely because the speaker appears to use the term to describe themselves and their social visibility. Yet even in this case, some respondents remain uncertain, especially where they read the adjective less as a self-claimed identity marker than as a descriptor of how the speaker is perceived by others.

Altogether, the comments collected in Item 2B reinforce three broader points. First, participants consistently treat reclamation as a contextual and relational phenomenon, rather than as a stable property of a lexical item in isolation. Second, the data confirm the analytical importance of metadiscourse: several comments show that respondents do not automatically equate talking about reclamation with performing it. Third, the comments on *frocia*—especially in #6 and #7—provide further support for the idea that some Italian forms follow language-specific and pragmatically unstable trajectories of reclamation, whose interpretation depends on more than speaker identity

alone.

7.2.4.2.3 Item 2C: Self-reported Similar Use Finally, Item 2C shifts the focus from the interpretation of the input sentence to participants' own reported linguistic practices. This question is especially relevant because it makes it possible to compare perceived reclamation with self-reported familiarity or participation in similar uses. In methodological terms, it therefore helps distinguish between recognising a reclamatory use and personally engaging in it—two dimensions that may overlap, but should not be assumed to coincide automatically. Table 7.2 reports the distribution of responses for each input and group.

Unlike Item 2A, where six inputs were overall recognised as reclamatory, Item 2C shows that recognition does not automatically translate into self-reported participation in similar practices. Across the overall sample, all ten inputs receive a majority of "No" responses. Even when a sentence is widely perceived as reclamatory, respondents are generally more cautious when asked whether they themselves use the highlighted word in a comparable way.

This is particularly clear for the six inputs previously classified as reclamatory (#1, #2, #5, #6, #7, and #10). Among these, the highest overall proportion of self-reported similar use is found for #6 (41.2%), followed by #1 (36.2%), whereas #5 and #7 remain below 30%, #2 reaches 21.5%, and #10 16.5%. The input most clearly interpreted as reclaimed in Item 2A (#2, with 91.4% "Yes") is therefore not the one participants most readily associate with their own practice.

The contrast between ingroup and outgroup respondents makes this pattern even clearer. Among outgroup participants, "No" is the majority response for every input, including those previously classified as reclamatory. By contrast, within the ingroup, three inputs receive a majority of "Yes" responses: #1 (59.1%), #6 (65.9%), and #7 (50.8%). A fourth case, #5, is almost evenly split (43.2% "Yes"; 44.7% "No"). This suggests that self-reported similar use is much more concentrated within the LGBT+ subgroup, but also that not all reclamatory items are equally close to participants' own repertoires.

This is especially relevant because the three ingroup-majority items (#1, #6, and #7) are also those in which *frocia*, or the pair *frocia/troia*, appears in

Input	Group	Yes	No	Idk
#1	*Ingroup (LGBT+)*	59.1% (78)	33.3% (44)	7.6% (10)
	Outgroup (non-LGBT+)	14.0% (19)	80.9% (110)	5.1% (7)
	Other	36.4% (4)	54.5% (6)	9.1% (1)
	Overall	36.2% (101)	**57.3% (160)**	6.5% (18)
#2	*Ingroup (LGBT+)*	33.3% (44)	64.4% (85)	2.3% (3)
	Outgroup (non-LGBT+)	10.3% (14)	87.5% (119)	2.2% (3)
	Other	18.2% (2)	72.7% (8)	9.1% (1)
	Overall	21.5% (60)	**76.0% (212)**	2.5% (7)
#3	*Ingroup (LGBT+)*	19.7% (26)	71.2% (94)	9.1% (12)
	Outgroup (non-LGBT+)	18.4% (25)	75.0% (102)	6.6% (9)
	Other	27.3% (3)	45.5% (5)	27.3% (3)
	Overall	19.4% (54)	**72.0% (201)**	8.6% (24)
#4	*Ingroup (LGBT+)*	11.4% (15)	79.5% (105)	9.1% (12)
	Outgroup (non-LGBT+)	6.6% (9)	89.0% (121)	4.4% (6)
	Other	0.0% (0)	90.9% (10)	9.1% (1)
	Overall	8.6% (24)	**84.6% (236)**	6.8% (19)
#5	*Ingroup (LGBT+)*	43.2% (57)	44.7% (59)	12.1% (16)
	Outgroup (non-LGBT+)	14.0% (19)	77.2% (105)	8.8% (12)
	Other	27.3% (3)	45.5% (5)	27.3% (3)
	Overall	28.3% (79)	**60.6% (169)**	11.1% (31)
#6	*Ingroup (LGBT+)*	65.9% (87)	29.5% (39)	4.5% (6)
	Outgroup (non-LGBT+)	18.4% (25)	78.7% (107)	2.9% (4)
	Other	27.3% (3)	63.6% (7)	9.1% (1)
	Overall	41.2% (115)	**54.8% (153)**	3.9% (11)
#7	*Ingroup (LGBT+)*	50.8% (67)	43.2% (57)	6.1% (8)
	Outgroup (non-LGBT+)	7.4% (10)	89.0% (121)	3.7% (5)
	Other	18.2% (2)	63.6% (7)	18.2% (2)
	Overall	28.3% (79)	**66.3% (185)**	5.4% (15)
#8	*Ingroup (LGBT+)*	5.3% (7)	93.2% (123)	1.5% (2)
	Outgroup (non-LGBT+)	5.9% (8)	91.9% (125)	2.2% (3)
	Other	18.2% (2)	63.6% (7)	18.2% (2)
	Overall	6.1% (17)	**91.4% (255)**	2.5% (7)
#9	*Ingroup (LGBT+)*	14.4% (19)	79.5% (105)	6.1% (8)
	Outgroup (non-LGBT+)	16.2% (22)	80.9% (110)	2.9% (4)
	Other	9.1% (1)	81.8% (9)	9.1% (1)
	Overall	15.1% (42)	**80.3% (224)**	4.7% (13)
#10	*Ingroup (LGBT+)*	25.0% (33)	65.9% (87)	9.1% (12)
	Outgroup (non-LGBT+)	8.8% (12)	82.4% (112)	8.8% (12)
	Other	9.1% (1)	90.9% (10)	0.0% (0)
	Overall	16.5% (46)	**74.9% (209)**	8.6% (24)

TABLE 7.2: Responses to Item 2C by Input and Group. For each Input, the Majority Vote for the General Population (Overall) is Highlighted in Bold.

particularly situated ways: within a restricted circle of permission (#1), as part of an event label (#6), or in adjectival self-description (#7). Taken together, these cases suggest that some uses involving *frocia* may be perceived, at least by ingroup participants, as closer to lived or imaginable practice than other reclaimed forms.

By contrast, #2 and #10 show that a use may be widely recognised as reclamatory without being reported as part of one's own repertoire. The same is true, in a different way, for #5: despite its explicit metadiscursive defence of reclamation, it does not generate a corresponding majority of self-reported similar use, not even within the ingroup. Inputs #4 and #8, instead, attract the lowest levels of self-reported similar use overall (8.6% and 6.1%, respectively), thereby confirming the contrast with the reclamatory cluster.

Taken together, Item 2C adds an important layer to the analysis. Whereas Item 2A and Item 2B show how respondents identify and explain reclamatory uses, Item 2C shows that recognition and personal use remain distinct.

7.2.4.2.4 Part 2. Word-level Reclamation Assessment Task (Discussion) Altogether, the results of Part 2 suggest that participants' judgements on reclamation are structured by a relatively coherent set of interpretive criteria.

First, Item 2A broadly confirms the preliminary binary annotation inherited from the ReCLAIM Project, with the important exception of input #1, which is reclassified by majority vote as reclamatory. This overall alignment suggests that, generally, participants do recognise a distinction between more clearly abusive and more plausibly reclaimed uses, even if this distinction remains unstable in some cases and, definitely, cannot be reduced to the lexical presence of a slur alone.

Second, the open explanations collected in Item 2B clarify how participants arrive at their judgements. Across the focal inputs, respondents repeatedly refer to speaker positioning, ingroup versus outgroup legitimacy, relational closeness, communicative intent, and the distinction between enacting a reclaimed use and merely commenting on it. This is particularly evident in the metadiscursive items, especially #3, but it also helps explain the instability of #1.

Third, Item 2C introduces a further distinction between recognising reclamation and reporting similar personal use. In the overall sample, no input

receives a majority of "Yes" responses, even when the same input is widely recognised as reclamatory in Item 2A. The contrast is especially clear for items such as #2 and #10, which are generally interpreted as reclaimed but are only rarely reported as part of respondents' own practice. This suggests that acceptability and personal use should be treated as non-equivalent dimensions.

The contrast between ingroup and outgroup respondents sharpens this picture further. While both groups often converge in recognising the overall orientation of the input, the ingroup is much more likely to report familiarity with similar uses, especially in cases such as #1, #6, and #7. This is particularly relevant because these are also the items in which *frocia*, or related forms, appears in highly situated ways. More broadly, the results support the view that some Italian forms of reclamation follow language-specific trajectories, whose interpretation may depend on both utterance configuration and community-internal circulation.

From a methodological point of view, Part 2 also confirms the value of combining closed and open questions. The closed responses make it possible to identify relatively stable tendencies across items and groups, whereas the open comments reveal the metalinguistic reasoning behind those choices. Considered together, the three items confirm that reclamation is understood as a socially negotiated practice whose recognition and use depend on a combination of different factors.

7.2.4.3 Part 1. Sentence-Level Linguistic Attribution Task (Results)

Although Part 1 appeared first in the questionnaire and therefore constituted participants' first contact with the input sentences, it is discussed here only after Part 4 and Part 2 for analytical reasons. As explained above, this order allows us first to define the respondent groups and then to establish which inputs were predominantly interpreted as reclamatory. Only at that point does it become methodologically more useful to return to the first-impression judgements collected in Part 1 and examine how participants attributed contextual and pragmatic properties to the same sentences. In this sense, Part 1 can be read as capturing an initial layer of interpretation, prior to the more explicit metalinguistic reflection elicited by the later tasks. However, for the purposes of the present chapter, we do not dwell on items 1A, 1C and 1D in

detail, as the most analytically informative results in this part concern the attribution of writer membership.

7.2.4.3.1 Item 1A: Context Attribution Since this item allowed multiple selections, its results are primarily informative at a descriptive level. Some tendencies are worth noting. Across the ten inputs, the *friendly context* was by far the most frequently selected option, and it was especially dominant for #4 (96.1%), #2 and #6 (93.9% each), #5 (91.4%), and #10 (90.7%). At the same time, some sentences activated a less bounded contextual reading. This was particularly the case for #3, #7, and #9, where more than half of the respondents selected more than one context (55.6%, 53.0%, and 57.7%, respectively), suggesting that these utterances were not anchored to a single interactional frame. Input #1 is also revealing in this respect, since it was associated not only with a friendly context (56.6%) but also quite strongly with an activist one (46.6%). As for the two cases involving *frocia* in adjectival position, #6 was read mainly as friendly (93.9%) and secondarily as activist (29.0%), whereas #7, while still predominantly friendly (85.3%), generated a somewhat more layered distribution, including activist (39.1%) and familial (27.6%) attributions. In methodological terms, this item is particularly relevant for those sentences whose interpretation depends strongly on contextual anchoring, such as metadiscursive inputs or cases involving *frocia* in adjectival position.

7.2.4.3.2 Item 1B: Membership Attribution Among the items in Part 1, this is the most analytically relevant for the present study, because it offers an indirect point of access to participants' assumptions about writer identity or community membership. As repeatedly shown in the literature on reclamation and confirmed by the focus group, the perceived legitimacy of a potentially reclaimed use is often tied to who is understood to be speaking. Item 1B is therefore especially useful because it allows us to observe from which factors (or combinations of them) factors participants infer LGBT+ membership.

The pattern emerging from Item 1B is remarkably consistent with the later reclamation assessment presented in Part 2. More specifically, the six inputs that were ultimately interpreted as predominantly reclamatory in Item 2A (#1, #2, #5, #6, #7, and #10) are also the six inputs for which the writer is attributed LGBT+ membership by majority vote in the overall sample. Conversely, the

Input	Group	Yes	No	Idk
#1	*Ingroup (LGBT+)*	68.2% (90)	3.0% (4)	28.8% (38)
	Outgroup (non-LGBT+)	34.6% (47)	14.7% (20)	50.7% (69)
	Other	36.4% (4)	27.3% (3)	36.4% (4)
	Overall	**50.5% (141)**	9.7% (27)	39.8% (111)
#2	*Ingroup (LGBT+)*	97.0% (128)	0.8% (1)	2.3% (3)
	Outgroup (non-LGBT+)	92.6% (126)	2.9% (4)	4.4% (6)
	Other	54.5% (6)	9.1% (1)	36.4% (4)
	Overall	**93.2% (260)**	2.2% (6)	4.7% (13)
#3	*Ingroup (LGBT+)*	0.0% (0)	95.5% (126)	4.5% (6)
	Outgroup (non-LGBT+)	1.5% (2)	91.2% (124)	7.4% (10)
	Other	0.0% (0)	54.5% (6)	45.5% (5)
	Overall	0.7% (2)	**91.8% (256)**	7.5% (21)
#4	*Ingroup (LGBT+)*	3.0% (4)	73.5% (97)	23.5% (31)
	Outgroup (non-LGBT+)	8.1% (11)	69.9% (95)	22.1% (30)
	Other	0.0% (0)	54.5% (6)	45.5% (5)
	Overall	5.4% (15)	**71.0% (198)**	23.7% (66)
#5	*Ingroup (LGBT+)*	81.8% (108)	3.0% (4)	15.2% (20)
	Outgroup (non-LGBT+)	81.6% (111)	6.6% (9)	11.8% (16)
	Other	45.5% (5)	9.1% (1)	45.5% (5)
	Overall	**80.3% (224)**	5.0% (14)	14.7% (41)
#6	*Ingroup (LGBT+)*	63.6% (84)	5.3% (7)	31.1% (41)
	Outgroup (non-LGBT+)	64.0% (87)	5.1% (7)	30.9% (42)
	Other	45.5% (5)	0.0% (0)	54.5% (6)
	Overall	**63.1% (176)**	5.0% (14)	31.9% (89)
#7	*Ingroup (LGBT+)*	87.9% (116)	2.3% (3)	9.8% (13)
	Outgroup (non-LGBT+)	88.2% (120)	2.2% (3)	9.6% (13)
	Other	54.5% (6)	0.0% (0)	45.5% (5)
	Overall	**86.7% (242)**	2.2% (6)	11.1% (31)
#8	*Ingroup (LGBT+)*	1.5% (2)	90.2% (119)	8.3% (11)
	Outgroup (non-LGBT+)	2.9% (4)	87.5% (119)	9.6% (13)
	Other	0.0% (0)	72.7% (8)	27.3% (3)
	Overall	2.2% (6)	**88.2% (246)**	9.7% (27)
#9	*Ingroup (LGBT+)*	2.3% (3)	59.8% (79)	37.9% (50)
	Outgroup (non-LGBT+)	0.7% (1)	61.8% (84)	37.5% (51)
	Other	0.0% (0)	36.4% (4)	63.6% (7)
	Overall	1.4% (4)	**59.9% (167)**	38.7% (108)
#10	*Ingroup (LGBT+)*	78.0% (103)	4.5% (6)	17.4% (23)
	Outgroup (non-LGBT+)	77.2% (105)	3.7% (5)	19.1% (26)
	Other	54.5% (6)	0.0% (0)	45.5% (5)
	Overall	**76.7% (214)**	3.9% (11)	19.4% (54)

TABLE 7.3: Responses to Item 1B by Input and Group. For each input, the majority vote for the general population (Overall) is highlighted in bold.

four items that remain predominantly non-reclamatory in Part 2 (#3, #4, #8, and #9) are also those for which the writer is not perceived as belonging to the LGBT+ community. This correspondence is methodologically important, because it suggests that writer membership is one of the main inferential cues through which participants orient their initial interpretation of potentially reclaimed language.

At the same time, the results also show that this attribution is not equally stable across all inputs. Input #2 is the clearest case, with 93.2% of the overall sample attributing LGBT+ membership to the writer, followed by #7 (86.7%), #5 (80.3%), and #10 (76.7%). In these cases, respondents appear to rely on relatively explicit textual cues, such as overt self-reference, mention of shared group membership, or discursive environments that are strongly associated with LGBT+ contexts. By contrast, input #1 is much more unstable: although it still receives a majority of "Yes" responses overall (50.5%), uncertainty is extremely high (39.8%), and the gap between ingroup and outgroup respondents is by far the largest in the table. While 68.2% of ingroup respondents attribute LGBT+ membership to the writer, only 34.6% of outgroup respondents do so, and more than half of the latter select "I do not know". This is highly consistent with what emerged in Part 2, where #1 was also the most weakly stabilised reclamatory item and the only one to overturn the preliminary annotation.

Input #6 occupies an intermediate position. Here too the writer is attributed LGBT+ membership by majority vote (63.1%), but uncertainty remains substantial (31.9%). This is particularly interesting because the sentence contains contextual cues such as "*Pride*" and "*serata frocia*", which appear sufficient for a majority reading, but not strong enough to eliminate doubt. In other words, the item seems to require participants to infer community membership from other cues than from explicit self-labelling. A related, though inverse, pattern can be observed in input #9: the writer is generally not perceived as belonging to the LGBT+ community (59.9% No), yet uncertainty is again very high (38.7%). This suggests that even when the evaluative orientation of the sentence is relatively clear, writer identity may remain partly opaque if it is not textually anchored.

Overall, these results indicate that perceived writer's membership is neither trivial nor automatically recoverable from the sentence alone. Rather, it is

reconstructed through a combination of lexical and pragmatic cues, and this reconstruction seems to play a major role in shaping whether a use is later recognised as reclamatory. In this respect, Item 1B provides one of the clearest links between the first-impression task resulting from Part 1 and the more explicit reclamation judgements deriving from Part 2.

7.2.4.3.3 Item 1C: Offensiveness As expected, the offensiveness scores broadly follow the distinction that later emerges more explicitly in Part 2. The highest overall mean values are found for inputs #8 (M = 4.72), #9 (M = 4.30), and #4 (M = 4.01), that is, the items that were also classified as predominantly non-reclamatory. By contrast, the lowest means are associated with inputs #2 (M = 1.30), #7 (M = 1.33), and #6 (M = 1.38), all of which were later recognised as predominantly reclamatory. Particularly interesting is input #3, whose mean value is intermediate (M = 2.86): this is consistent with its meta-discursive character, since the sentence explicitly discusses differential legitimacy of use without straightforwardly performing either a reclaimed or an abusive use. Overall, Item 1C suggests that participants' first-pass judgements about offensiveness already anticipate much of the later distinction between more clearly harmful and more plausibly reclaimed uses.

7.2.4.3.4 Item 1D: Hatefulness The hatefulness scores show a very similar pattern, though in slightly more conservative form. Again, the highest overall mean values are found for #8 (M = 4.60), #9 (M = 3.95), and #4 (M = 3.76), while the lowest are associated with #2 (M = 1.35), #6 (M = 1.36), and #7 (M = 1.52). The fact that the ranking is so close to that observed for offensiveness suggests that participants do not radically separate the two dimensions at this early stage, but rather use them as related indicators of negative stance.

At the same time, hatefulness scores remain, overall, slightly lower than offensiveness scores. This may suggest that some sentences are perceived as offensive without necessarily being interpreted as fully hateful. Still, this item was designed with reference to the strand of research on HS detection within which the present thesis is situated. The results do not allow us to conclude whether, or to what extent, respondents were identifying hate speech as a broader phenomenon in its own right.

7.2.4.3.5 Part 1. Sentence-Level Linguistic Attribution Task (Discussion)

Collectively, the items in Part 1 help reconstruct the interpretive background against which the later judgement on reclamation is made. Above all, they show that participants' first readings of the inputs are shaped by assumptions about context and, even more clearly, by inferences about writer membership. Among the dimensions considered here, perceived writer membership appears especially important, because it seems to function as a key inferential step between the sentence-level task and the later assessment of reclamation. The fact that the overall majority pattern of Item 1B aligns so closely with the later Item 2A results strongly suggests that participants rely, at least in part, on a prior judgement about who is speaking when deciding whether a use is reclaimed.

From the point of view of the present thesis, this is methodologically significant. If the interpretation of a potentially reclaimed use depends, at least in part, on the writer being inferred as a community member, then reclamation-sensitive annotation must also take into account the unstable, and often only implicit, process through which speaker or writer identity is reconstructed in context. This becomes especially clear in the more uncertain cases, such as #1 and #6, where the sentence seems to invite an LGBTQ+ reading without ever making it fully explicit. In this sense, Part 1 reinforces one of the central claims of the chapter: reclamation is not recognised through lexical material alone, but through an interpretative process that is already pragmatic from the first encounter with the utterance.

7.2.4.4 Part 3. Respondents' General Opinions and Insights (Results)

After examining sentence-level interpretations and word-level assessments, we finally turn to the third part of the questionnaire, which moves from stimulus-bound judgements to respondents' broader opinions and self-reported experiences. Compared with the previous sections, this part is less closely tied to the ten input sentences taken individually and instead offers a more general perspective. For this reason, it is particularly useful in two respects. On the one hand, it allows us to test whether the patterns observed in the more localised tasks also reappear when participants are asked to reflect on the phenomenon at a more general level. On the other hand, it provides explicitly metalinguistic and, at times, meta-methodological material that helps

situate the questionnaire within the broader aims of this case study. As in the previous sections, the discussion below is selective and focuses primarily on those results that are most relevant to the research questions and to the interpretive issues already highlighted by the focus group, the ReCLAIM Project, and the literature reviewed in Chapter 3.

7.2.4.4.1 Item 3A: Perceived Exposure With regard to the Perceived Exposure Question, the data suggest that exposure to reclamatory uses of slurs is not distributed evenly across the sample, even though the overall proportion of affirmative responses is high. In the general population, 243 respondents out of 279 (87.1%) answered "Yes", while 21 (7.5%) answered "No" and 15 (5.4%) selected "I do not know". This confirms that the phenomenon is widely recognised within the sample, but also that such recognition is not uniform across all socio-demographic dimensions.

When responses are considered by age group, individuals aged 55 and above are more likely to answer negatively to the question asking whether they encounter reclamatory language in everyday life. More specifically, negative responses are proportionally more frequent among respondents aged 55–64 and over 65, whereas the younger groups show a more balanced distribution and, in absolute terms, a stronger concentration of affirmative responses. Although the majority response remains "Yes" across all age brackets, the older groups appear less likely to report direct familiarity with reclamatory language in everyday experience.

Group belonging is also relevant. LGBT+ respondents are more likely than non-LGBT+ respondents to report familiarity with reclamatory uses of language: 128 out of 132 ingroup participants (97.0%) answered "Yes", compared with 106 out of 136 outgroup participants (77.9%). Negative and uncertain responses are correspondingly more frequent in the outgroup. This difference is analytically important, as it suggests that reclamatory language is not only more visible within the LGBT+ subgroup, but may also be more easily recognised as a meaningful and recurrent practice by those who are within the community.

This tendency appears to intersect with age, since the LGBT+ group is more strongly represented in the younger age brackets, particularly among respondents aged 25–34 and 35–44, where affirmative responses are especially

frequent. At the same time, the data also suggest some geographical variation. Among respondents who answered "Yes", ingroup participants show a higher proportion of affirmative responses from the North (69.5%) and from abroad (7.0%), whereas outgroup affirmative responses are relatively more frequent in the South and Islands (27.4% vs. 13.3% in the ingroup). However, these geographical patterns should be interpreted with caution, since the sample does not mirror Italian demographics evenly and some areas—especially Northern regions, and particularly Piedmont—are more strongly represented than others.

These results suggest that perceived exposure to reclamatory language is shaped by both group belonging and age. More specifically, the phenomenon appears more familiar and more readily identifiable among LGBT+ respondents and among younger participants, while older respondents and non-LGBT+ participants are comparatively more likely to report limited exposure or uncertainty. In methodological terms, this is relevant because it indicates that participants do not approach reclamation from a uniform experiential background: for at least part of the sample, the phenomenon is not merely an abstract object of judgement, but something already encountered in everyday life.

7.2.4.4.2 Item 3B: Legitimacy The Legitimacy Question provides a more articulated picture than the previous items, both because it was structured as a checklist allowing multiple selections and because respondents could also add short open-ended specifications. For this reason, in order to distinguish firmer positions from more articulated ones, responses were first divided into *exclusive* and *broader/multiple* responses.

For analytical purposes, the coding of the Legitimacy Question took into account not only the checklist selections (i.e. "Persone LGBT+", "Chiunque", or "Nessuna persona"), but also short free-text qualifications where these remained conceptually aligned with one primary position. Accordingly, *exclusive* responses include both straightforward single-option answers and briefly qualified responses that did not introduce additional actor categories. References to art, literature, and entertainment were likewise identified through a broader qualitative coding that included both explicit selections and clearly equivalent free-text formulations.

This coding choice makes it possible to compare more categorical stances with responses in which legitimacy is explicitly distributed across several social actors or interactional contexts.

Group	Exclusive responses	Broader/multiple responses
Ingroup (LGBT+)	25.8% (34)	74.2% (98)
Outgroup (non-LGBT+)	39.7% (54)	60.3% (82)

TABLE 7.4: Distribution of exclusive versus broader/multiple responses to the Legitimacy Question by group.

As Table 7.4 shows, LGBT+ respondents were less likely to adopt an exclusive stance and more likely to provide broader or multiple responses. This suggests that, within the ingroup, legitimacy is more often treated as a nuanced and context-dependent issue rather than as a matter of fixed entitlement. By contrast, outgroup respondents still frequently gave articulated answers, but were comparatively more likely to choose a single, more categorical position.

Group	LGBT+ only	Anyone	No one
Ingroup (within exclusive responses)	47.1% (16)	26.5% (9)	26.5% (9)
Outgroup (within exclusive responses)	22.2% (12)	46.3% (25)	31.5% (17)

TABLE 7.5: Distribution of positions within the exclusive-response category for the Legitimacy Question.

Within the LGBT+ group, 34 respondents out of 132 (25.8%) provided an exclusive response, whereas the majority (98, i.e. 74.2%) gave broader or multiple responses. Among the exclusive responses, 9 participants (6.8% of the ingroup) stated that no one should be allowed to use such terms, 16 (12.1%) restricted legitimacy to LGBT+ individuals, and 9 (6.8%) selected a position aligned with "Anyone". The first of these ingroup-restrictive stances was sometimes explicitly linked to the very definition of reclamation as a form of self-referential or community-based reuse, as in the following comment:

> altrimenti come regge la definizione di "riappropriazione" che è stata presentata? "l'utilizzo di termini dispregiativi per esprimere

> la PROPRIA identità e/o l'appartenenza ad una certa comunità"[20]

Another respondent similarly restricted legitimacy to

> Persone LGBT che siano risolte e non riversino il loro odio per se stessi verso gli altri[21]

By contrast, the more permissive exclusive responses were often qualified by the condition that the terms should not be used offensively, as in "Purché non siano veicolo di offesa"[22] or "Chiunque purché non vengano utilizzati con toni offensivi"[23].

The majority of ingroup respondents, however, fell into the broader/multiple category. These responses consistently suggest that legitimacy is more often treated as context-dependent than absolute. In most cases, respondents recognised the acceptability of use within a relatively restricted circle—typically centred on LGBT+ individuals and closely connected actors such as activists, friends, and family members. Using a broader qualitative coding that included both explicit category selections and clearly equivalent free-text formulations, 19 participants also extended legitimacy to entertainment and/or art and literature, an inclusion that is effectively consistent with Bianchi (2015)'s definition of *community uses* (see Section 3.2.1). Among them, 10 referred to both domains, whereas 9 referred only to art and literature. The associated comments are particularly revealing because they show that these extensions were not understood as unconditional. For instance, one respondent wrote:

> Chiunque riesca ad utilizzare determinate parole ma in un contesto ironico. Non persone quali professori, istituzioni pubbliche ecc.[24]

Another specified:

[20] Our translation: otherwise, how does the definition of "reclamation" that was presented hold up? the use of derogatory terms to express ONE'S OWN identity and/or belonging to a certain community.

[21] Our translation: LGBT people who have come to terms with their identity and do not project their self-hatred onto others.

[22] Our translation: provided that they do not function as vehicles of offence.

[23] Our translation: by anyone, provided that they are not used in an offensive way

[24] Our translation: Anyone who can use certain words, but in an ironic context. Not people such as teachers, public institutions, etc.

> Solo per sè e persone con cui si è affrontato l'argomento direttamente[25]

Particularly significant are also those comments that distinguish between fictional characters and their authors ("personaggi di opere di finzione o meno in libri, cinema etc.; non automaticamente i loro autori"[26]), or that restrict legitimacy within artistic and literary domains to queer speakers ("persone del mondo dell'arte o della letteratura se sono queer"[27]). A further respondent argued that such uses might be acceptable more broadly only if their reclamatory meaning were fully explicit and not open to misunderstanding:

> Penso che tutti possano utilizzarle se il tono non è dispregiativo ma anzi riappropriativo, ovviamente nel mondo dello spettacolo, arte ecc ecc è necessario che ciò sia esplicito e non fraintendibile dal pubblico[28].

Others, again, framed artistic and literary legitimacy as acceptable only for educational or activist purposes: "nell'arte e letteratura se a fini divulgativi/attivismo e non per offendere[29]".

The non-LGBT+ group displays a partly similar but not identical pattern. Here, 54 respondents out of 136 (39.7%) gave an exclusive response, whereas 82 (60.3%) provided broader or multiple responses. Among the exclusive responses, 17 participants (12.5% of the outgroup) stated that no one should be allowed to use these terms. One of them explicitly justified this stance by arguing that such words should be avoided altogether because of their vulgarity.

A further 12 respondents (8.8% of the outgroup) restricted legitimacy to LGBT+ individuals, sometimes with narrow exceptions, as in "Altre persone in

[25]Our translation: Only for oneself and people with whom the topic has been directly discussed.

[26]Our translation: fictional or non-fictional characters in books, films etc.; not automatically their authors.

[27]Our translation: figures from the worlds of art and literature, insofar as they are themselves queer.

[28]Our translation: I think everyone can use them as long as the tone is not derogatory but rather reclamatory; obviously, in entertainment, art, etc., this needs to be explicit and not open to misunderstanding by the audience.

[29]Our translation: in art and literature if for dissemination/activist purposes and not to offend.

un contesto 'protetto'"[30]. The most frequent orientation within the exclusive-response area was, however, the one aligned with "Anyone", comprising 25 respondents in total. Methodologically, this figure includes both standalone "Anyone" responses and short free-text formulations clearly expressing the same primary position. Several of these responses emphasised the importance of depoliticising or neutralising derogatory force ("Tali termini vanno usati pulendoli dell'accezione dispregiativa"[31]), or framed legitimacy in terms of non-offensive intent and mutual understanding, as in "Chiunque non abbia intento offensivo"[32] and "Chiunque purché ci sia consenso da ambo le parti e sia chiaro il modo in cui viene usato un dato termine".[33] Others adopted a more pragmatic approach, focusing on consequences rather than entitlement:

> Chiunque può farlo, bisogna solo valutare le conseguenze. Permettere o vietare l'uso di parole, che potrebbero essere definite dispregiative, ad una comunità piuttosto che ad un'altra, potrebbe non essere la soluzione. Bisogna più educare all'uso dei termini e consapevolizzare il loro significato. Informare per eliminare l'ignoranza, perché il rifiuto, la paura, la fobia è spesso alimentata da ignoranza e pregiudizi. Solo con queste basi sarà possibile limitare o concedere l'uso di linguaggio e/o termini.[34]

Finally, some respondents rejected the idea that such words belong exclusively to the LGBT+ community while maintaining a more neutral or even mildly disapproving stance towards them:

> Ciascuno può parlare come crede, certamente questi termini riappropriativi non sono appannaggio della sola comunità LGBT+.

[30] Our translation: other people in a "protected" context

[31] Our translation: Such terms should be used after stripping them of their derogatory meaning.

[32] Our translation: Anyone who does not have offensive intent.

[33] Our translation: Anyone, provided that there is consent on both sides and that the way a given term is being used is clear.

[34] Our translation: Anyone can do it; one simply has to consider the consequences. Allowing or prohibiting the use of words that could be considered derogatory for one community rather than another may not be the right solution. Instead, we should focus on educating people on language use and raising awareness of its meaning. Misinformation fuels fear and prejudice. Only by addressing ignorance can we establish meaningful guidelines on the use of certain terms.

> Personalmente, non le trovo gradevoli ma non mi disturbano nemmeno.[35]

The broader/multiple responses in the outgroup remain far from unrestricted. Most respondents in this category still concentrated legitimacy within a relatively restricted circle, typically comprising LGBT+ individuals and people socially close to them. Using the same broader qualitative coding adopted above, 14 respondents also referred to art, literature, and/or entertainment as potentially acceptable domains of use. Of these, 5 restricted their approval to art and literature specifically. Here too, the open comments repeatedly stress the importance of contextual appropriateness. Another linked legitimacy to positive, supportive, or carefully managed humorous intent:

> Chiunque lo faccia con chiaro intento positivo, di supporto, o comico (ma comico fatto bene, non Pio ed Amedeo)[36].

Finally, some comments invoked broader ethical notions such as common sense and sensitivity, for instance:

> In generale va usato il "buon senso", ossia usare la terminologia riappropriativa senza ledere la sensibilità delle persone coinvolte. Non è facile, ma basta un po' di tatto e sapersi adattare al contesto in cui ci si trova.[37]

Overall, these results suggest that the legitimacy of reclamatory language is rarely understood in rigid or absolute terms. The data do confirm that LGBT+ speakers are widely regarded as the most legitimate users of reclaimed slurs, especially within the ingroup. At the same time, however, they also complicate any overly dogmatic version of targetism as discussed in Chapter 3, since respondents in both groups frequently extend at least some degree of legitimacy to allies or specific domains such as art and entertainment. What

[35]Our translation: Everyone can speak as they see fit; certainly, these reclaimed terms are not the preserve of the LGBT+ community alone. Personally, I do not find them pleasant, but they do not bother me either.

[36]Our translation: Anyone who does so with a clear positive, supportive, or comedic intent (but comedic done well, not Pio and Amedeo)

[37]Our translation: In general, "common sense" should be used, that is, reclaimed terminology should be used without hurting the sensitivity of the people involved. It is not easy, but a bit of tact is enough, together with the ability to adapt to the context one is in.

emerges most clearly from the open comments is the centrality of three closely related notions—context, intention, and appropriateness—which repeatedly function as the criteria through which participants negotiate the boundaries of legitimate reclamatory use.

7.2.4.4.3 Item 3C: Use The Use Question further clarifies the distinction between the two main groups by shifting from judgements about exposure and legitimacy to self-reported general practice. In the overall sample, 151 respondents out of 279 (54.1%) reported that they do or have at some point used words or expressions with a reclamatory intent, while 88 (31.5%) answered "No" and 40 (14.3%) selected "I do not know". Even at this general level, therefore, the phenomenon is not marginal within the sample; at the same time, the distribution of responses is much less homogeneous than in the Perceived Exposure Question, where affirmative answers were far more widespread. This already suggests that being familiar with reclamatory language and personally using it are related but distinct dimensions.

The difference between LGBT+ and non-LGBT+ respondents is especially marked. Among ingroup participants, the large majority reported having used reclamatory expressions at least once (103 out of 132, i.e. 78.0%), whereas only 18 respondents (13.6%) answered "No" and 11 (8.3%) remained uncertain. By contrast, the outgroup displays a much more cautious profile: only 43 out of 136 respondents (31.6%) reported such use, while 68 (50.0%) stated that they had never done so and 25 (18.4%) selected "I do not know". In other words, reclamatory language appears much more readily available as an actual linguistic practice within the LGBT+ subgroup than outside it.

Age is also significantly associated with self-reported use in the current dataset, although the pattern is less linear than in the Perceived Exposure Question. The highest proportion of affirmative responses is found among respondents aged 35–44 (69.8%), followed by those aged 25–34 (55.1%) and 18–24 (52.4%). By contrast, negative responses become more frequent among respondents aged 55–64, where "No" is the majority answer (57.7%), and the oldest group also shows a comparatively high proportion of uncertainty. These results still support the broader impression that younger and middle-aged respondents are more likely to report direct involvement in reclamatory

practices, whereas older participants are more likely to position themselves at a distance from them.

This item is particularly informative when read in relation to both Item 3A and Item 2C. Compared with the Perceived Exposure Question, self-reported use is clearly more restrictive: while 87.1% of the sample reported having encountered reclamatory language in everyday life, only 54.1% reported using it themselves. This confirms that exposure does not automatically entail participation. At the same time, the contrast with Item 2C is equally revealing. In Item 2C, respondents were asked whether they had ever used the *specific highlighted word* in a way similar to that found in each input sentence, and in the overall sample all ten inputs received a majority of "No" responses. By contrast, Item 3C shows that more than half of the respondents report using reclamatory language in general. The difference is methodologically important: it indicates that respondents may recognise themselves as users of reclamatory language at a broad level, while being much more cautious when this practice is tied to a particular linguistic expression or to a more clearly delimited context of use.

Overall, these three items point in the same direction. First, they suggest that reclamatory language is more visible, more intelligible, and more actively used within the LGBT+ community than outside it. Second, they show that age also plays a role, especially with regard to perceived exposure, but also—though somewhat less linearly—with respect to self-reported use, with older respondents tending more often to report limited familiarity or lower involvement in reclamatory practices. Third, and perhaps most importantly, the open answers make clear that participants do not evaluate reclamation in purely categorical terms. Across both ingroups and outgroups, judgements repeatedly revolve around a set of closely related notions—context, intention, consent, appropriateness, and possible misinterpretation.

7.2.4.4.4 Item 3D: Additional Reclaimed Words This item is not analysed here in a fully systematic way, since the present thesis does not aim to provide a comprehensive lexical mapping of reclaimed forms. Nevertheless, the responses are methodologically valuable and will be taken up again in future work, as they offer a first exploratory inventory of further terms and domains that participants associate with reclamatory use. Although many respondents

reported not knowing additional examples, a substantial subset used this open field to mention further words, often extending the discussion beyond the specific items included in the questionnaire.

The material suggests, first of all, that participants frequently remain within the LGBTQ+ domain, mentioning forms such as *queer*, *finocchio*, *frocio* (and its gender and grammatical configurations), *checca*, *ricchione*, *lesbica*, and *lella*. Some responses also point to more local or community-specific variants, for instance *puppo* in the dialect of Catania, *femminiello* from the Neapolitan culture, or other dialectal forms such as the Genoan *buliccio*. Particularly interesting are those comments that explicitly frame reclamation as uneven across terms, as in the observation that "*lesbica*, nella sua semplicità, ha ancora bisogno di essere riappropriato come termine"[38], or that *queer* in Italian may now be perceived as relatively "sdoganato" (eng. "normalised").

At the same time, many respondents broadened the scope considerably and referred to other semantic fields. A recurrent cluster concerns misogynistic or slut-shaming terms, such as *troia*, *zoccola*, *puttana*, *mignotta*, *bagascia*, and related forms. In several cases, these words are explicitly connected to female friendship or ironic in-group use, for instance: "Utilizzo la parola *troia* e sinonimi con le mie amiche, dandovi un significato positivo"[39]. Another recurring domain concerns racialised or ethnic slurs, especially references to the "n-word", *negro*, *nigga*, and, more sporadically, terms such as *zingaro*.

A smaller but still noteworthy set of responses mentions regional or class-marked terms such as *terrone*, *polentone*, or *montanaro*, while a few respondents also refer to words associated with mental health, neurodivergence, or disability, such as *pazzo*, *matto*, *autistic**, or *neurospicy*.

Nonetheless, these answers are relevant for this thesis because they suggest that respondents do not conceptualise reclamation as confined to a single community or lexical field. Rather, they seem to understand it as a broader sociolinguistic process that may involve different histories of stigma, different communities of practice, different contexts of use, and different degrees of stabilisation across terms. In this sense, Item 3D provides useful exploratory material for future research, especially for the possible expansion of reclamation-sensitive resources beyond the set of forms directly examined in the

[38] Our translation: *lesbica*, in its simplicity, still needs to be reclaimed as a term.

[39] Our translation: I use the word *troia*, and similar terms, with my [female] friends, giving it a positive meaning.

present study.

7.2.4.4.5 Item 3E: Final Comments The final open comment field was not completed by all respondents, yet it still yielded a substantial amount of qualitatively rich material. After excluding empty entries and purely procedural fillers, we retained 64 substantive comments overall. These comments are not treated here as a systematically coded corpus in their own right; rather, they are used as an additional interpretive layer that helps clarify how respondents framed the phenomenon beyond the more structured questions. Two broad dimensions emerge with particular clarity: content-related reflections on reclamation itself and methodological feedback on the questionnaire and on the difficulties of interpreting decontextualised utterances.

From the point of view of content, the most recurrent theme is the **centrality of context and tone**. Across both ingroup and outgroup comments, respondents repeatedly stress that reclamation cannot be judged independently of who is speaking, to whom, with what intention, and in which interactional setting. In this sense, the comments strongly reinforce the patterns already observed both in Part 1 (Par. 7.2.4.3.5) and in Part 2 (Par. 7.2.4.2.4), where situational context emerged as the crucial interpretive variable and writer membership becomes less important. Here follows some examples:

> L'utilizzo di questi termini in maniera riappropriativa non è legittimato dall'appartenenza o meno alla comunità LGBTQ+, ma dall'intento.[40]

> I contesti e i toni nei quali sono usati i termini di cui ci si potrebbe riappropriare sono condizione essenziale tra l'essere offensivi o semplicemente definizioni.[41]

> Senza intonazioni o linguaggio del corpo a volte ho avuto difficoltà a cogliere il senso autentico delle frasi.[42]

[40] Our translation: The reclamatory use of these terms is not legitimised by belonging or not belonging to the LGBTQ+ community, but by intent.

[41] Our translation: The contexts and tones in which terms that one might reclaim are used are an essential condition in distinguishing between being offensive and being simply descriptive.

[42] Our translation: Without intonation or body language, I sometimes found it difficult to grasp the authentic meaning of the sentences.

> Credo che per capire se la parola è usata in senso riappropriativo o meno sia fondamentale il contesto ed il background del parlante.[43]

Nonetheless, a second major cluster, concerning **the legitimacy of use and the boundaries of reclamation**, complicates any simple or categorical answer. Even if some respondents place primary emphasis on intent, others insist that legitimacy remains largely internal to the relevant community or even to the specific subject position involved. Particularly notable are those comments that explicitly reject the idea that belonging to the broader LGBTQ+ umbrella is always sufficient for using any reclaimed term, thus pointing to internal differentiation within the community itself. The following are some examples:

> Il contesto dev'essere ironico, dissacratorio, e/o orgoglioso. Per questo non reputo accettabile che un amico o un familiare, per quanto ally, usi un termine offensivo.[44]

> Essendo una persona LGBT ma non del tutto "gay" (bisessuale) ci sono termini riappropriativi che ritengo non possano essere mai rivendicati se non toccano specificatamente la tua categoria. Es. i termini per le persone transessuali.[45]

> So che le stanno usando per riappropriarsene ma io non posso usarle perché, in qualche modo, mi sembrerebbe una riappropriazione indebita.[46]

> Vedrei come offensivo l'utilizzo di questo tipo di linguaggio da parte di persone non appartenenti alla comunità LGBT.[47]

[43] Our translation: I believe that, in order to understand whether the word is used in a reclamatory sense or not, the context and the speaker's background are fundamental.

[44] Our translation: The context must be ironic, irreverent, and/or proud. For this reason, I do not consider it acceptable for a friend or family member, however much of an ally they may be, to use an offensive term.

[45] Our translation: As an LGBT person but not entirely "gay" (I am bisexual), there are reclaimed terms that I believe can never be claimed unless they specifically concern your category. For example, terms for trans people.

[46] Our translation: I know they are using them in order to reclaim them, but I cannot use them because, in some way, it would feel like an improper reclamation.

[47] Our translation: I would see the use of this kind of language by people who do not belong to the LGBT community as offensive.

This cluster of comments partially undermines our conception of reclamation as inherited from the philosophical framework presented in this thesis, where amicable contexts are considered within the definition of *community uses* (see Section 3.2.1).

A small cluster of comments extends this stance by emphasising the idea that reclamation is more acceptable when it is **self-referential or takes place within a subjective-restricted or profoundly safe context**. Such comments explicitly frame reclamation as something that should primarily concern one's own community or small circles in which mutual understanding is already in place. This gives a very specific condition, but allows a broader interpretation of legitimacy. The context seems to be read in terms of solidarity and protection rather than in terms of pure intention. This point is especially relevant in light of the results of Item 2C, where self-reported similar use was much more selective than recognition of reclamation in principle. The following are some examples:

> Credo che il grosso discrimine sia se la parola sia usata per indicare se stessə (e una qualsiasi situazione in cui si è inclusi e si è partecipi) e non gli altri.[48]

> In quanto soggetto lesbico tendo a usare e riappropriami di termini che fanno riferimento a questa soggettività piuttosto che agli uomini gay o alle soggettività trans. Penso che anche all'interno della comunità LGBT debbano esserci sfumature.[49]

> Si possono usare parole offensive in senso riappropriativo solo riferendosi a se stessi o a membri della stessa comunità che non lo considerano sgradevole.[50]

A third recurrent theme is **the ambivalent and affectively charged character of reclamation**. The comments are not polarised into a simple opposition

[48]Our translation: I believe that the main dividing line is whether the word is used to refer to oneself (and to any situation in which one is included and participating) rather than to others.

[49]Our translation: As a lesbian subject, I tend to use and reclaim terms that refer to this subjectivity rather than to gay men or trans subjectivities. I think there must also be nuances within the LGBT community itself.

[50]Our translation: Offensive words can be used in a reclamatory sense only when referring to oneself or to members of the same community who do not find this unpleasant.

between approval and rejection. Rather, several respondents—including members of the ingroup—describe a position of hesitation and discomfort. This is particularly important because it shows that reclamation is not experienced as uniformly empowering even by those who understand its political or symbolic rationale, and the emotional experience plays a relevant role. The historical and emotional weight of terms such as *frocio*, *frocia*, *checca*, and *ricchione* remains palpable in many comments. A few examples are given below:

> Comprendo bene il valore dell'utilizzo riappropriativo di alcuni termini, ma non fa parte della mia esperienza e in alcuni casi faccio molta fatica ad accoglierlo.[51]

> Ho un costante dubbio su come mi faccia sentire la riappropriazione di tali termini, a volte preferirei non venissero usati e basta.[52]

> Anche se usate con un altro intento, io continuo a provare fastidio nel sentirle.[53]

> In generale non mi piacciono i termini frocio/a, checca, ricchione (probabilmente perché per troppo tempo hanno avuto un'accezione dispregiativa ed offensiva), quindi non riesco a vederci il senso riappropriativo ed una speranza di utilizzo degli stessi in modo neutro.[54]

A further thematic nucleus, instead, stresses the subversive potential of slurs by presenting **reclamation as a political and identity-related move**. Here respondents describe it as a practice of self-determination and symbolic resistance, which can weakens the offensive power of slurs. This more affirmative reading is visible in both groups, although it is articulated in more explicitly identity-based terms within the ingroup. Some examples follow:

[51] Our translation: I fully understand the value of the reclamatory use of certain terms, but it is not part of my own experience and, in some cases, I find it very difficult to accept.

[52] Our translation: I constantly doubt how the reclamation of such terms makes me feel; at times I would rather they were simply not used at all.

[53] Our translation: Even if they are used with a different intent, I still feel discomfort when I hear them.

[54] Our translation: In general, I do not like the terms *frocio/a*, *checca*, *ricchione* (probably because for too long they have had a derogatory and offensive meaning), so I cannot really see their reclamatory sense or any hope of their being used neutrally.

> La riappropriazione invece mostra proprio un uso dinamico del linguaggio. Decostruisce da dentro ogni forma di rigidità e attraverso la mimesi parodizza le forme di violenza rendendole grottesche.[55]

> Io uso spesso (per me stessa) le parole sopra per rivendicare la mia identità fluida e la mia libertà e autodeterminazione sessuale.[56]

> È importante utilizzare termini spesso usati in modo dispregiativo per ridurne il potere offensivo.[57]

> Penso che la riappropriazione sia un atto di ribellione e penso che i fenomeni linguistici siano importanti e compiano delle piccole grandi rivoluzioni.[58]

Some comments also explicitly foreground internal and generational differences. These are particularly useful because they caution against treating either the LGBT+ group or the non-LGBT+ group as internally homogeneous blocks, suggesting that an intersectional and positionality-based reading is definitely needed. In a few cases, age is directly invoked as a relevant factor in shaping what feels acceptable or intelligible, and therefore reclaimable. A few examples are given:

> Da donna lesbica 50 enne mi rendo conto che la mia percezione di ciò che è ammissibile / offensivo è molto diversa da quella dei / delle più giovani.[59]

> Le mie risposte non sono da tenere in considerazione perché ai miei tempi (sono nata nel 1930) un certo modo di parlare era inimmaginabile in nessun contesto e anche oggi mi infastidisce

[55] Our translation: Reclamation, instead, shows a genuinely dynamic use of language. It deconstructs every form of rigidity from within and, through mimesis, parodies forms of violence, making them grotesque.

[56] Our translation: I often use the words above (for myself) in order to reclaim my fluid identity and my sexual freedom and self-determination.

[57] Our translation: It is important to use terms that are often used derogatorily in order to reduce their offensive power.

[58] Our translation: I think reclamation is an act of rebellion, and I think linguistic phenomena are important and carry out small great revolutions.

[59] Our translation: As a 50-year-old lesbian woman, I realise that my perception of what is acceptable/offensive is very different from that of younger people.

> molto soprattutto per quanto è diffuso specie tra i giovani per me è solo turpiloquio.[60]

Finally, one comment is particularly worth highlighting because it points directly to a meta-discursive pattern that is highly relevant for the present thesis and for the interpretation of some of the input sentences discussed earlier, especially those involving *frocia*.

> Ho giusto un'osservazione da fare che magari potrebbe tornarvi utile per le prossime fasi dello studio. Noto che spesso gli slur riappropriati vengono usati dalle persone queer in contesti in cui stanno descrivendo le reazioni o le percezioni che altre persone hanno di loro (vedasi l'esempio di "ho i capelli viola e sono evidentemente frocia"). Sarebbe carino approfondire questo uso, per capire quanto sia comune e se ci siano delle differenze con gli altri contesti riappropriativi![61]

The respondent here notes that reclaimed slurs are often used by queer speakers precisely when they are describing how they are seen or perceived by others. This observation is important because it suggests that some uses may be reclamatory while still being embedded in reported or perception-oriented discourse rather than in simple self-labelling.

Alongside these content-related reflections, the final comments also contain a number of **explicitly methodological observations on the questionnaire itself**. These responses are useful because they help identify the main difficulties encountered by participants, many of which concern precisely the dimensions that are most challenging for reclamation-sensitive annotation: lack of contextual information, absence of prosodic or bodily cues, ambiguity of some formulations, and uncertainty about speaker identity. In this respect,

[60] Our translation: My answers should not be taken into consideration because, in my time (I was born in 1930), a certain way of speaking was unimaginable in any context, and even today it bothers me greatly; especially given how widespread it is among younger people, to me it is simply vulgar language.

[61] Our translation: I just have one observation that might be useful for the next stages of the study. I notice that reclaimed slurs are often used by queer people in contexts where they are describing the reactions or perceptions that other people have of them (see the example "I have purple hair and I am obviously *frocia*"). It would be interesting to investigate this use further, in order to understand how common it is and whether it differs from other reclamatory contexts.

the comments both provide feedback on the survey instrument and indirectly confirm the complexity of the phenomenon under investigation.

Overall, the final comments are especially valuable because they make two points visible at once. On the one hand, they confirm that respondents conceptualise reclamation through a dense interplay of factors, such as context, legitimacy, self-reference, historical sedimentation, and internal differentiation. On the other hand, they show that the questionnaire itself made participants confront the limits of interpreting isolated utterances without access to fuller contextual information.

7.2.4.4.6 Part 3. Respondents' General Opinions and Insights (Discussion) Collectively, the results of Part 3 provide a broader and more reflective perspective on the phenomenon already explored through the input-specific tasks.

First, they confirm that reclamatory language is widely recognised within the sample, but not evenly distributed in terms of legitimacy or familiarity. *Perceived exposure* is high overall, yet significantly more frequent among LGBT+ respondents and lower among older participants. *Legitimacy* is rarely framed in absolute terms and is instead negotiated through context, intention, relational proximity, and perceived membership. *Use*, finally, is much more selective than the perceived exposure, and is reported far more often within the ingroup than outside it. In this sense, the three items on perceived exposure, legitimacy, and use do not simply repeat the same finding from different angles, but rather map three distinct levels of engagement with reclamatory language: encountering it, judging it, and taking part in it.

Second, the final comments help clarify the interpretive logic underlying these more structured results. Across both ingroups and outgroups, respondents repeatedly stress that reclamation cannot be reduced to the lexical item alone. Rather, it depends on who is speaking, in which context, with what tone, towards whom, and with what degree of shared understanding. This confirms, at a more general level, what had already emerged in Parts 1 and 2: namely, that writer or speaker identity, legitimacy, and discourse configuration are central to the interpretation of potentially reclaimed language. Particularly significant, in this respect, are the comments that foreground self-reference, restricted circles of use, and the persistence of emotional or

historical discomfort even in the presence of non-offensive intent. These responses show that reclamation is not experienced as a stable or uniformly accepted phenomenon, but as one whose boundaries need to be negotiated, as they are affectively loaded and internally differentiated. Rejection is minoritarian, but still present—even if often motivated by particular social or individual positions—and may be read as a partial adhesion of the community to a *deflationary perspective*. However, most comments on legitimacy reinforce the idea that reclamation is not always exclusively reserved for the target community. The recurrent emphasis on appropriateness, irony, political meaning, and consent complicates any rigid opposition between "allowed" and "forbidden" uses. Likewise, the references to terms such as *frocio/frocia*, and especially to uses in which queer speakers describe how they are perceived by others, suggest that some reclamatory practices may be embedded in metadiscursive or reported configurations. This is methodologically important, because it highlights a type of use that may be especially difficult to capture through binary annotation alone, moreover when it appears to be highly relevant for the Italian case.

Finally, the methodological feedback contained in the open comments confirms one of the central premises of this case study: that decontextualised sentences, while analytically useful, make visible the very limits that a reclamation-sensitive approach must confront. Participants explicitly point to missing contextual information, ambiguous wording, absence of prosodic or bodily cues, and uncertainty about the speaker's background—precisely the same dimensions that repeatedly shape their substantive judgements. In this sense, Part 3 closes the analytical loop of the questionnaire: while broadening the empirical picture of reclamation in the LGBTQ+ Italian context, it also shows why a socially situated and participatory approach remains necessary for such phenomena to be first understood, and then annotated and eventually modelled in a way that is both linguistically and ethically adequate.

7.3 Final Discussion

Starting from the final comments discussed above, what emerges most clearly throughout our survey is the constitutive complexity of reclamation. The phenomenon cannot be reduced to a stable definition, and cannot be adequately

captured by a single philosophical framework or the insights of a community of practice. It appears instead as an ongoing process, historically layered and pragmatically unstable, in which legitimacy, emotional acceptability, political value, and actual usage do not always coincide, and whose trajectories can be unpredictable. The questionnaire confirms this repeatedly: respondents may recognise reclamation in principle and still restrict its acceptable use; they may acknowledge its subversive force and still experience discomfort towards the terms involved; they may appeal to community membership and yet deny that the broader LGBT+ umbrella is sufficient in every case. This multi-faceted plurality, however, is one of the main findings of the study, because it shows that reclamation remains socially negotiated and internally contested even among those who are closest to it.

For the purposes of this thesis, this has an immediate consequence. Any attempt to describe reclamation must resist the temptation to settle too quickly on a single definition or on a single criterion of legitimacy. What the present case study makes visible is not only the breadth of the social and philosophical issues involved, but also the ethical difficulty of holding them together: community involvement, positionalities, methodological accuracy, and technical limits all matter, yet they do not always point in the same direction. A socially situated approach does not dissolve this tension, but suggests the need to keep it visible and to treat it as part of the object of study itself.

7.3.1 Methodological Lesson Learnt

From a methodological standpoint, several issues stem from the exploratory nature of this study. At the time of design, the research team had not yet defined a fully delimited set of hypotheses; therefore, a strategic decision was made to include a wide range of items, with the understanding that selection and refinement would occur in later stages. This choice generated a valuable breadth of data but also introduced certain imprecisions and redundancies that should be resolved in future studies.

Although the total number of participants (N = 279) appears before our eyes as a solid sample for an exploratory investigation, the absolute numbers remain relatively modest, whereas certain socio-demographic dimensions—such as geographical origin and residence, gender identity, age—are sparsely

represented. These asymmetries constrain the external validity of the findings and reflect the limits of the questionnaire's dissemination.

A further limitation—again connected with the dissemination of the survey—concerns the composition of the respondent pool, which appears to comprise predominantly individuals already familiar with or proximal to LGBTQ+-related themes. This probably reflects the socio-digital "bubble" through which the questionnaire circulated and may have shaped participants' familiarity with reclamation discourse as well as their comfort with the thematic content. In this respect, the limitation is evident in the results of the item assessing closeness to the LGBT+ community, but rather in the socio-cultural clustering of those who elected to participate.

One imprecision relates to the age item. While the survey employed conventional age brackets, these do not align with recognised generational cohorts such as Baby Boomers, Generation X, Millennials, and Generation Z and so on. Although generational boundaries are culturally constructed and not without limitations, aligning with cohort-based definitions—or simply requesting respondents' exact year of birth—would enable more analytically meaningful comparisons, particularly in light of digitalisation-related differences that shape exposure to linguistic practices across generations.

Another methodological challenge concerns the aforementioned five-point scale measuring self-reported closeness to the LGBT+ community. Such a scale is vulnerable to social desirability bias (Massara et al., 2012; Kwak, Holtkamp, and Kim, 2019; Kwak, Ma, and Kim, 2021) and may be too generic to capture the multifaceted nature of community proximity (e.g., relational, cultural, situational, or activist involvement). In the present study, this construct was kept analytically separate from gender identity and sexual/romantic orientation—a distinction that should be preserved—but future research may require a more nuanced operationalisation of this issue.

The overall length and density of the questionnaire further represents a significant limitation. As a pilot study, the instrument was intentionally comprehensive in order to test the feasibility and the scope of a future standardised tool. However, this design unavoidably increased respondent burden and may have negatively affected completion rates. Future studies should aim to streamline the instrument, possibly through shorter targeted questionnaires or sub-question structures.

Additional complexities result from the use of multiple-response items with optional "Other: ___" fields, as in Items 1A and 3B. While intended to preserve inclusivity and allow for self-definition, these formats generated substantial difficulties during coding and interpretation due to the heterogeneity of participants' additions. A more restrictive design may be warranted unless multiple selections or self-descriptions are demonstrably necessary for the construct being examined, that was not always the case in our study.

A critical issue emerged in relation to the items on gender identity and sexual/romantic orientation. Although open-ended options are consistent with best practices in inclusive demographic design (Hughes et al., 2022), they produced 11 cases in which responses could not be unequivocally classified into either ingroup or outgroup categories according to the adopted criteria. These participants were excluded from most analyses, although their data were presented descriptively for transparency.

Finally, the operationalisation of geographical context through population-size categories, while methodologically defensible, may not capture socially meaningful patterns as effectively as the rural–suburban–urban tripartition recommended by Hughes et al. (2022). Adopting such a framework could enhance comparability with international research and better reflect respondents' lived environmental contexts.

To summarise, in our post-analysis view future research should rely on wider dissemination channels, refine demographic measurements, reduce reliance on open-ended or multi-response items, adopt more modular and targeted survey designs, and integrate clearer procedural safeguards to support both analytical rigour and participant well-being.

Overall, these limitations underscore the challenges inherent in the attempt of balancing inclusivity, exploratory aims, ethical concerns, and methodological precision. At the same time, not all these difficulties should be read merely as shortcomings of the instrument. Some of them arise because the phenomenon itself resists neat operationalisation. The questionnaire became difficult to simplify precisely because reclamation is entangled with all the broader questions mentioned so far. In this sense, part of the methodological instability encountered here reflects the constitutive instability of the object under investigation.

7.3.2 Implications and Future Directions for NLP

Despite its limitations, this pilot study has yielded insights that are directly relevant to future NLP work on slur reclamation in Italian. Above all, it shows that reclamation cannot be treated as a property of the lexical item alone. The questionnaire data suggest instead that respondents rely on a more complex interpretive process, in which writer positioning, inferred community membership, relational proximity, and discourse function shape the final judgement. From this point of view, one main—so far very clear—implication of the study is that a future annotation schema for reclamatory language should not reduce the task to a binary decision on the lexical item itself, but should make part of that interpretive process more visible.

More concretely, the survey points to a number of dimensions that could be integrated into a more socially grounded annotation schema. One concerns **speaker or writer membership**: as the results of Part 1 and Part 2 have shown, respondents often decide whether a use is reclamatory only after inferring whether the writer is likely to belong to the target community. Another concerns **discourse function**: the comments collected for Item 2B repeatedly show that respondents distinguish between directly reclaiming a term and merely talking about its use. Input #1 is especially revealing in this respect, since many respondents interpreted it not as a reclaimed use proper, but as a statement about who may use the term and under which relational conditions. Taken together, these results suggest that a future dataset should retain at least some information about inferred membership and discourse context, as localised conditions that allow annotators to evaluate a use as legitimate or illegitimate.

The questionnaire also shows that **morphosyntactic and pragmatic configuration matter**. The comments on *frocia* in inputs #6 and #7 make this especially clear: adjectival uses are often harder to interpret than nominal self-labelling, because they may frame an event, a scenario, or a way of appearing rather than directly naming a person. In the same way, the results confirm that not every occurrence of a slur carries the same evaluative force, and that the reclamatory reading changes depending on pragmatics. Sentences #5 or #10 are read very differently from, respectively, #4 or #9, even though they contain the same slurs in similar discursive configurations. This means that a reclamation-sensitive annotation schema should be able to record how the term is

functioning locally, whether the utterance is self-directed or other-directed, and whether the surrounding stance supports self-affirmation, distance, or hostility. It would also be useful, especially in uncertain cases, to preserve annotator hesitation or disagreement rather than immediately collapsing it into a single majority label.

This last point is especially important in light of what the questionnaire reveals about the social life of reclamation. **The phenomenon is not only context-dependent**: it is also affectively charged and still in the making. The final comments show that even within the LGBT+ respondents there is no stable consensus about who may reclaim what and under which conditions. Some respondents stress intention above all; others privilege group membership; others again accept reclamation only in self-referential or strongly protected contexts. These contrasts are not noise to be eliminated. They indicate that reclamation **remains an ongoing process**, and that, again, its annotation cannot be fully separated from the social disagreements through which it is negotiated. For this reason, disagreement should be treated, in a perspectivist and community-reflective framework, as a signal that an utterance occupies a boundary zone between competing readings.

In practical terms, then, the main potential of the present approach lies in its ability to inform annotation before annotation begins. The questionnaire helps identify which cues ordinary interpreters actually rely on and where ambiguity tends to arise, thus signalling, case by case, **which distinctions should be made explicit in the annotation design**. Building on these results, a next step would be the development of a community-reflective annotated dataset specifically designed for reclamation in Italian. Such a resource should also include information about the context and all the local pragmatic cues that may shape interpretation. **Annotators should be selected so as to reflect different relationships to the LGBTQ+ community**, and **the guidelines should be refined through discussion of difficult cases**, especially those involving metadiscourse, adjectival uses, and context-dependent legitimacy. In this sense, the broader goals of FATA would be pursued more concretely, since community knowledge would shape the annotation task from the outset rather than being added only afterwards.

More generally, the present study suggests that the detection of reclamatory language is one of those cases in which socially situated annotation is

not an optional methodological refinement, but a basic condition for building computational resources that are both credible and fair. If future NLP systems are expected to distinguish harmful uses from community-internal or self-affirming ones, they will need training data that already encode this complexity in a principled way. The present study does not yet provide such a dataset, nor operationalised categories, but it does indicate more clearly what such a dataset should look like and which dimensions it should preserve.

At the same time, the study also suggests that no annotation design will ever fully stabilise the phenomenon once and for all. Reclamation remains tied to changing histories of use and to forms of social negotiation that cannot be entirely formalised. What a queer perspective within NLP can do is to document this instability better and organise it more transparently, with the aim of reducing, as much as possible, the risk that computational simplification entails social flattening or, even worse, reproduces existing structures of oppression in the algorithm.

Chapter 8

CONCLUSIONS

This thesis has investigated how a more socially situated, interdisciplinary, queer, and context-aware approach can contribute to the study of harmful language in NLP, with particular attention to gender stereotypes and LGBTQ+ slur reclamation. Across the conceptual and methodological chapters, the preliminary study on nomina agentis, and the two case studies, the central claim has been that the computational elaboration of socially sensitive language cannot be reduced to a purely technical operation. Instead, it must be understood as an interpretative and socially situated practice, shaped by context, speaker position, community norms, and the epistemic conditions under which judgements are produced.

In this final chapter, we bring together the main research lines of the thesis in order to answer the research questions outlined in Chapter 1, reflect on the principal limitations of the work, make explicit its ethical and positional dimensions, and summarise its main contributions, outputs, and future directions.

8.1 Answering the Research Questions

The work developed in this thesis was guided by four main research questions. They were intended to clarify whether social-science-informed approaches can contribute to NLP research and how such approaches may reshape annotation design when the object of analysis is socially sensitive and interpretatively unstable, as deeply embedded in relations of power. In the following subsections, we try to answer our research questions in order.

8.1.1 RQ1. How and to what extent can methods and practices informed by socio-linguistic and philosophical knowledge contribute to NLP research?

The thesis suggests that methods and practices informed by socio-linguistic and philosophical knowledge can contribute to NLP research at multiple and mutually reinforcing levels.

First, they provide renovated hermeneutical perspectives for understanding language as a socially situated practice rather than a decontextualised textual object. In this respect, Sociolinguistics, Queer Linguistics, and Philosophy of Language have made it possible to foreground definitions and dimensions such as positionality, identity, social norms, and interpretative asymmetries.

Second, social-science-informed approaches contribute methodologically. The use of focus groups, surveys, iterative annotation refinement, post-annotation questionnaires, and explicit attention to annotator feedback demonstrates that socially sensitive tasks require more than a top-down assignment of labels. Instead, they help in exploring processes that make room for participants' and annotators' own interpretative frameworks, thereby helping the researcher identify ambiguities, unstable categories, and blind spots that would otherwise remain implicit.

Third, the work shows that social-science-informed practices contribute epistemologically. They do not merely enrich NLP with "external" context, but help redefine what counts as relevant evidence in the first place. In this sense, the contribution of these socio-linguistic and philosophical traditions is not ancillary: it concerns the very framing of the task, the conditions of annotation, and, subsequently, the interpretability of the resulting categories. The FATA paradigm synthesises this contribution by proposing a re-orientation of the standard NLP pipeline in which community consultation and situated judgement are treated as foundational. This orientation is also close to the perspectivist line of work that has inspired the present thesis throughout, especially in its refusal to treat disagreement as mere noise and in its insistence that different judgements may reflect different but still socially grounded ways of reading the same utterance.

Crucially, the contribution is reciprocal: the computational lens has in turn

refined and stress-tested the socio-linguistic and philosophical questions at stake, so that the two traditions are treated throughout as equal partners and not one being placed at the service of the other.

8.1.2 RQ2. What are the possibilities and limits of adopting a queer perspective in a study aimed at developing an annotation schema?

The thesis suggests that adopting a queer perspective is both productive and demanding. Its main possibility lies in its capacity to unsettle fixed and normative assumptions about gender, sexuality, target groups, speaker legitimacy, and even the apparent stability of linguistic categories themselves. In this work, a queer perspective has been particularly useful in resisting reductive binary oppositions, in questioning the naturalisation of ingroup/outgroup distinctions, and in approaching slur reclamation and gendered nomination as phenomena whose meaning cannot be stabilised independently of context, identity, social positioning, legitimation, and conscious uptake.

At the same time, the thesis also makes visible the limits of such a perspective when it is brought into operational contexts such as annotation design. A queer-informed methodology cannot simply dissolve categories, because annotation work requires some degree of operationalisation, explicitness, and analytical delimitation. The challenge, then, is not to avoid categories altogether, but to consider and apply them reflexively and accountably, as well as provisionally. Therefore, the possibilities of a queer perspective in annotation reside in making a schema—and its categories—more self-aware, i.e., more attentive to marginalised voices, more cautious with normative assumptions, and more open to ambiguity, disagreement, complexity, sensitivity, and category instability. At the same time, its limits are found in the fact that empirical and computational work still requires groupings, label boundaries, complexity reduction, and clear-cut choices, which must be made explicit exactly because they remain contestable. Such contestability is not a residual inconvenience to be managed at the margins of the schema. It is often part of the phenomenon itself. This is especially clear in the case of reclamation, but the same pressure can also be seen in the treatment of stereotypes and in

the socially meaningful use of gendered forms: the closer one looks, the less stable categories appear.

8.1.3 RQ3. How can stereotypes and reclamation inform both Linguistics and NLP?

The work has treated stereotypes and reclamation as complementary sites through which broader questions about language and its interpretation become particularly visible. From a linguistic point of view, both phenomena show that meaning cannot be fully inferred from textual form alone. Stereotypes often operate through implicitness, presupposition, and generic claims that appear descriptive while carrying essentialising effects. Reclamation, by contrast, shows how a lexical form historically associated with derogation may be reused in ways that are, instead, identity-affirming or politically subversive, depending on complex pragmatic factors.

From an NLP point of view, both stereotypes and reclamation expose the limitations of models and datasets that rely too heavily on surface cues or assume that labels can be assigned independently of social interpretation. In both cases, the problem is not merely one of classification difficulty, but of annotation design: what is being annotated, under which assumptions, by whom, and for which purpose. The thesis therefore suggests that stereotypes and reclamation should be seen as phenomena that force a reconsideration of how socially sensitive language is operationalised for computational purposes.

More broadly, the parallel treatment of stereotypes and reclamation has allowed the thesis to argue that NLP-oriented research benefits when apparently different harmful-language phenomena are analysed in relation to shared methodological and epistemic issues.

8.1.4 RQ4. How and to what extent can gender stereotypes and LGBTQ+ slur reclamation be framed so as to inform the development of a fair, community-centred, and participatory annotation schema for automatic detection?

The central answer proposed by this thesis is that gender stereotypes and LGBTQ+ slur reclamation can inform annotation design only if they are framed not as fixed categories to be mechanically recognised, but as socially mediated and interpretatively demanding phenomena. In this sense, the case studies do not offer a single final schema ready for large-scale use; rather, they provide a principled basis for designing such schemas more responsibly.

The first case study has shown that an annotation schema on gender stereotypes in public discourse benefits from iterative development, explicit discussion with annotators, gradual refinement of categories, and post-hoc reflection on what worked and what remained unstable. More specifically, it suggests that such stereotypes are not always encountered when or to the extent in which they are expected. A multi-layered schema on this topic can help to recognise patterns and correlations among dimensions that may not be hypothesised before.

The second case study has shown that reclamation cannot be meaningfully operationalised without consulting participants directly about context, speaker identity, offensiveness, legitimacy, and self-reported linguistic practice. Moreover, it suggests that even the involvement of the community can be not enough if the idea that a social group is not homogeneous is not taken into consideration methodologically. In this respect, one of the clearest lessons of the thesis is that community involvement should not be imagined as a way of obtaining a single authoritative answer once and for all. What it can offer, rather, is a better grounded account of where disagreement lies and which distinctions are framed as relevant. The point is not to search for an idealised community consensus, but to make the construction of categories answerable to the social complexity from which they emerge.

Together, the two studies indicate that fairness in annotation is not only achieved through label balance or technical de-biasing, but also through methodological attention to the prior design of the task, the selection of categories, the treatment of disagreement, and, when possible, through the

inclusion of the viewpoints of those most affected by the phenomena under study.

Within this broader trajectory, the FATA paradigm offers the clearest synthesis of the thesis' methodological answer to RQ4. By placing contextualisation and participatory design earlier in the research pipeline, it provides a way of translating theoretical commitments into annotation practice.

8.2 Limitations

Several limitations of this work should be acknowledged. Some of them stem from the exploratory and interdisciplinary nature of the thesis, while others arise from the specific empirical designs adopted in the two case studies.

A first limitation concerns scope. This thesis is primarily conceptual and methodological in orientation. Its overall contribution resides in the design, critique, and rethinking of annotation processes rather than in the development of a final computational model. For this reason, the work should be read as laying the groundwork for future modelling and not as delivering a complete end-to-end computational solution.

A second limitation concerns data and representativeness. The preliminary study on nomina agentis is intentionally circumscribed and functions as a methodological bridge rather than as a fully autonomous corpus study. The Rackete-Schettino case study is tied to two specific events and to a historically situated discourse on a specific social media platform. The survey on slur reclamation, while robust for an exploratory investigation, remains limited by the composition of the respondent pool and the subsequent under-representation of some identities and territorial areas.

A third limitation concerns annotation itself. The thesis repeatedly argues that disagreement is meaningful, yet some phases of the empirical work still rely on procedures such as majority voting or category consolidation. These choices are methodologically justified, but they also involve a reduction of plurality. More generally, several dimensions in both case studies remain difficult to operationalise cleanly, especially when they depend on missing contextual cues. This difficulty is not merely technical. It also reflects a more basic mismatch between socially dense phenomena and the formats through which computational work often needs to render them manageable. When

context is sparse and texts are pragmatically demanding, the pressure to simplify becomes methodologically necessary and at the same time analytically risky. However, in this sense, some of the technical limits encountered in the thesis are inseparable from the social complexity of the object itself.

A fourth limitation concerns transferability. Although the methodological proposals advanced here are intended to be relevant beyond Italian, the empirical core of the thesis is strongly situated in the Italian context, both linguistically and socio-culturally. Future work will therefore need to test whether the same conceptual and methodological architecture can be extended to other languages and other communities, without flattening their specificities.

Finally, there is a productive tension running through the whole thesis that should itself be recognised as a limit: the work critiques rigid normative categories while also needing to use categories in order to conduct empirical and computationally oriented research. Far from fully resolving this tension, instead, the thesis has tried to make it visible and methodologically manageable.

8.3 Ethical Considerations and Positionality Statement

This thesis deals with socially sensitive materials, including derogatory expressions and potentially harmful forms of discourse. It also involves participants' and annotators' judgements on identity-related and politically charged issues. For these reasons, ethical reflection has been treated as an integral dimension of the research design and not as an external requirement.

At the level of data and methods, the work has sought to minimise harm through anonymisation, explicit warnings where needed, careful attention to informed consent, and a constant effort to avoid presenting communities as internally homogeneous. At the same time, the thesis has treated disagreement and uncertainty as ethically and epistemically significant aspects of research on harmful language.

A further ethical issue concerns the aforementioned reduction of plurality. As also noted in a recent work on annotation of hate speech toward

LGBTQIA+ people by Damo et al. (2025), any evaluation procedure based on a harmonised or aggregated gold standard risks obscuring the subjectivity of annotator perspectives. This thesis has repeatedly tried to confront this problem by preserving disaggregated viewpoints where possible, discussing disagreement explicitly, maintaining an ongoing dialogue with the annotators, and incorporating post-annotation and post-survey reflections into the analysis. This has also meant accepting that ethical adequacy can limit the production of a fully harmonised account of the phenomenon. In research of this kind, a major difficulty lies in an essential necessity: to decide how to keep different positions in view without pretending that they naturally converge and without reducing them too quickly to a single evaluative outcome. The issue is not only epistemic. It also concerns responsibility towards the people whose experience or discomfort makes the phenomenon intelligible in the first place.

8.3.1 Positionality Statement

As already stated in the *Preliminary Remark*, and as suggested throughout the thesis by the argumentative structure itself, this dissertation is written from within an interdisciplinary research trajectory situated at the intersection of Linguistics—including Sociolinguistics and Applied Linguistics—Digital Humanities, and NLP. My academic background has shaped the conceptual and methodological orientation of the work, especially in its emphasis on language as social practice, on the interpretative nature of annotation, and on the need for dialogue between computational and social-scientific approaches.

At the same time, my positionality is also shaped by my own social location. During the years of the PhD, I was in the 30–34 age range. I identify as a cisgender gay man. I grew up in a small town of around 20,000 inhabitants in the province of Salerno, in Southern Italy, and moved away at the age of 22. I currently live in Turin. This biographical trajectory is also significant because, as a Southern Italian living in Northern Italy, I am situated within a social experience historically shaped by internal migration, due to socio-economic asymmetries between different areas of the country, and still affected by stereotypes, prejudices, discrimination, and hate speech. I have also lived for shorter periods in several European countries and in Argentina. These elements are relevant because they have informed the kinds of questions I

ask, the aspects of harmful language to which I am more attentive, and, more in general, the ways in which I read the relation between language, contexts, power dynamics, and social experience.

Beyond its academic dimension, this thesis has also been shaped by forms of engagement developed outside the university context, including previous and ongoing experience in activism. In particular, this background includes involvement in projects and events promoted by the *Council of Europe*, especially through its Youth Department; continuous collaboration with *No Hate Speech Movement Italia* as a member of the digital activist group; participation in Erasmus+ non-formal learning activities, also as a facilitator; collaboration with *A.P.I.C.E.*, a no-profit organisation based in South Italy that promotes youth participation and human-rights-based approaches to the contrast of hate speech and discrimination; and activism within the LGBTQ+ association *Maurice* in Turin. These experiences are relevant sources of situated knowledge, as they have contributed to shaping my perspective and consolidating my awareness of how harmful language, hate speech, stereotypes, discriminatory discourse, and reclamation function in real-world settings.

My relation to the topic is also grounded in personal experience. I have myself experienced homophobia. At the same time, in amicable contexts, I do sometimes use slurs to refer to myself, to my partner, and to LGBTQ+ friends. This does not mean that I treat reclamation as self-evident, uniform, or unproblematic; rather, it makes especially clear to me that reclamatory practices are context-dependent and unevenly distributed even within the same broad community.

For this reason, the aim of making my positionality explicit is not to claim any automatic epistemic privilege, since whatever epistemic advantage may arise from a given social position must be understood as contingent rather than guaranteed (Rolin, 2009). Nor is it to speak on behalf of LGBTQ+ people as a homogeneous whole. On the contrary, to recall Edmondson (2021), one of the central ideas of this thesis is precisely that no community is internally uniform, that interpretative authority is always partial and negotiated, and that the researcher's own assumptions must be visible, in order to become more accountable and revisable. My positionality is therefore presented as one of the conditions that shape the production of knowledge in this dissertation.

Methodologically, this positionality has had several consequences for the

design of the research. First, it has reinforced the decision to treat participants and annotators as situated knowers whose feedback, emotions, choices, and suggestions help determine what becomes visible and analytically relevant. Second, it has motivated the use of specific sociolinguistic techniques and iterative practices, and accompanied their implementation, especially in terms of group facilitation and the creation of a *safe(r) space*. Third, it has encouraged a cautious and reflexive attitude towards categories such as gender, identity, social group, legitimacy, and reclamation. In this sense, positionality in the present work is not intended as an autobiographical addition, but as a methodological commitment to the whole research.

8.4 Research Contributions

The contributions of this thesis can be summarised at four levels: conceptual, methodological, empirical, and resource-oriented.

8.4.1 Conceptual contributions

At a conceptual level, the thesis has argued for a socially situated approach to harmful language annotation in NLP. It has shown that annotation should be understood not as the extraction of stable textual properties, but as a socially mediated interpretative practice in which meaning emerges through the interaction of semantics, pragmatics, contextual assumptions, positionality, and community norms. In doing so, it has offered an interdisciplinary synthesis connecting Sociolinguistics, Queer Linguistics, and Philosophy of Language to contemporary debates in socially aware NLP.

8.4.2 Methodological contributions

At a methodological level, the thesis has proposed and operationally explored a reorientation of annotation design through the FATA paradigm. This contribution lies in showing how community consultation, reflexive design, and participant-centred methods can be integrated into the earlier phases of NLP research. The thesis has also foregrounded the methodological value of disagreement, post-annotation reflection, and iterative category revision in the construction of socially sensitive datasets.

8.4.3 Empirical contributions

At an empirical level, the thesis has produced three connected case-based investigations. The preliminary study on Italian nomina agentis has shown how gendered lexical and morphological choices can carry socially meaningful evaluative effects and metalinguistic reflections, and can therefore inform annotation design. The first case study has documented the iterative construction, testing, revision, and analysis of an annotation schema for gender stereotypes in public discourse, using the Rackete-Schettino corpus as a (historically) situated case. The second case study has provided an exploratory but methodologically rich investigation of LGBTQ+ slur reclamation in Italian, combining a preliminary focus group, a sociolinguistic survey, and multi-layered interpretative analysis.

Importantly, these empirical contributions do not only inform NLP. They also return something to the social sciences from which much of the thesis has drawn its conceptual and methodological orientation, by offering new empirical materials, although necessarily limited, and, hopefully, further points of reflection on the social life of stereotypes, gendered nomination, and reclamatory language.

8.5 Resources Produced

Beyond its argumentative contributions, this thesis has also produced a set of methodological and empirical resources that may be valuable for future research.

- a conceptual and methodological framework for socially situated annotation design, for which the thesis itself represents an application of the FATA paradigm, even if partial;
- the preliminary annotation framework developed for the study of Italian nomina agentis;
- the multi-layer annotation schema and guideline versions developed for the Rackete-Schettino corpus;
- the post-annotation questionnaire used to collect annotators' feedback on the first case study;

- the preliminary focus group design and the full sociolinguistic survey instrument used for the second case study on slur reclamation;
- the organised outputs derived from the analysis of the survey on slur reclamation, including response groupings, analytical distinctions, and coding structures for the interpretation of the results.

8.6 Relevant Publications

Parts of this thesis have developed in dialogue with publications, workshop papers, conference presentations, and other dissemination activities produced during the Ph.D. In what follows, these outputs are organised into three groups: publications directly connected to the dissertation, other related publications that belong to the broader research trajectory, and presentations or posters through which parts of the work were discussed in academic contexts or disseminated in schools.

8.6.1 Publications directly connected to the thesis

- Marra, Andrea and Cristina Bosco (2024). "Orsù dunque... avvocato? Osservazione del maschile sovraesteso nei nomina agentis su Twitter". In: *Lingua inclusiva: forme, funzioni, atteggiamenti e percezioni*. Ed. by Anna-Maria De Cesare and Giuliana Giusti. Vol. 6. LiVVaL. Venezia: Edizioni Ca' Foscari – Venice University Press, pp. 203–248. DOI: 10.30687/978-88-6969-866-8/008. URL: `http://edizionicafoscari.it/it/edizioni4/libri/978-88-6969-867-5/orsu-dunque-avvocato/`
- Ferrando, Chiara, Lia Draetta, Andrea Marra, Angela Zottola, Cristina Bosco, and Viviana Patti (2025). "AEQUITAS 2025 Fairness and Bias in AI". In: *Proceedings of the 3rd Workshop on Fairness and Bias in AI co-located with the 28th European Conference on Artificial Intelligence (ECAI 2025)*. Vol. 4147. Workshop AEQUITAS 2025, October 26th, 2025. Bologna, Italy: CEUR Workshop Proceedings, pp. 1–12. URL: `https://ceur-ws.org/Vol-4147/`

- Marra, Andrea, Chiara Ferrando, Lia Draetta, Bianca Cepollaro, and Viviana Patti (2026). "How is the reclamation of slurs perceived in Italian?: A sociolinguistic survey to inform future NLP studies". In: *Linguistik online* 144.3, pp. 209–225. URL: https://bop.unibe.ch/linguistik-online/article/view/13550

8.6.2 Other related publications

- Rosola, Martina, Simona Frenda, Alessandra Teresa Cignarella, Matteo Pellegrini, Andrea Marra, and Mara Floris (2023). "Beyond Obscuration and Visibility: Thoughts on the Different Strategies of Gender-Fair Language in Italian". In: *Proceedings of the Ninth Italian Conference on Computational Linguistics CLiC-it 2023*. Venezia, 30 novembre–2 dicembre 2023. Torino: Accademia University Press, pp. 370–379. URL: https://ceur-ws.org/Vol-3596/

- Schmeisser-Nieto, Wolfgang S., Alessandra Teresa Cignarella, Tom Bourgeade, Simona Frenda, Alejandro Ariza-Casabona, Mario Laurent, Paolo Giovanni Cicirelli, Andrea Marra, Giuseppe Corbelli, Farah Benamara, Cristina Bosco, Véronique Moriceau, Marinella Paciello, Viviana Patti, Mariona Taulé, and Francesca D'Errico (2024). "StereoHoax: a multilingual corpus of racial hoaxes and social media reactions annotated for stereotypes". In: *Language Resources and Evaluation*, pp. 1–39. URL: https://doi.org/10.1007/s10579-024-09791-3

- Cignarella, Alessandra Teresa, Manuela Sanguinetti, Simona Frenda, Andrea Marra, Cristina Bosco, and Valerio Basile (2024). "QUEEREOTYPES: A Multi-Source Italian Corpus of Stereotypes towards LGBTQIA+ Community Members". In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 13429–13441. URL: https://aclanthology.org/2024.lrec-main.1176

- Cignarella, Alessandra Teresa, Elisa Chierchiello, Chiara Ferrando, Simona Frenda, Soda Marem Lo, and Andrea Marra (2024). "From Hate Speech to Societal Empowerment: A Pedagogical Journey Through Computational Thinking and NLP for High School Students". In:

Proceedings of the Sixth Workshop on Teaching NLP, pp. 54–65. URL: https://aclanthology.org/2024.teachingnlp-1.7

8.6.3 Presentations and dissemination activities

- *EPITHETS & STAL-2025 Workshop*, Università degli Studi di Genova, Italy, 7–8 May 2025. Poster presentation: "The Pragmatic Shift of Slurs: A Preliminary Linguistic Analysis of Reclamation in Digital Spaces".
- *Frontiere e confini della ricerca: saperi e metodologie nello spazio digitale e tecnologico* (1st Doctoral Conference of the University of Turin), Università degli Studi di Torino, Italy, 9–11 December 2024. Oral presentation: "Participating communities: a proposal for a fair methodological approach in NLP".
- *Gender-inclusive language in a multilingual Europe. Institutional policies, their applications and AI-related developments (LaGendA)*, Università Ca' Foscari Venezia, Italy, 3–4 October 2024. Poster presentation: "Sociolinguistics informs Natural Language Processing: the semantic reappropriation of slurs in the Italian LGBT+ community".
- *2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (LREC-COLING 2024), Università degli Studi di Torino, Italy, 20–25 May 2024. Poster presentation: "QUEEREOTYPES: A Multi-Source Italian Corpus of Stereotypes towards LGBTQIA+ Community Members".
- *9th Italian Conference on Computational Linguistics* (CLiC-it 2023), Università Ca' Foscari Venezia, Italy, 30 November–2 December 2023. Poster presentation: "Beyond Obscuration and Visibility: Thoughts on the Different Strategies of Gender-Fair Language in Italian".
- *Lectures on Computational Linguistics* (LCL 2023), Università di Pisa, Italy, 29–31 May 2023. Poster presentation: "Exploring gender stereotypes: an interdisciplinary proposal to study gender-related stereotypical representation in language".
- *Laboratorio #DeactivHate* Turin, Italy, 2023–2026. I conducted several editions of the #DeactivHate workshop, promoted by the Department

of Computer Science at the University of Turin and delivered in upper secondary schools. The activities were designed to foster critical awareness of the digital technologies students use on a daily basis, with the aim of addressing online hate through the introduction of basic methods and practices drawn from Computational Linguistics and Artificial Intelligence.

8.7 Future Work and Final Remarks

The work developed in this thesis opens several directions for future research.

A first line of development concerns annotation itself. Future work could further test the schemas and procedures proposed here on new corpora and tasks, with a stronger integration of disaggregated annotations, soft labels, text genres, and contexts. This would make it possible to preserve interpretative plurality more systematically and to evaluate computational models against richer representational targets.

A second line concerns computational modelling. The thesis has repeatedly argued that stereotypes, reclamation, and socially meaningful gendered nomination are not well captured by surface-based approaches alone. Future work could therefore explore models that incorporate richer contextual information, speaker-related cues, explanation-based prompting, or forms of metadata-aware and perspectivist learning.

A third line concerns expansion across phenomena, languages, and communities. The methodological architecture proposed here should be tested on additional forms of harmful language, on other target groups, and in languages beyond Italian. Such extensions would help determine which aspects of the present work are language-specific and which may generalise as part of a broader framework for socially situated NLP.

A fourth line concerns participation. The FATA paradigm has been advanced here as a principled orientation and partially operationalised through the two case studies. Future work could expand this orientation by involving affected communities even more directly in task definition, schema revision, interpretative validation, and dataset documentation. This would also help address one of the central tensions highlighted in the thesis: the need to balance computational operationalisation with social and interpretative complexity.

It is worth adding that this trajectory is itself still in progress. The broader work currently developing around *FAVOLOSA*, together with the attempt to apply FATA more systematically across tasks and resources, should be understood as an ongoing process rather than as a completed methodological solution. In this respect, it is also significant that some of the researchers who contributed to the present thesis have recently been involved in the organisation of *MultiPRIDE*, a shared task at EVALITA 2026 devoted to the multilingual automatic detection of slur reclamation in LGBTQ+ contexts. The task addressed both text-only detection and contextual detection with user biographies, across Italian, Spanish, and English, and therefore brings into a shared evaluation setting some of the same questions explored in this thesis, especially those concerning context, membership, legitimacy, and the need to of model reclamation across different linguistic and cultural settings.

Future work will also need to confront a more general issue that has emerged throughout the thesis, and especially in the study of reclamation: some socially sensitive phenomena remain open-ended in a strong sense. Their meanings shift, their legitimacy is renegotiated, and even within the relevant communities there may be no stable agreement about acceptable use. For this reason, a better methodology has to document such instability more carefully, specify where it matters, and prevent simplified analytical decisions from being treated as if they fully captured the phenomenon under investigation.

In conclusion, the thesis has argued that the study of harmful language in NLP cannot be reduced to a problem of recognising pre-existing labels in text. It is, at the same time, a problem of deciding how such labels come into being, whose perspectives shape them, which forms of disagreement are preserved or suppressed, and what kinds of social reality are reproduced through annotation practices. What is at stake, then, is not only the descriptive adequacy of labels, but also the risk that computational simplification may flatten social difference or reproduce, at the algorithmic level, existing structures of oppression. If there is one broader lesson that cuts across the whole thesis, it is that socially sensitive language does not become clearer simply because it is formalised. On the contrary, formalisation often brings into view the very tensions that ordinary categorisation tends to hide: between use and mention, between recognition and acceptance, between political value and affective

discomfort, between the need to annotate and the impossibility of framing exhaustively what a community may mean by a term. In summary, a socially situated approach is valuable not because it promises to resolve these tensions once and for all, but because it makes it harder to mistake simplification for understanding.

From this standpoint—and accordingly with the FATA paradigm, of which this work represents a first attempt of consistent application—the main contribution of the present thesis resides in making clear that more reflexive, context-aware, community-centred, and socially grounded paths for annotation design are both necessary and possible.

Therefore, aware of the challenges these paths pose, we can only try to understand them, or at least position ourselves on a different site, from which they may be seen, then observed, from ever-new points of view.

Bibliography

Abbatecola, Emanuela (2016). "Sessismo a parole". In: *Genere e linguaggio. I segni dell'uguaglianza e della diversità*. Ed. by Fabio Corbisiero, Pietro Maturi, and Elisabetta Ruspini. Milano: FrancoAngeli, pp. 138–158. ISBN: 9788891727183. URL: https://boa.unimib.it/handle/10281/102554.

Abbott, Andrew Delano (2004). *Methods of discovery: Heuristics for the social sciences*. W. W. Norton & Company.

Adamo, Sergia, Elisabetta Tigani Sava, and Giulia Zanfabro (2019). *Non esiste solo il maschile. Teorie e pratiche per un linguaggio non discriminatorio da un punto di vista di genere*. EUT Edizioni Università di Trieste.

Anderson, Luvell and Michael Randall Barnes (2022). "Hate Speech". In: *Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. The Metaphysics Research Lab, Philosophy Department, Stanford University. URL: https://plato.stanford.edu/.

Azzalini, Monia and Giuliana Giusti (2019). "Lingua e genere fra grammatica e cultura". In: *Economia della cultura* 29.4, pp. 537–546.

Baker, Paul (2008). *Sexed texts: Language, gender and sexuality*. University of Toronto Press.

Baldo, Michela, Fabio Corbisiero, and Pietro Maturi (2016). "Ricostruire il genere attraverso il linguaggio: per un uso della lingua (italiana) non sessista e non omotransfobico". In: *gender/sexuality/italy* 3, pp. XII–XVII.

Banaji, Mahzarin R and Curtis D Hardin (1996). "Automatic stereotyping". In: *Psychological science* 7.3, pp. 136–141.

Basile, Pierpaolo, Annalina Caputo, Tommaso Caselli, Pierluigi Cassotti, and Rossella Varvara (2020). "A diachronic Italian corpus based on "L'Unita"". In: *Italian Conference on Computational Linguistics 2020*. CEUR Workshop Proceedings (CEUR-WS. org).

Basile, Valerio, Mirko Lai, and Manuela Sanguinetti (2019). "Long-term Social Media Data Collection at the University of Turin". In: *CEUR WORKSHOP PROCEEDINGS*. Vol. 2253. Accademia University Press, pp. 40–45.

Bassignana, Elisa, Valerio Basile, and Viviana Patti (2018). "Hurtlex: A multilingual lexicon of words to hurt". In: *CEUR Workshop proceedings*. Vol. 2253. CEUR-WS, pp. 1–6.

Bazzanella, Carla, Orsola Fornara, and Manuela Manera (2006). "Indicatori linguistici e stereotipi al femminile". In: *Linguaggio e genere. Grammatica e usi*. Ed. by Silvia Luraghi and Anna Olita. Vol. 1. Contributo in volume — IRIS UniTO handle 2318/11063. Roma: Carocci, pp. 155–169. URL: https://hdl.handle.net/2318/11063.

Berruto, Gaetano (2004). *Prima lezione di sociolinguistica*. Laterza.

Bianchi, Claudia (2014). "Slurs and appropriation: An echoic account". In: *Journal of Pragmatics* 66, pp. 35–44.

— (2015). "Il lato oscuro delle parole: epiteti denigratori e riappropriazione". In: *Sistemi intelligenti* 27.2, pp. 285–302.

Bianchi, Mauro, Andrea Carnaghi, Fabio Fasoli, Patrice Rusconi, and Carlo Fantoni (2024). "From self to ingroup reclaiming of homophobic epithets: A replication and extension of Galinsky et al.'s (2013) model of reappropriation". In: *Journal of Experimental Social Psychology* 111, pp. 1–8.

Birhane, Abeba, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed (2022). "Power to the people? Opportunities and challenges for participatory AI". In: *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–8.

Bolukbasi, Tolga, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai (2016). "Man is to computer programmer as woman is to homemaker? debiasing word embeddings". In: *Advances in neural information processing systems* 29.

Bosco, Cristina, Felice Dell'Orletta, Fabio Poletto, Manuela Sanguinetti, and Maurizio Tesconi (2018). "Overview of the evalita 2018 hate speech detection task". In: *Ceur workshop proceedings*. Vol. 2263. CEUR, pp. 1–9.

Bosco, Cristina, Viviana Patti, Simona Frenda, Alessandra Teresa Cignarella, Marinella Paciello, and Francesca D'Errico (2023). "Detecting racial stereotypes: An Italian social media corpus where psychology meets NLP". In: *Information Processing & Management* 60.1, p. 103118.

Brambilla, Marina, Valentina Crestani, et al. (2020). "Il genere nelle denominazioni di persona: grammatiche pedagogiche dell'italiano e del tedesco". In: *Italiano LinguaDue* 12.1, pp. 210–242.

Brontsema, Robin (2004). "A queer revolution: Reconceptualizing the debate over linguistic reclamation". In: *Colorado Research in Linguistics*.

Bucholtz, Mary and Kira Hall (2004). "Theorizing identity in language and sexuality research". In: *Language in society* 33.4, pp. 469–515.

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

Cameron, Deborah (2012). *Verbal hygiene*. Routledge.

— (2014). "Gender and language ideologies". In: *The handbook of language, gender, and sexuality*, pp. 279–296.

Cameron, Deborah and Don Kulick (2003). *Language and sexuality*. Cambridge University Press.

Caselli, Tommaso, Roberto Cibin, Costanza Conforti, Enrique Encinas, and Maurizio Teli (2021). "Guiding principles for participatory design-inspired natural language processing". In: *Proceedings of the 1st Workshop on NLP for Positive Impact*. Association for Computational Linguistics (ACL), pp. 27–35.

Casola, Silvia, Simona Frenda, Soda Marem Lo, Erhan Sezerer, Antonio Uva, Valerio Basile, Cristina Bosco, Alessandro Pedrani, Chiara Rubagotti, and Viviana Patti (2024). "MultiPICo: Multilingual perspectivist irony corpus". In: *PROCEEDINGS OF THE CONFERENCE-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. MEETING*. Vol. 1. Association for Computational Linguistics (ACL), pp. 16008–16021.

Cassotti, Pierluigi, Andrea Iovine, Pierpaolo Basile, Marco De Gemmis, and Giovanni Semeraro (2021). "Emerging trends in gender-specific occupational titles in italian newspapers". In: *Proceedings of the Eighth Italian Conference on Computational Linguistics (CLiC-it 2021)*, pp. 368–373.

Cavagnoli, Stefania (2013). *Linguaggio giuridico e lingua di genere: una simbiosi possibile*. Edizioni dell'Orso.

Cella, Federico and Martina Rosola (2024). "Fuorvianti e resistenti: i generici tra asimmetria inferenziale, scivolosità ed essenzialismo sociale". In: *Rivista di filosofia* 115.2, pp. 341–360.

Cepollaro, Bianca and Paolo Labinaz (2019). "Social network e comunicazione". In: *SISTEMI INTELLIGENTI* 2019.3, pp. 385–632.

Cepollaro, Bianca and Dan López de Sa (2022). "Who reclaims slurs?" In: *Pacific Philosophical Quarterly* 103.3, pp. 606–619.

— (2023). "The successes of reclamation". In: *Synthese* 202.6, p. 205.

Chiril, Patricia, Farah Benamara, and Véronique Moriceau (Nov. 2021). ""Be nice to your wife! The restaurants are closed": Can Gender Stereotype Detection Improve Sexism Classification?" In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 2833–2844. DOI: 10.18653/v1/2021.findings-emnlp.242. URL: https://aclanthology.org/2021.findings-emnlp.242.

Cignarella, Alessandra Teresa, Anastasia Giachanou, and Els Lefever (2025). "A Survey on Stereotype Detection in Natural Language Processing". In: *ACM Computing Surveys* 58.5, pp. 1–33.

Cignarella, Alessandra Teresa, Mirko Lai, Andrea Marra, and Manuela Sanguinetti (2021). "" La ministro è incinta": A Twitter Account of Women's Job Titles in Italian". In: *Proceedings of the 8th Italian Conference on Computational Linguistics (CLiC-it 2021)*. Vol. 3033. CEUR-WS. org, pp. 1–7.

Cignarella, Alessandra Teresa, Manuela Sanguinetti, Simona Frenda, Andrea Marra, Cristina Bosco, and Valerio Basile (2024). "QUEEREOTYPES: A Multi-Source Italian Corpus of Stereotypes towards LGBTQIA+ Community Members". In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 13429–13441.

Collins, Patricia Hill (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge.

Costa-jussà, Marta R (2019). "An analysis of gender bias studies in natural language processing". In: *Nature Machine Intelligence* 1.11, pp. 495–496.

Council of Europe (2016). *Combating Sexist Hate Speech*. Accesed: June 6, 2025. URL: https://rm.coe.int/1680651592.

Crenshaw, Kimberlé (1989). "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics". In: *The University of Chicago Legal Forum* 140, pp. 139–167.

— (1991). "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color". In: *Stanford Law Review* 43.6, pp. 1241–1299. ISSN: 00389765. URL: http://www.jstor.org/stable/1229039 (visited on 02/01/2025).

Cuccarini, Marco, Lia Draetta, Chiara Ferrando, Liam James, and Viviana Patti (2024). "ReCLAIM Project: Exploring Italian Slurs Reappropriation with Large Language Models". In: *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*. CEUR.

Cuddy, Amy JC, Susan T Fiske, and Peter Glick (2008). "Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map". In: *Advances in experimental social psychology* 40, pp. 61–149.

Damo, Greta, Alessandra Teresa Cignarella, Tommaso Caselli, Viviana Patti, and Debora Nozza (2025). "HODIAT: A Dataset for Detecting Homotransphobic Hate Speech in Italian with Aggressiveness and Target Annotation". In: *Proceedings of the The 9th Workshop on Online Abuse and Harms (WOAH)*, pp. 124–135.

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 Conference on Web and Social Media*. AAAI, pp. 512–515.

Davis, Kathy (2008). "Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful". In: *Feminist theory* 9.1, pp. 67–85.

De Mauro, Tullio (1982). *Minisemantica dei linguaggi non verbali e delle lingue*. Saggi tascabili Laterza (87). Prima edizione. Roma-Bari: Laterza.

Doleschal, Ursula (2009). "Linee guida e uguaglianza linguistica". In: *Mi fai male*, pp. 135–148.

Eckert, Penelope and Keith Brown (2006). "Communities of practice". In: *Concise encyclopedia of pragmatics*, pp. 109–112.

Eckert, Penelope and Sally McConnell-Ginet (1992). "Think practically and look locally: Language and gender as community-based practice". In: *Annual review of anthropology*, pp. 461–490.

— (2007). "Putting communities of practice in their place". In: *Gender and language* 1.1, pp. 27–37.

Eckert, Penelope and Etienne Wenger (2005). "Communities of practice in sociolinguistics." In: *Journal of sociolinguistics* 9.4.

Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A Bail (2020). "Computational social science and sociology". In: *Annual review of sociology* 46.1, pp. 61–81.

Edmondson, Daniel (2021). "Word norms and measures of linguistic reclamation for LGBTQ+ slurs". In: *Pragmatics & Cognition* 28.1, pp. 193–221.

Esposito, Eleonora, Carolina Pérez-Arredondo, and Angela Zottola (2024). "Intersecting inequalities: towards a critical discursive approach". In: *Journal of Gender Studies*, pp. 1–10.

Ferrando, Chiara, Lia Draetta, Andrea Marra, Angela Zottola, Cristina Bosco, and Viviana Patti (2025). "AEQUITAS 2025 Fairness and Bias in AI". In: *Proceedings of the 3rd Workshop on Fairness and Bias in AI co-located with the 28th European Conference on Artificial Intelligence (ECAI 2025)*. Vol. 4147. Workshop AEQUITAS 2025, October 26th, 2025. Bologna, Italy: CEUR Workshop Proceedings, pp. 1–12. URL: https://ceur-ws.org/Vol-4147/paper7.pdf.

Ferrando, Chiara, Marco Madeddu, Beatrice Antola, Sveva Silvia Pasini, Giulia Telari, Mirko Lai, and Viviana Patti (2024). "Exploring YouTube Comments Reacting to Femicide News in Italian". In: *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*. CEUR.

Field, Anjalie, Su Lin Blodgett, Zeerak Waseem, and Yulia Tsvetkov (2021). "A survey of race, racism, and anti-racism in NLP". In: *arXiv preprint arXiv:2106.11410*.

Fiske, Susan T (1998). "Stereotyping, prejudice, and discrimination." In: *The handbook of social psychology*.

Fiske, Susan T, Amy JC Cuddy, Peter Glick, and Jun Xu (2002). "A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition." In: *Journal of personality and social psychology* 82.6, p. 878.

Formato, Federica (2016). "Linguistic markers of sexism in the Italian media: a case study of ministra and ministro". In: *Corpora* 11.3, pp. 371–399.

— (2017). "Ci sono troie in giro in Parlamento che farebbero di tutto". In: *Gender and Language* 11.3, pp. 389–414.

— (2019). *Gender, Discourse and Ideology in Italian*. Palgrave Macmillan.

Fortuna, Paula and Sérgio Nunes (2018). "A survey on automatic detection of hate speech in text". In: *Acm Computing Surveys (Csur)* 51.4, pp. 1–30.

Francesconi, Chiara, Cristina Bosco, Fabio Poletto, and Manuela Sanguinetti (2019). "Error analysis in a hate speech detection task: The case of HaSpeeDe-TW at EVALITA 2018". In: *CEUR WORKSHOP PROCEEDINGS*. Vol. 2481. CEUR-WS, pp. 1–6.

Frenda, Simona, Gavin Abercrombie, Valerio Basile, Alessandro Pedrani, Raffaella Panizzon, Alessandra Teresa Cignarella, Cristina Marco, and Davide Bernardi (2024). "Perspectivist approaches to natural language processing: a survey". In: *Language Resources and Evaluation*, pp. 1–28.

Frenda, Simona, Alessandra Teresa Cignarella, Valerio Basile, Cristina Bosco, Viviana Patti, and Paolo Rosso (2022). "The unbearable hurtfulness of sarcasm". In: *Expert Systems with Applications* 193, p. 116398.

Frisina, Annalisa (2010). *Focus group. Una guida pratica*. Collana Itinerari. Il mulino.

Fumagalli, Corrado and Martina Rosola (2024). "The case for a duty to use gender-fair language in democratic representation". In: *The Philosophical Quarterly* 74.4, pp. 1159–1181.

Fusco, Fabiana (2019). "Il genere femminile tra norma e uso nella lingua italiana: qualche riflessione". In: *Non esiste solo il maschile. Teorie e pratiche per un linguaggio non discriminatorio da un punto di vista di genere*. Ed. by Sergia Adamo, Elisabetta Tigani Sava, and Giulia Zanfabro. Trieste: EUT Edizioni Università di Trieste, pp. 27–50. URL: http://hdl.handle.net/10077/30982.

Galeandro, Simona (2021). "Femminilizzazione versus neutralizzazione". In: *Testo e Senso* 23, pp. 65–73.

Galinsky, Adam D, Cynthia 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". In: *Psychological science* 24.10, pp. 2020–2029.

García-Sánchez, Rubén, Carmen Almendros, Begoña Aramayona, María Jesús Martín, María Soria-Oliver, Jorge S López, and José Manuel Martínez (2019). "Are sexist attitudes and gender stereotypes linked? A critical feminist approach with a Spanish sample". In: *Frontiers in psychology* 10, p. 2410.

Giusti, Giuliana (2022). "Inclusività nella lingua italiana: come e perché. Fondamenti teorici e proposte operative". In: *DEP. Deportate, esuli, profughe* 48.1, pp. 1–19.

Glick, Peter and Susan T Fiske (1997). "Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women". In: *Psychology of women quarterly* 21.1, pp. 119–135.

Grieve, Jack, Sara Bartl, Matteo Fuoli, Jason Grafmiller, Weihang Huang, Alejandro Jawerbaum, Akira Murakami, Marcus Perlman, Dana Roemling, and Bodo Winter (2025). "The sociolinguistic foundations of language modeling". In: *Frontiers in Artificial Intelligence* 7, p. 1472411.

Guillén-Pacho, Ibai, Arianna Longo, Marco Antonio Stranisci, Viviana Patti, and Carlos Badenes-Olmedo (2024). "The Vulnerable Identities Recognition Corpus (VIRC) for Hate Speech Analysis". In: *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*. CEUR.

Gygax, Pascal, Ute Gabriel, Oriane Sarrasin, Jane Oakhill, and Alan Garnham (2008). "Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men". In: *Language and cognitive processes* 23.3, pp. 464–485.

Gygax, Pascal Mark, Daniel Elmiger, Sandrine Zufferey, Alan Garnham, Sabine Sczesny, Lisa Von Stockhausen, Friederike Braun, and Jane Oakhill (2019). "A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men". In: *Frontiers in psychology* 10, p. 1604.

Haraway, Donna (1988). "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective". In: *Feminist Studies* 14.3, pp. 575–599.

Harding, Sandra (1991). *Whose Science? Whose Knowledge?: Thinking from Women's Lives*. Ithaca: Cornell University Press.

Harrington, Kate, Lia Litosseliti, Helen Sauntson, and Jane Sunderland, eds. (2008). *Gender and Language Research Methodologies*. Basingstoke: Palgrave Macmillan.

Haslanger, Sally (2000). "Gender and Race: (What) Are They? (What) Do We Want Them To Be?" In: *Noûs* 34.1, pp. 31–55. DOI: https://doi.org/10.1111/0029-4624.00201. eprint: https://onlinelibrary.wiley.com/

doi/pdf/10.1111/0029-4624.00201. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/0029-4624.00201.

— (2011). "Ideology, Generics, and Common Ground". In: *Feminist Metaphysics: Explorations in the Ontology of Sex, Gender and the Self*. Ed. by Charlotte Witt. Feminist Philosophy Collection. Dordrecht: Springer, pp. 179–208.

— (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.

Hovy, Dirk (2015). "Demographic Factors Improve Classification Performance". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*. Beijing, China: Association for Computational Linguistics, pp. 752–762.

— (2018). "The social and the neural network: How to make natural language processing about people again". In: *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pp. 42–49.

Hughes, Jennifer L., Abigail A. Camden, Tenzin Yangchen, Gabrielle P. A. Smith, Melanie M. Domenech Rodríguez, Steven V. Rouse, C. P. McDonald, and Stella Lopez (2022). "Guidance for Researchers When Using Inclusive Demographic Questions for Surveys: Improved and Updated Questions". In: *Psi Chi Journal of Psychological Research* 27.4, pp. 232–255. DOI: 10.24839/2325-7342.JN27.4.232.

Jahan, Israt and Nurul Md Haque (2023). "The Impact Of Gendered Language On Our Communication And Perception Across Contexts And Domains". In: *Journal of Language and Linguistic Studies* 17.4, pp. 3523–3534.

Jahan, Md Saroar and Mourad Oussalah (2023). "A systematic review of hate speech automatic detection using natural language processing". In: *Neurocomputing* 546, p. 126232.

Jeshion, Robin (2020). "Pride and prejudiced: On the reclamation of slurs". In: *Grazer Philosophische Studien* 97.1, pp. 106–137.

Kumar, Sushant, Sumit Datta, Vishakha Singh, Deepanwita Datta, Sanjay Kumar Singh, and Ritesh Sharma (2024). "Applications, Challenges, and Future Directions of Human-in-the-Loop Learning". In: *IEEE Access*.

Kurrek, Jana, Haji Mohammad Saleem, and Derek Ruths (2020). "Towards a comprehensive taxonomy and large-scale annotated corpus for online slur usage". In: *Proceedings of the Fourth Workshop on Online Abuse and Harms*, pp. 138–149.

Kwak, Dong-Heon, Philipp Holtkamp, and Sung S Kim (2019). "Measuring and controlling social desirability bias: Applications in information systems research". In: *Journal of the Association for Information Systems* 20.4, p. 5.

Kwak, Dong-Heon Austin, Xiao Ma, and Sumin Kim (2021). "When does social desirability become a problem? Detection and reduction of social desirability bias in information systems research". In: *Information & Management* 58.7, p. 103500.

Labov, William (1973). *Sociolinguistic patterns*. 4. University of Pennsylvania press.

Landis, J Richard and Gary G Koch (1977). "An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers". In: *Biometrics*, pp. 363–374.

Lave, Jean and Etienne Wenger (1991). "Learning in doing: Social, cognitive, and computational perspectives". In: *Situated learning: Legitimate peripheral participation* 10, pp. 109–155.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. (2009). "Computational social science". In: *Science* 323.5915, pp. 721–723.

Leap, William (1996). *Word's out: Gay men's English*. University of Minnesota Press.

Lippmann, W. (2004). *Public Opinion*. Pelican books. Project Gutenberg. ISBN: 9781412832403.

Litosseliti, Lia (2003). *Using Focus Groups in Research*. London: Continuum.

Lykke, Nina (2010). *Feminist studies: A guide to intersectional theory, methodology and writing*. Routledge.

Maass, Anne, Luciano Arcuri, and Caterina Suitner (2014). "Shaping intergroup relations through language." In.

Malik, Jitendra Singh, Hezhe Qiao, Guansong Pang, and Anton van den Hengel (2024). "Deep learning for hate speech detection: a comparative study". In: *International Journal of Data Science and Analytics*, pp. 1–16.

Marjanovic, Sara, Karolina Stańczak, and Isabelle Augenstein (2022). "Quantifying gender biases towards politicians on Reddit". In: *PloS one* 17.10, e0274317.

Marra, Andrea and Cristina Bosco (2024). "Orsù dunque… avvocato? Osservazione del maschile sovraesteso nei nomina agentis su Twitter". In: *Lingua inclusiva: forme, funzioni, atteggiamenti e percezioni*. Ed. by Anna-Maria De Cesare and Giuliana Giusti. Vol. 6. LiVVaL. Venezia: Edizioni Ca' Foscari – Venice University Press, pp. 203–248. DOI: 10.30687/978-88-6969-866-8/008. URL: http://edizionicafoscari.it/it/edizioni4/libri/978-88-6969-867-5/orsu-dunque-avvocato/.

Marra, Andrea, Chiara Ferrando, Lia Draetta, Bianca Cepollaro, and Viviana Patti (2026). "How is the reclamation of slurs perceived in Italian?: A sociolinguistic survey to inform future NLP studies". In: *Linguistik online* 144.3, pp. 209–225.

Massara, Francesco, Fabio Ancarani, Michele Costabile, and Francesco Ricotta (2012). "Social desirability in virtual communities". In: *International Journal of Business Administration* 3.6, pp. 93–100.

Maturi, Pietro (2020). "Qual è il tuo pronome? Riflessioni su questioni di genere nelle lingue europee". In: *Fuori Luogo Journal of Sociology of Territory, Tourism, Technology* 8.2, pp. 67–74.

Milroy, James (2001). "Language ideologies and the consequences of standardization". In: *Journal of Sociolinguistics* 5.4, pp. 530–555. DOI: https://doi.org/10.1111/1467-9481.00163. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9481.00163. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9481.00163.

Milroy, Lesley and Matthew Gordon (2008). *Sociolinguistics: Method and interpretation*. John Wiley & Sons.

Motschenbacher, Heiko (2010). *Language, Gender and Sexual Identity: Poststructuralist Perspectives*. John Benjamins.

— (2018). "Corpus linguistics in language and sexuality studies: Taking stock and looking ahead". In: *Journal of Language and Sexuality* 7.2, pp. 145–174.

Motschenbacher, Heiko (2019). "Methods in language, gender and sexuality studies: An overview". In: *Wiener Slawistischer Almanach* 84.1, pp. 43–70.

Mutlak, Nisreen (2024). "The Power of Words: Unpacking the Sociolinguistic Impact of Slurs". In: *Aydın İnsan ve Toplum Dergisi* 10.2, pp. 159–182.

Nardone, Chiara (2016). "Asimmetrie semantiche di genere: un'analisi sull'italiano del corpus itWaC". In: *gender/sexuality/italy* 3.

Nguyen, Dong, Laura Rosseel, and Jack Grieve (2021). "On learning and representing social meaning in NLP: a sociolinguistic perspective". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, pp. 603–612.

Nossem, Eva (2019). "Queer, Frocia, Femminiellə, Ricchione et al.–Localizing "Queer" in the Italian Context". In: *gender/sexuality/italy* 6.

Nozza, Debora, Alessandra Teresa Cignarella, Greta Damo, Caselli Tommaso, and Viviana Patti (2023). "HODI at EVALITA 2023: Overview of the EVALITA 2023 task on Homotransphobia Detection in Italian Tweets". In: *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023)*. CEUR-WS.org, (in press).

Office of the United Nations High Commissioner for Human Rights (2013). *Gender stereotyping*. Accessed: June 6, 2025. URL: https://www.ohchr.org/en/women/gender-stereotyping.

Pachinger, Pia, Allan Hanbury, Julia Neidhardt, and Anna Planitzer (May 2023). "Toward Disambiguating the Definitions of Abusive, Offensive, Toxic, and Uncivil Comments". In: *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*. Ed. by Sunipa Dev, Vinodkumar Prabhakaran, David Ifeoluwa Adelani, Dirk Hovy, and Luciana Benotti. Dubrovnik, Croatia: Association for Computational Linguistics, pp. 107–113. DOI: 10.18653/v1/2023.c3nlp-1.11. URL: https://aclanthology.org/2023.c3nlp-1.11/.

Pamungkas, Endang Wahyu, Valerio Basile, and Viviana Patti (2020). "Do You Really Want to Hurt Me? Predicting Abusive Swearing in Social Media". English. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno,

Jan Odijk, and Stelios Piperidis. Marseille, France: European Language Resources Association, pp. 6237–6246. ISBN: 979-10-95546-34-4. URL: https://aclanthology.org/2020.lrec-1.765.

— (2023). "Towards multidomain and multilingual abusive language detection: a survey". In: *Personal and Ubiquitous Computing* 27.1, pp. 17–43.

Pasini, Sveva Silvia, Marco Madeddu, Chiara Ferrando, Chiara Zanchi, and Viviana Patti (2025). "Analyzing Femicide Reactions in YouTube Comments: a Comparative Study of Giulia Cecchettin and Carol Maltesi". In: *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, pp. 886–898.

Pepponi, Elena (2024). *Parole arcobaleno: storia del lessico LGBT+ in Italia*. Mimesis.

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

Rescigno, Argentina Anna, Eva Vanmassenhove, Johanna Monti, and Andy Way (2020). "A case study of natural gender phenomena in translation a comparison of Google Translate, Bing Microsoft Translator and DeepL for English to Italian, French and Spanish". In: *Computational Linguistics CLiC-it 2020* 359.

Robustelli, Cecilia (2012). *Linee guida per l'uso del genere nel linguaggio amministrativo*. Progetto Genere e Linguaggio. Parole e immagini della Comunicazione.

Rolin, Kristina (2009). "Standpoint theory as a methodology for the study of power relations". In: *Hypatia* 24.4, pp. 218–226.

Rosola, Martina, Simona Frenda, Alessandra Teresa Cignarella, Matteo Pellegrini, Andrea Marra, and Mara Floris (2023). "Beyond Obscuration and Visibility: Thoughts on the Different Strategies of Gender-Fair Language in Italian". In: *Proceedings of the Ninth Italian Conference on Computational Linguistics CLiC-it 2023*. Venezia, 30 novembre–2 dicembre 2023. Torino: Accademia University Press, pp. 370–379.

Sabatini, Alma (1986). *Raccomandazioni per un uso non sessista della lingua italiana. Per la scuola e per l'editoria scolastica*. Roma: Presidenza del Consiglio dei Ministri.

Sabatini, Alma (1987). *Il sessismo nella lingua italiana. Raccomandazioni per un uso non sessista della lingua italiana*. Rome: Istituto Poligrafico e Zecca dello Stato.

— (1993). *Ricerca sulla formulazione degli annunci di lavoro*. Roma: Commissione nazionale per la parità e le pari opportunità tra uomo e donna.

Sandri, Marta, Elisa Leonardelli, Sara Tonelli, and Elisabetta Ježek (2023). "Why don't you do it right? analysing annotators' disagreement in subjective tasks". In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 2428–2441.

Sanguinetti, Manuela, Gloria Comandini, Elisa Di Nuovo, Simona Frenda, Marco Stranisci, Cristina Bosco, Tommaso Caselli, Viviana Patti, and Irene Russo (2020). "Haspeede 2@ evalita2020: Overview of the evalita 2020 hate speech detection task". In: *Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*.

Sap, Maarten, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith (2019). "The Risk of Racial Bias in Hate Speech Detection". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*. Florence, Italy: Association for Computational Linguistics, pp. 1668–1678.

Schleef, Erik (2013). "Written surveys and questionnaires in sociolinguistics". In: *Research methods in sociolinguistics: A practical guide*, pp. 42–57.

Schmeisser-Nieto, Wolfgang S, Alessandra Teresa Cignarella, Tom Bourgeade, Simona Frenda, Alejandro Ariza-Casabona, Mario Laurent, Paolo Giovanni Cicirelli, Andrea Marra, Giuseppe Corbelli, Farah Benamara, Cristina Bosco, Véronique Moriceau, Marinella Paciello, Viviana Patti, Mariona Taulé, and Francesca D'Errico (2024a). "Stereohoax: a multilingual corpus of racial hoaxes and social media reactions annotated for stereotypes". In: *Language Resources and Evaluation*, pp. 1–39.

Schmeisser-Nieto, Wolfgang S., Montserrat Nofre, and Mariona Taulé (2022). "Criteria for the annotation of implicit stereotypes". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*, pp. 753–762.

Schmeisser-Nieto, Wolfgang S., Giacomo Ricci, Simona Frenda, Mariona Taulé, and Cristina Bosco (2024b). "Implicit Stereotypes: A Corpus-Based Study for Italian". In: *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*, pp. 997–1004.

Sloane, Mona, Emanuel Moss, Olaitan Awomolo, and Laura Forlano (2020). "Participation is not a design fix for machine learning (pp. 1–7)". In: *Proceedings of the International Conference on Machine Learning, Vienna, Austria*.

Somma, Anna Lisa and Gabriele Maestri, eds. (2020). *Il sessismo nella lingua italiana: trent'anni dopo Alma Sabatini*. Blōnk.

Stańczak, Karolina and Isabelle Augenstein (2021). "A survey on gender bias in natural language processing". In: *arXiv preprint arXiv:2112.14168*.

Sulis, Gigliola and Vera Gheno (2022). "The debate on language and gender in Italy, from the visibility of women to inclusive language (1980s-2020s)". In: *The Italianist* 42.1, pp. 153–183. DOI: 10.1080/02614340.2022.2125707.

Sun, Tony, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang (2019). "Mitigating gender bias in natural language processing: Literature review". In: *arXiv preprint arXiv:1906.08976*.

Thornton, Anna Maria (2016). "Designare le donne: preferenze, raccomandazioni e grammatica". In: *Genere e linguaggio. I segni dell'uguaglianza e della diversità*. Ed. by Fabio Corbisiero, Pietro Maturi, and Elisabetta Ruspini. Milano: FrancoAngeli, pp. 15–33.

— (2022). "Genere e igiene verbale: l'uso di forme con ə in italiano". In: *Annali del Dipartimento di Studi Letterari, Linguistici e Comparati. Sezione linguistica* 11, pp. 11–54.

Uscinski, Joseph E and Lilly J Goren (2011). "What's in a name? Coverage of senator Hillary Clinton during the 2008 democratic primary". In: *Political Research Quarterly* 64.4, pp. 884–896.

Villani, Paola (2020). "Il femminile come "genere del disprezzo". Il caso di presidenta: parola d'odio e fake news". In: *Italiano digitale* 14.3, pp. 111–133.

Violi, Patrizia (1986). *L'infinito singolare. Considerazioni sulla differenza sessuale nel linguaggio*. Verona: Essedue.

Voghera, Miriam and Debora Vena (2016). "Forma maschile, genere femminile: si presentano le donne". In: *Genere e linguaggio. I segni dell'uguaglianza e della diversità*. Ed. by Fabio Corbisiero, Pietro Maturi, and Elisabetta Ruspini. Milano: FrancoAngeli, pp. 34–52.

Wang, Jiayu, Guangyu Jin, and Wenhua Li (2023). "Changing perceptions of language in sociolinguistics". In: *Humanities and Social Sciences Communications* 10.1, pp. 1–9.

Yang, Diyi, Dirk Hovy, David Jurgens, and Barbara Plank (2025). "Socially Aware Language Technologies: Perspectives and Practices". In: *Computational Linguistics* 51.2, pp. 689–703.

Yip, Sun Yee (2024). "Positionality and reflexivity: negotiating insider-outsider positions within and across cultures". In: *International Journal of Research & Method in Education* 47.3, pp. 222–232.

Youngquist, Hunter (2023). "Analyzing Binary Relationships of Identity Labels Using Distributional Semantic Models: A Critical Queer Linguistic Analysis of an English Language Subreddit". In: *Iperstoria* 22.

Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang (2018). "Gender bias in coreference resolution: Evaluation and debiasing methods". In: *arXiv preprint arXiv:1804.06876*.

Zsisku, Eszter, Arkaitz Zubiaga, and Haim Dubossarsky (2024). "Hate Speech Detection and Reclaimed Language: Mitigating False Positives and Compounded Discrimination". In: *Proceedings of the 16th ACM Web Science Conference*, pp. 241–249.

Appendices

The following five appendices reproduce the materials referenced throughout the thesis.

Appendices A–C document the three successive versions of the annotation guidelines for the Rackete–Schettino corpus (before the pilot, after the pilot, and after the feedback meeting) presented in Chapter 6. To track the evolution of the guidelines, changes introduced in Appendix B are highlighted in violet, while further modifications in Appendix C are marked in blue. Appendices D and E include the two questionnaires administered for the research presented in Chapters 6 and 7, respectively.

All appendices are in Italian, the language in which they were administered.

Appendix A

Annotation Guidelines: First Version (*Before-Pilot*)

First version of the annotation schema and guidelines, drafted before the researchers' testing pilot (see Chapter 6).

Linee guida per l'annotazione pilota di tweet relativi a Rackete e Schettino

Presentazione dei personaggi

Francesco Schettino è stato il comandante[1] della nave da crociera Costa Concordia, coinvolta in un tragico naufragio avvenuto il 13 gennaio 2012 nei pressi dell'Isola del Giglio, in Italia. L'incidente, causato da una manovra azzardata di avvicinamento alla costa per il cosiddetto "inchino", portò la nave a urtare uno scoglio, provocando il parziale affondamento e la morte di 32 persone, oltre a numerosi feriti. La gestione dell'emergenza da parte di Schettino fu oggetto di dure critiche, soprattutto per il ritardo nell'ordinare l'evacuazione e per il suo abbandono della nave mentre molti passeggeri erano ancora a bordo. L'episodio suscitò indignazione internazionale e aprì un dibattito sulla sicurezza marittima. Nel 2015, Schettino è stato condannato in via definitiva a 16 anni di reclusione per omicidio colposo, naufragio e abbandono della nave, diventando il simbolo di uno dei più gravi disastri marittimi della storia recente.

Personaggi secondari:

- **Gregorio De Falco** (Capitano della Capitaneria di Porto di Livorno): coordinò i soccorsi e divenne celebre per il suo duro ordine a **Francesco Schettino** di risalire a bordo ("Torni a bordo, cazzo!").
- **Ciro Ambrosio** (primo ufficiale della Costa Concordia): era in plancia con Schettino al momento dell'impatto e contribuì alla gestione dell'emergenza.
- **Domnica Cemortan** (ex ballerina moldava): si trovava a bordo e la sua presenza accanto a Schettino sollevò interrogativi sulla distrazione del comandante.
- **Roberto Bosio** (comandante in seconda): contribuì ai soccorsi e fu tra i primi ufficiali ad aiutare i passeggeri.
- **Gianni Onorato** (direttore generale di Costa Crociere): rilasciò le prime dichiarazioni ufficiali sulla tragedia, difendendo la compagnia.
- **Magistratura italiana**: il **PM di Grosseto, Francesco Verusio**, condusse l'accusa contro Schettino, che fu poi condannato a 16 anni di carcere.
- **Protezione civile, Croce Rossa Italiana e gruppi volontari di vario tipo nonché aziende specifiche**, in generale tutti gli enti e le persone che dal momento della tragedia alle settimane successive si occuparono della gestione umana e ambientale

Carola Rackete è una comandante di navi e attivista tedesca, nota per il suo ruolo nel salvataggio di migranti nel Mediterraneo. Nel giugno 2019, al comando della Sea-Watch 3,

[1] Comandante è inteso come titolo di funzione, riferito a chi è al comando di una nave, un aereo o un'unità militare. Ad esempio, Francesco Schettino era il comandante della Costa Concordia. Capitano è inteso come grado nelle forze armate o un titolo professionale nella marina mercantile. Ad esempio, Gregorio De Falco era Capitano di Fregata della Capitaneria di Porto di Livorno. In sintesi, Schettino era *comandante* della nave, mentre De Falco era *capitano* per grado militare. **Tuttavia, nei tweet da annotare i due termini possono essere utilizzati in maniera impropria e interscambiabile, sia per Rackete che per Schettino.**

nave battente bandiera dei Paesi Bassi e gestita dalla ONG tedesca Sea-Watch, soccorse 53 migranti al largo della Libia. Dopo il trasferimento di alcuni naufraghi per ragioni mediche, rimase a bordo con 42 persone, attendendo per giorni un porto sicuro che le autorità italiane rifiutavano di concedere. Di fronte al peggioramento delle condizioni, decise di entrare senza autorizzazione nel porto di Lampedusa. L'episodio si svolse nel contesto del Decreto Sicurezza Bis, promosso dal Ministro dell'Interno Matteo Salvini, che inaspriva le sanzioni contro le ONG impegnate nei salvataggi. Rackete fu arrestata per violazione delle leggi marittime, ma successivamente rilasciata quando il giudice riconobbe che aveva agito in stato di necessità, rispettando le convenzioni internazionali sul soccorso in mare. Il caso sollevò un acceso dibattito sulle politiche migratorie e umanitarie in Europa.

Personaggi secondari:

- **Matteo Salvini** (all'epoca Vicepremier e Ministro dell'Interno): principale oppositore della Rackete, promosse il "decreto sicurezza" per bloccare gli sbarchi e definì l'azione della capitana un atto illegale.
- **Giuseppe Conte** (all'epoca Presidente del Consiglio): cercò di mediare tra le parti, pur sostenendo la linea del governo.
- **Luigi Di Maio** (all'epoca Vicepremier e Ministro dello Sviluppo Economico): mantenne una posizione intermedia, criticando sia Rackete che l'atteggiamento rigido della Lega.
- **Magistratura italiana**: il **GIP di Agrigento, Alessandra Vella**, dispose la scarcerazione di Rackete, ritenendo che la sua azione fosse giustificata dallo stato di necessità. Il **procuratore di Agrigento, Luigi Patronaggio**, aprì inizialmente l'indagine sulla Rackete per favoreggiamento dell'immigrazione clandestina e resistenza a pubblico ufficiale.
- **Attivisti e ONG**: sostennero la Rackete e denunciarono le politiche migratorie italiane.
- **Forze dell'ordine e Guardia di Finanza**: eseguirono i controlli sulla nave e ne impedirono inizialmente l'attracco.

Etichette

Target

Indica se il target dei messaggi è il personaggio *principale* (Rackete o Schettino) oppure se è un personaggio *secondario* (vedi le due liste nella presentazione) oppure se si fa riferimento al personaggio principale in modo *figurativo*, ad esempio se il personaggio viene citato come termine di paragone per altri eventi o situazioni, anche all'interno di liste di personaggi o fatti, oppure ancora come simbolo di un fenomeno (ad esempio, Schettino come simbolo della decadenza morale italiana).

Esempi:

'#CostaConcordia Dopo le immagini del tg 5 di stasera si e' capito che #schettino oltre ad un incapace e' un delinquente ! In galera a vita !' (principale)

L'Italia non è Schettino ma le centinaia di professionisti che stanno lavorando senza sosta e scamperanno un disastro ambientale' (secondario)

#Morandi fai alla #schettino Abbandona la nave...fidati,rischi di più rimanendo a bordo #occupysanremo' (figurativo)

Responsabilità

Indica se dalla lettura del tweet si evince un'attribuzione di responsabilità rispetto all'accaduto (l'inchino e l'abbandono della nave da parte di Schettino; la violazione del Decreto Sicurezza Bis e l'ingresso senza autorizzazione nel porto di Lampedusa da parte di Rackete). Se presente, indica se la responsabilità viene attribuita a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *società*
- *vittime*
- *altro*

Offensività

Indica l'eventuale presenza di offensività nel tweet. Se presente, indica se viene diretta a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *società*
- *vittime*
- *altro*

Stance

Indica se il tweet è *pro*, *contro* o *neutrale* rispetto all'operato o più in generale alla persona target principale (*Schettino* o *Rackete*).

Nota: considera anche l'eventuale deresponsabilizzazione delle azioni compiute dalla persona target o l'attenuazione di alcune sue caratteristiche negative come elemento a favore (pro), ad esempio attraverso l'empatizzazione verso la persona o l'accaduto.

Humor

Indica se il commento veicola un contenuto umoristico (utilizzando ironia, giochi di parole, iperbole, sarcasmo etc.). Se presente, indica se viene diretta a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *società*
- *vittime*
- *altro*

Tono del discorso

Indica se il discorso riguardante i personaggi e/o l'accaduto viene affrontato in maniera *leggera*, *grave* o *neutrale*.

Capacità cognitive e professionali

Se possibile, indica con quale giudizio viene valutato o percepito il target principale (Schettino o Rackete) in base alle sue competenze e capacità di tipo professionale o intellettivo:

- Negativo
- Positivo
- Non applicabile

Capacità emotive e morali

Se possibile, indica con quale giudizio viene valutato o percepito il target principale (Schettino o Rackete) in base alle sue competenze e capacità di tipo affettivo-emotivo e/o etico-morale:

- Negativo
- Positivo
- Non applicabile

Giudizio estetico

Indica se dal tweet emerge *un giudizio estetico* (su fisico, parti del corpo, abbigliamento etc.) nei confronti del *target principale*.

Stereotipo

Indica l'eventuale presenza di contenuto stereotipico (sia che questo sia implicito che esplicito) legato al target principale (*Sì/No/Non so*).

Se presente, indica **quale categoria di persone riguarda** tra:

- per il caso Rackete:
    - Donne
    - Donne alla guida di una nave
    - Persone migranti
    - Persone africane o Africa o Paesi dell'Africa in generale
    - ONG
    - Altro

- per il caso Schettino:
    - Uomini
    - Uomini alla guida di una nave
    - Donne
    - Donne dell'Est
    - Compagnie di crociera o persone che vanno in crociera
    - Altro

Se presente, indica **a quale sfera afferisce** tra:

- Corpo

- Comportamento o attitudine
- Attività o ruolo
- Sessualità
- Provenienza/Nazionalità

Chi?

Domanda legata rispettivamente alle annotazioni di ***responsabilità***, ***offensività*** e ***humor***. Per ognuna di queste tre categorie, puoi indicare più di una risposta a seconda che il commento sia diretto o riferito a:

- target principale (**Schettino** o **Rackete**)
- **nave o azienda/ONG**
  - per Schettino, **nave** da intendersi come equipaggio o insieme di persone che lavoravano o svolgevano attività sulla Costa Concordia oppure **azienda** (Costa Crociere);
  - per Rackete, **nave**, da intendersi come equipaggio o insieme di persone che lavoravano o svolgevano attività sulla Sea-Watch 3 oppure **ONG** (Sea-Watch);
- **istituzioni**, da intendersi sia come istituzioni o enti pubblici, come ad esempio la magistratura e i suoi membri, oppure come politica, ossia personaggi politici (eventualmente segnalando nelle note se il riferimento è a persone specifiche) e partiti, incluso se il riferimento è al dibattito politico in generale
- **società**, da intendersi in senso ampio (ad es. educazione, mezzi di informazione, dibattito pubblico, utenti del web etc.)
- **vittime**, da intendere come persone che si trovavano sulla nave e hanno subito le decisioni del target (e a prescindere del fatto che queste abbiano arrecato loro morte o danni fisici), ma anche come soggetti o gruppi che non erano sulla nave ma hanno subito conseguenze negative in seguito agli accaduti (es. parenti, gruppi minoritari, etc.). La donna moldava, Domnica, che ha dichiarato di trovarsi con Schettino durante l'accaduto è da considerarsi vittima alla pari delle altre persone presenti sulla nave.
- **altro**

Appendix B

Annotation Guidelines: Second Version (*After-Pilot*)

Second version of the schema and guidelines, revised following the testing pilot and adopted for the first block of annotation (see Chapter 6).

Linee guida per l'annotazione pilota di tweet relativi a Rackete e Schettino

Presentazione dei personaggi

Francesco Schettino è stato il comandante[1] della nave da crociera Costa Concordia, coinvolta in un tragico naufragio avvenuto il 13 gennaio 2012 nei pressi dell'Isola del Giglio, in Italia. L'incidente, causato da una manovra azzardata di avvicinamento alla costa per il cosiddetto "inchino", portò la nave a urtare uno scoglio, provocando il parziale affondamento e la morte di 32 persone, oltre a numerosi feriti. La gestione dell'emergenza da parte di Schettino fu oggetto di dure critiche, soprattutto per il ritardo nell'ordinare l'evacuazione e per il suo abbandono della nave mentre molti passeggeri erano ancora a bordo. L'episodio suscitò indignazione internazionale e aprì un dibattito sulla sicurezza marittima. Nel 2015, Schettino è stato condannato in via definitiva a 16 anni di reclusione per omicidio colposo, naufragio e abbandono della nave, diventando il simbolo di uno dei più gravi disastri marittimi della storia recente.

Personaggi secondari:

- **Gregorio De Falco** (Capitano della Capitaneria di Porto di Livorno): coordinò i soccorsi e divenne celebre per il suo duro ordine a **Francesco Schettino** di risalire a bordo ("Torni a bordo, cazzo!").
- **Ciro Ambrosio** (primo ufficiale della Costa Concordia): era in plancia con Schettino al momento dell'impatto e contribuì alla gestione dell'emergenza.
- **Domnica Cemortan** (ex ballerina moldava): si trovava a bordo e la sua presenza accanto a Schettino sollevò interrogativi sulla distrazione del comandante.
- **Roberto Bosio** (comandante in seconda): contribuì ai soccorsi e fu tra i primi ufficiali ad aiutare i passeggeri.
- **Gianni Onorato** (direttore generale di Costa Crociere): rilasciò le prime dichiarazioni ufficiali sulla tragedia, difendendo la compagnia.
- **Magistratura italiana**: il **PM di Grosseto, Francesco Verusio**, condusse l'accusa contro Schettino, che fu poi condannato a 16 anni di carcere.
- **Protezione civile, Croce Rossa Italiana e gruppi volontari di vario tipo nonché aziende specifiche**, in generale tutti gli enti e le persone che dal momento della tragedia alle settimane successive si occuparono della gestione umana e ambientale

Carola Rackete è una comandante di navi e attivista tedesca, nota per il suo ruolo nel salvataggio di migranti nel Mediterraneo. Nel giugno 2019, al comando della Sea-Watch 3,

[1] Comandante è inteso come titolo di funzione, riferito a chi è al comando di una nave, un aereo o un'unità militare. Ad esempio, Francesco Schettino era il comandante della Costa Concordia. Capitano è inteso come grado nelle forze armate o un titolo professionale nella marina mercantile. Ad esempio, Gregorio De Falco era Capitano di Fregata della Capitaneria di Porto di Livorno. In sintesi, Schettino era *comandante* della nave, mentre De Falco era *capitano* per grado militare. **Tuttavia, nei tweet da annotare i due termini possono essere utilizzati in maniera impropria e interscambiabile, sia per Rackete che per Schettino.**

nave battente bandiera dei Paesi Bassi e gestita dalla ONG tedesca Sea-Watch, soccorse 53 migranti al largo della Libia. Dopo il trasferimento di alcuni naufraghi per ragioni mediche, rimase a bordo con 42 persone, attendendo per giorni un porto sicuro che le autorità italiane rifiutavano di concedere. Di fronte al peggioramento delle condizioni, decise di entrare senza autorizzazione nel porto di Lampedusa. L'episodio si svolse nel contesto del Decreto Sicurezza Bis, promosso dal Ministro dell'Interno Matteo Salvini, che inaspriva le sanzioni contro le ONG impegnate nei salvataggi. Rackete fu arrestata per violazione delle leggi marittime, ma successivamente rilasciata quando il giudice riconobbe che aveva agito in stato di necessità, rispettando le convenzioni internazionali sul soccorso in mare. Il caso sollevò un acceso dibattito sulle politiche migratorie e umanitarie in Europa.

Personaggi secondari:

- **Matteo Salvini** (all'epoca Vicepremier e Ministro dell'Interno): principale oppositore della Rackete, promosse il "decreto sicurezza" per bloccare gli sbarchi e definì l'azione della capitana un atto illegale.
- **Giuseppe Conte** (all'epoca Presidente del Consiglio): cercò di mediare tra le parti, pur sostenendo la linea del governo.
- **Luigi Di Maio** (all'epoca Vicepremier e Ministro dello Sviluppo Economico): mantenne una posizione intermedia, criticando sia Rackete che l'atteggiamento rigido della Lega.
- **Magistratura italiana**: il **GIP di Agrigento, Alessandra Vella**, dispose la scarcerazione di Rackete, ritenendo che la sua azione fosse giustificata dallo stato di necessità. Il **procuratore di Agrigento, Luigi Patronaggio**, aprì inizialmente l'indagine sulla Rackete per favoreggiamento dell'immigrazione clandestina e resistenza a pubblico ufficiale.
- **Attivisti e ONG**: sostennero la Rackete e denunciarono le politiche migratorie italiane.
- **Forze dell'ordine e Guardia di Finanza**: eseguirono i controlli sulla nave e ne impedirono inizialmente l'attracco.

Etichette

Target

Indica se il target dei messaggi è il personaggio *principale* (Rackete o Schettino) oppure se è un personaggio *secondario* (vedi le due liste nella presentazione) oppure se si fa riferimento al personaggio principale in modo *figurativo*, ad esempio se il personaggio viene citato come termine di paragone per altri eventi o situazioni, anche all'interno di liste di personaggi o fatti, oppure ancora come simbolo di un fenomeno (ad esempio, Schettino come simbolo della decadenza morale italiana).
Se il target non rientra nell'elenco dei personaggi secondari, seleziona Altro.

Esempi:

'#CostaConcordia Dopo le immagini del tg 5 di stasera si e' capito che #schettino oltre ad un incapace e' un delinquente ! In galera a vita !' (principale)

L'Italia non è Schettino ma le centinaia di professionisti che stanno lavorando senza sosta e scamperanno un disastro ambientale' (secondario)

#Morandi fai alla #schettino Abbandona la nave...fidati,rischi di più rimanendo a bordo #occupysanremo' (figurativo)

Responsabilità

Indica se dalla lettura del tweet si evince un'attribuzione di responsabilità rispetto all'accaduto (l'inchino e l'abbandono della nave da parte di Schettino; la violazione del Decreto Sicurezza Bis e l'ingresso senza autorizzazione nel porto di Lampedusa da parte di Rackete). Se presente, indica se la responsabilità viene attribuita a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *politica*
- *società*
- *vittime*
- *altro*

Offensività

Indica l'eventuale presenza di offensività nel tweet. Se presente, indica se viene diretta a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *politica*
- *società*
- *vittime*
- *altro*

Stance

Indica se il tweet è *pro*, *contro* o *neutrale* rispetto all'operato o più in generale alla persona target principale (*Schettino* o *Rackete*).

Nota: considera anche l'eventuale deresponsabilizzazione delle azioni compiute dalla persona target o l'attenuazione di alcune sue caratteristiche negative come elemento a favore (pro), ad esempio attraverso l'empatizzazione verso la persona o l'accaduto.

Humor

Indica se il commento veicola un contenuto umoristico (utilizzando ironia, giochi di parole, iperbole, sarcasmo etc.). Se presente, indica se viene diretta a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *politica*
- *società*
- *vittime*

- *altro*

Tono del discorso
Indica se il discorso riguardante i personaggi e/o l'accaduto viene affrontato in maniera *leggera*, *seria* o *neutrale*.

Competenze professionali
Se possibile, indica con quale giudizio viene valutato o percepito il target principale (Schettino o Rackete) in base alle sue competenze e capacità professionali (emotive, morali, intellettive,...):
- Negativo
- Positivo
- Non applicabile

Giudizio estetico
Indica se dal tweet emerge *un giudizio estetico* (su fisico, parti del corpo, abbigliamento etc.) nei confronti del *target principale*.

Stereotipo
Indica l'eventuale presenza di qualsiasi contenuto stereotipico, sia che questo sia implicito che esplicito (Sì/No/Non so), anche se non è legato direttamente al target principale. In tal caso, se nelle opzioni successive non trovi la categoria più adeguata, seleziona Altro.
Se lo stereotipo è presente, indica quale categoria di persone riguarda tra:

Se presente, indica **quale categoria di persone riguarda** tra:

- per il caso Rackete:
  - Donne
  - Donne alla guida di una nave
  - Persone migranti
  - Persone africane o Africa o Paesi dell'Africa in generale
  - ONG
  - Altro

- per il caso Schettino:
  - Uomini
  - Uomini alla guida di una nave
  - Donne
  - Donne dell'Est
  - Compagnie di crociera o persone che vanno in crociera
  - Altro

Se presente, indica **a quale sfera afferisce** tra:
- Corpo
- Comportamento o attitudine
- Attività o ruolo
- Sessualità

- Provenienza/Nazionalità

Chi?

Domanda legata rispettivamente alle annotazioni di ***responsabilità***, ***offensività*** e ***humor***. Per ognuna di queste tre categorie, puoi indicare più di una risposta a seconda che il commento sia diretto o riferito a:

- target principale (**Schettino** o **Rackete**)
- **nave o azienda/ONG**
  - per Schettino, **nave** da intendersi come equipaggio o insieme di persone che lavoravano o svolgevano attività sulla Costa Concordia oppure **azienda** (Costa Crociere);
  - per Rackete, **nave**, da intendersi come equipaggio o insieme di persone che lavoravano o svolgevano attività sulla Sea-Watch 3 oppure **ONG** (Sea-Watch);
- **istituzioni**, da intendersi sia come istituzioni o enti pubblici, come ad esempio la magistratura e i suoi membri
- **politica**, ossia personaggi politici (eventualmente segnalando nelle note se il riferimento è a persone specifiche) e partiti, incluso se il riferimento è al dibattito politico in generale
- **società**, da intendersi in senso ampio (ad es. educazione, mezzi di informazione, dibattito pubblico, utenti del web etc.)
- **vittime**, da intendere come persone che si trovavano sulla nave e hanno subito le decisioni del target (e a prescindere del fatto che queste abbiano arrecato loro morte o danni fisici), ma anche come soggetti o gruppi che non erano sulla nave ma hanno subito conseguenze negative in seguito agli accaduti (es. parenti, gruppi minoritari, etc.). La donna moldava, Domnica, che ha dichiarato di trovarsi con Schettino durante l'accaduto è da considerarsi vittima alla pari delle altre persone presenti sulla nave.
- **altro**

Appendix C

Annotation Guidelines: Third and Final Version (*After Feedback Meeting*)

Third and final version of the guidelines, produced after the feedback meeting and constituting the definitive annotation framework used for the subsequent annotation and analyses (see Chapter 6).

Linee guida per l'annotazione pilota di tweet relativi a Rackete e Schettino

Presentazione dei personaggi

Francesco Schettino è stato il comandante[1] della nave da crociera Costa Concordia, coinvolta in un tragico naufragio avvenuto il 13 gennaio 2012 nei pressi dell'Isola del Giglio, in Italia. L'incidente, causato da una manovra azzardata di avvicinamento alla costa per il cosiddetto "inchino", portò la nave a urtare uno scoglio, provocando il parziale affondamento e la morte di 32 persone, oltre a numerosi feriti. La gestione dell'emergenza da parte di Schettino fu oggetto di dure critiche, soprattutto per il ritardo nell'ordinare l'evacuazione e per il suo abbandono della nave mentre molti passeggeri erano ancora a bordo. L'episodio suscitò indignazione internazionale e aprì un dibattito sulla sicurezza marittima. Nel 2015, Schettino è stato condannato in via definitiva a 16 anni di reclusione per omicidio colposo, naufragio e abbandono della nave, diventando il simbolo di uno dei più gravi disastri marittimi della storia recente.

Personaggi secondari (elenco esemplificativo, non esaustivo):

- **Gregorio De Falco** (Capitano della Capitaneria di Porto di Livorno): coordinò i soccorsi e divenne celebre per il suo duro ordine a **Francesco Schettino** di risalire a bordo ("Torni a bordo, cazzo!").
- **Ciro Ambrosio** (primo ufficiale della Costa Concordia): era in plancia con Schettino al momento dell'impatto e contribuì alla gestione dell'emergenza.
- **Domnica Cemortan** (ex ballerina moldava): si trovava a bordo e la sua presenza accanto a Schettino sollevò interrogativi sulla distrazione del comandante.
- **Roberto Bosio** (comandante in seconda): contribuì ai soccorsi e fu tra i primi ufficiali ad aiutare i passeggeri.
- **Gianni Onorato** (direttore generale di Costa Crociere): rilasciò le prime dichiarazioni ufficiali sulla tragedia, difendendo la compagnia.
- **Magistratura italiana**: il **PM di Grosseto, Francesco Verusio**, condusse l'accusa contro Schettino, che fu poi condannato a 16 anni di carcere.
- **Protezione civile, Croce Rossa Italiana e gruppi volontari di vario tipo nonché aziende specifiche**, in generale tutti gli enti e le persone che dal momento della tragedia alle settimane successive si occuparono della gestione umana e ambientale.
- **Altre figure:** tutte le persone o gli enti che non rientrano nell'elenco precedente e/o che non sono legati ai personaggi principali.

---

[1] Comandante è inteso come titolo di funzione, riferito a chi è al comando di una nave, un aereo o un'unità militare. Ad esempio, Francesco Schettino era il comandante della Costa Concordia. Capitano è inteso come grado nelle forze armate o un titolo professionale nella marina mercantile. Ad esempio, Gregorio De Falco era Capitano di Fregata della Capitaneria di Porto di Livorno. In sintesi, Schettino era *comandante* della nave, mentre De Falco era *capitano* per grado militare. **Tuttavia, nei tweet da annotare i due termini possono essere utilizzati in maniera impropria e interscambiabile, sia per Rackete che per Schettino.**

Carola Rackete è una comandante di navi e attivista tedesca, nota per il suo ruolo nel salvataggio di migranti nel Mediterraneo. Nel giugno 2019, al comando della Sea-Watch 3, nave battente bandiera dei Pa
esi Bassi e gestita dalla ONG tedesca Sea-Watch, soccorse 53 migranti al largo della Libia. Dopo il trasferimento di alcuni naufraghi per ragioni mediche, rimase a bordo con 42 persone, attendendo per giorni un porto sicuro che le autorità italiane rifiutavano di concedere. Di fronte al peggioramento delle condizioni, decise di entrare senza autorizzazione nel porto di Lampedusa. L'episodio si svolse nel contesto del Decreto Sicurezza Bis, promosso dal Ministro dell'Interno Matteo Salvini, che inaspriva le sanzioni contro le ONG impegnate nei salvataggi. Rackete fu arrestata per violazione delle leggi marittime, ma successivamente rilasciata quando il giudice riconobbe che aveva agito in stato di necessità, rispettando le convenzioni internazionali sul soccorso in mare. Il caso sollevò un acceso dibattito sulle politiche migratorie e umanitarie in Europa.

Personaggi secondari (elenco esemplificativo, non esaustivo):

- **Matteo Salvini** (all'epoca Vicepremier e Ministro dell'Interno): principale oppositore della Rackete, promosse il "decreto sicurezza" per bloccare gli sbarchi e definì l'azione della capitana un atto illegale.
- **Giuseppe Conte** (all'epoca Presidente del Consiglio): cercò di mediare tra le parti, pur sostenendo la linea del governo.
- **Luigi Di Maio** (all'epoca Vicepremier e Ministro dello Sviluppo Economico): mantenne una posizione intermedia, criticando sia Rackete che l'atteggiamento rigido della Lega.
- **Magistratura italiana**: il **GIP di Agrigento, Alessandra Vella**, dispose la scarcerazione di Rackete, ritenendo che la sua azione fosse giustificata dallo stato di necessità. Il **procuratore di Agrigento, Luigi Patronaggio**, aprì inizialmente l'indagine sulla Rackete per favoreggiamento dell'immigrazione clandestina e resistenza a pubblico ufficiale.
- **Attivisti e ONG**: sostennero la Rackete e denunciarono le politiche migratorie italiane.
- **Forze dell'ordine e Guardia di Finanza**: eseguirono i controlli sulla nave e ne impedirono inizialmente l'attracco.
- **Altri personaggi legati alla politica**: Movimento 5 Stelle, Partito Democratico, etc…
- **Altre figure:** tutte le persone o gli enti che non rientrano nell'elenco precedente e/o che non sono legati ai personaggi principali.

Etichette

Target

Indica se il target dei messaggi è il personaggio *principale* (Rackete o Schettino) oppure se è un personaggio *secondario* (puoi trovare un elenco di figure ricorrenti nelle due liste nella presentazione) oppure se si fa riferimento al personaggio principale in modo *figurativo*, ad esempio se il personaggio viene citato come termine di paragone per altri eventi o situazioni, anche all'interno di liste di personaggi o fatti, oppure ancora come simbolo di un fenomeno (ad esempio, Schettino come simbolo della decadenza morale italiana).

~~Se il target non rientra nell'elenco dei personaggi secondari, seleziona Altro.~~

Esempi:

'#CostaConcordia Dopo le immagini del tg 5 di stasera si e' capito che #schettino oltre ad un incapace e' un delinquente ! In galera a vita !' (principale)

L'Italia non è Schettino ma le centinaia di professionisti che stanno lavorando senza sosta e scamperanno un disastro ambientale' (secondario)

#Morandi fai alla #schettino Abbandona la nave...fidati,rischi di più rimanendo a bordo #occupysanremo' (figurativo)

Responsabilità

Indica se dalla lettura del tweet si evince un'attribuzione di responsabilità rispetto all'accaduto (l'inchino e l'abbandono della nave da parte di Schettino; la violazione del Decreto Sicurezza Bis e l'ingresso senza autorizzazione nel porto di Lampedusa da parte di Rackete). L'attribuzione di responsabilità può prescindere dal giudizio di chi scrive rispetto al personaggio (ossia, il *sentiment* del tweet può comunque apparire positivo anche in presenza di un'attribuzione di responsabilità). Se presente, indica se la responsabilità viene attribuita a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *politica*
- *società*
- *vittime*
- *altro*

Offensività

Indica l'eventuale presenza di offensività nel tweet. Se presente, indica se viene diretta a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *politica*
- *società*
- *vittime*
- *altro*

Stance

Indica se il tweet è *pro*, *contro* o *neutrale* rispetto all'operato o più in generale alla persona target principale (*Schettino* o *Rackete*).

Nota: considera anche l'eventuale deresponsabilizzazione delle azioni compiute dalla persona target o l'attenuazione di alcune sue caratteristiche negative come elemento a favore (pro), ad esempio attraverso l'empatizzazione verso la persona o l'accaduto.

Humor
Indica se il commento veicola un contenuto umoristico (utilizzando ironia, giochi di parole, iperbole, sarcasmo etc.). Se presente, indica se viene diretta a (risposta multipla):

- *Schettino* oppure *Rackete*
- *azienda/nave* oppure *ONG/nave*
- *istituzioni*
- *politica*
- *società*
- *vittime*
- *altro*

~~Tono del discorso~~
~~Indica se il discorso riguardante i personaggi e/o l'accaduto viene affrontato in maniera *leggera*, *seria* o *neutrale*.~~

Competenze professionali
Se possibile, indica con quale giudizio viene valutato o percepito il target principale (Schettino o Rackete) in base alle sue competenze e capacità professionali (emotive, morali, intellettive,...):

- Negativo
- Positivo
- Non applicabile

~~Giudizio estetico~~
~~Indica se dal tweet emerge *un giudizio estetico* (su fisico, parti del corpo, abbigliamento etc.) nei confronti del *target principale*.~~

Stereotipo
Indica l'eventuale presenza di qualsiasi contenuto stereotipico, sia che questo sia implicito che esplicito (Sì/No/Non so), anche se non è legato direttamente al target principale. In tal caso, se nelle opzioni successive non trovi la categoria più adeguata, seleziona Altro.
Se lo stereotipo è presente, indica quale categoria di persone riguarda tra:

Se presente, indica **quale categoria di persone riguarda** tra:

- per il caso Rackete:
  - Donne
  - Donne alla guida di una nave
  - Uomini
  - Persone migranti
  - Persone africane o Africa o Paesi dell'Africa in generale
  - ONG
  - Società
  - Altro

- per il caso Schettino:
  - Uomini

- Uomini alla guida di una nave
- Donne
- Donne dell'Est
- Compagnie di crociera o persone che vanno in crociera
- Società
- Altro

Se presente, indica **a quale sfera afferisce** tra:

- Corpo/Sessualità
- Comportamento/Attitudine
- Attività/Ruolo ricoperto
- ~~Sessualità~~
- Provenienza/Nazionalità

Chi?

Domanda legata rispettivamente alle annotazioni di ***responsabilità***, ***offensività*** e ***humor***. Per ognuna di queste tre categorie, puoi indicare più di una risposta a seconda che il commento sia diretto o riferito a:

- target principale (**Schettino** o **Rackete**)
- **nave o azienda/ONG**
  - per Schettino, **nave** da intendersi come equipaggio o insieme di persone che lavoravano o svolgevano attività sulla Costa Concordia oppure **azienda** (Costa Crociere);
  - per Rackete, **nave**, da intendersi come equipaggio o insieme di persone che lavoravano o svolgevano attività sulla Sea-Watch 3 oppure **ONG** (Sea-Watch);
- **istituzioni**, da intendersi sia come istituzioni o enti pubblici, come ad esempio la magistratura e i suoi membri
- **politica**, ossia personaggi politici (eventualmente segnalando nelle note se il riferimento è a persone specifiche) e partiti, incluso se il riferimento è al dibattito politico in generale
- **società**, da intendersi in senso ampio (ad es. educazione, mezzi di informazione, dibattito pubblico, utenti del web etc.)
- **vittime**, da intendere come persone che si trovavano sulla nave e hanno subito le decisioni del target (e a prescindere del fatto che queste abbiano arrecato loro morte o danni fisici), ma anche come soggetti o gruppi che non erano sulla nave ma hanno subito conseguenze negative in seguito agli accaduti (es. parenti, gruppi minoritari, etc.). La donna moldava, Domnica, che ha dichiarato di trovarsi con Schettino durante l'accaduto è da considerarsi vittima alla pari delle altre persone presenti sulla nave.
- **altro**

Appendix D

Post-Annotation Questionnaire

Reflective questionnaire administered via Google Form to the three annotators at the end of the annotation process (see Chapter 6).

Questionario post-annotazione Corpus Rackete/Schettino

Questo questionario costituisce l'ultimo step del processo di annotazione che ha visto il tuo coinvolgimento. Ti ringraziamo ancora una volta per il tempo, la disponibilità e la partecipazione attiva che hai dedicato a questo progetto.

Il questionario raccoglie alcune informazioni socio-demografiche e include domande riflessive sulla tua esperienza con il processo e lo schema di annotazione. Le risposte contribuiranno a contestualizzare le analisi condotte e a rendere il processo più trasparente e riflessivo.

Le informazioni saranno trattate in maniera riservata ed esclusivamente per finalità di ricerca, nel rispetto del Regolamento (UE) 2016/679 (GDPR) sulla protezione dei dati personali.

* Indica una domanda obbligatoria

1. Email *

Sezione A - Informazioni personali

2. In quale fascia d'età rientri? *

Contrassegna solo un ovale.

- 18-24
- 25-34
- 35-44
- 45+

3. Quale di queste definizioni ti rappresenta meglio? *

Contrassegna solo un ovale.

- ◯ Donna cis (mi riconosco nel genere assegnato alla nascita)
- ◯ Donna trans (non mi riconosco nel genere assegnato alla nascita)
- ◯ Uomo cis (mi riconosco nel genere assegnato alla nascita)
- ◯ Uomo trans (non mi riconosco nel genere assegnato alla nascita)
- ◯ Persona non binaria / Genderqueer / Agender / Genderfluid
- ◯ Preferisco non rispondere
- ◯ Altro: _____

4. Qual è il tuo background di studi/formazione principale? *

Seleziona tutte le voci applicabili.

- ☐ Linguistica
- ☐ Linguistica computazionale
- ☐ Informatica / AI / Data Science
- ☐ Scienze sociali/umane
- ☐ Altro: _____

Sezione B - Familiarità con lo schema

5. Qual era la tua esperienza in attività di annotazione linguistica **prima** di questo progetto? *

Contrassegna solo un ovale.

- ◯ Nessuna
- ◯ Limitata (1-2 progetti)
- ◯ Moderata (3-5 progetti)
- ◯ Estesa (più di 5 progetti)

6. Hai svolto altre attività di annotazione linguistica **parallelamente** a questa? *

Contrassegna solo un ovale.

- ◯ No, nessuna
- ◯ Sì, una
- ◯ Sì, più di una

7. Se sì, senti che l'aver svolto altre attività di annotazione parallele **abbia influenzato il tuo processo di annotazione** in questo progetto? In che modo?

_____

All'inizio del progetto, **quanta familiarità sentivi di avere con i seguenti temi?** (Risposta su scala 1–5: 1 = nessuna familiarità, 5 = molta familiarità)

8. Stereotipi di genere *

Contrassegna solo un ovale.

	1	2	3	4	5	
ness	◯	◯	◯	◯	◯	molta familiarità

9. Discorsi sulla migrazione *

Contrassegna solo un ovale.

	1	2	3	4	5	
ness	◯	◯	◯	◯	◯	molta familiarità

10. Intersezionalità *

Contrassegna solo un ovale.

	1	2	3	4	5	
ness	◯	◯	◯	◯	◯	molta familiarità

11. Schettino e il naufragio della Costa Concordia *

Contrassegna solo un ovale.

	1	2	3	4	5	
ness	◯	◯	◯	◯	◯	molta familiarità

12. Rackete e le vicende della SeaWatch3 *

Contrassegna solo un ovale.

	1	2	3	4	5	
ness	◯	◯	◯	◯	◯	molta familiarità

13. Ritieni che **la tua sensibilità personale verso questi temi abbia influenzato** le tue decisioni di annotazione?

Contrassegna solo un ovale.

- ◯ Per nulla
- ◯ Poco
- ◯ Abbastanza
- ◯ Molto

Sezione C - Processo di annotazione

Indica, **per ogni categoria riportata**, su una scala da 1 a 5 (1=molto facile; 5=molto difficile), che **livello di difficoltà hai percepito mentre annotavi**:

14. **Target** [*Principale*, *Secondario*, *Figurativo*] *

Contrassegna solo un ovale.

	1	2	3	4	5	
molt	◯	◯	◯	◯	◯	molto difficile

15. **Responsabilità** [*Sì, No, Non so*]. *
Se sì:
Verso chi? [*Schettino/Rackete, Azienda-Nave/ONG-Nave, Istituzioni, Politica, Società, Vittime, Altro*]

Contrassegna solo un ovale.

	1	2	3	4	5	
molt	◯	◯	◯	◯	◯	molto difficile

16. **Offensività** [*Sì, No, Non so*]. *
Se sì:
Verso chi? [*Schettino/Rackete, Azienda-Nave/ONG-Nave, Istituzioni, Politica, Società, Vittime, Altro]*

Contrassegna solo un ovale.

	1	2	3	4	5	
molt	◯	◯	◯	◯	◯	molto difficile

17. **Stance** [*Pro, Contro, Neutrale*] *

Contrassegna solo un ovale.

	1	2	3	4	5	
molt	◯	◯	◯	◯	◯	molto difficile

18. **Humor** [*Sì, No, Non so*]. *
Se sì:
Verso chi? [*Schettino/Rackete, Azienda-Nave/ONG-Nave, Istituzioni, Politica, Società, Vittime, Altro*]

Contrassegna solo un ovale.

	1	2	3	4	5	
molt	◯	◯	◯	◯	◯	molto difficile

19. **Competenza professionale (cognitive, morali, emotive)** [*Negativo, Positivo, Non Applicabile*] *

Contrassegna solo un ovale.

	1	2	3	4	5	
molt	◯	◯	◯	◯	◯	molto difficile

20. **Stereotipo** [*Sì, No, Non so*]. *
Se sì:
Chi riguarda lo stereotipo? [SCHETTINO: *Uomini, Uomini alla guida di una nave, Donne, Donne dell'Est, Compagnie di crociera o persone che vanno in crociera, Società, Altro;* RACKETE: *Donne, Donne alla guida di una nave, Uomini, Persone migranti, Persone africane o Africa o Paesi dell'Africa in generale, ONG, Società, Altro*]
A quale sfera afferisce lo stereotipo? [*Corpo/Sessualità, Comportamento/Attitudine, Attività/Ruolo ricoperto, Provenienza/Nazionalità*]

Contrassegna solo un ovale.

	1	2	3	4	5	
molt	◯	◯	◯	◯	◯	molto difficile

21. Ricordi di aver provato **incertezza verso alcune etichette specifiche?** Se sì, quali? E come l'hai affrontata (quali strategie pensi di aver adottato)? *

22. Alla luce del confronto con le altre persone che hanno partecipato a questo progetto di annotazione e della successiva revisione dello schema, **quali ritieni che possano essere le fonti principali di disaccordo** tra annotatori/annotatrici? *

Seleziona tutte le voci applicabili.

- [ ] Ambiguità dei testi
- [ ] Mancanza di informazioni (es. poco contesto, testi incompleti...)
- [ ] Soggettività (es. differenze di interpretazione personale, bias...)
- [ ] Scarsa chiarezza delle Linee Guida
- [ ] Altro: _____

23. Se vuoi, motiva brevemente la tua risposta:

24. Se vuoi, lascia un commento sulla tua esperienza (ad es., su come miglioreresti lo schema di annotazione):

Appendix E

The *Linguaggio e Riappropriazione* Questionnaire

Sociolinguistic questionnaire of the FAVOLOSA project, administered via Google Form (see Chapter 7).

Questionario "Linguaggio e Riappropriazione"

Il sondaggio che stai per compilare è **completamente anonimo** ed è stato approvato dal Comitato di Bioetica dell'Università di Torino, dove il progetto *F.A.V.O.L.O.S.A. - Fair Automatization and Visibility Of LGBT+ community On Slur (re)Appropriation* è stato sottoposto a revisione etica per garantire che la raccolta, l'analisi e l'archiviazione dei dati siano conformi al EU GDPR (puoi controllare i dettagli qui: Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali).

ATTENZIONE: nel questionario sono presenti frasi tratte da social network, che potrebbero veicolare contenuti offensivi ed urtare la tua sensibilità.

Non esistono risposte giuste o sbagliate! Dal momento che le frasi vengono presentate senza il loro contesto, potrebbe risultare difficile capirle e valutarle. Ti chiediamo tuttavia di non preoccuparti, ma di basare le risposte esclusivamente sulla tua percezione e sulla tua esperienza!

Se desideri ricevere maggiori informazioni sul sondaggio o sulla ricerca che esso informerà, invia un'email a andrea.marra@unito.it.
Le tue risposte rimarranno completamente anonime.
La compilazione del questionario richiederà circa 25 minuti.

Grazie per la tua considerazione!

* Indica una domanda obbligatoria

Consenso

1. Seleziona una delle due opzioni: *

   Contrassegna solo un ovale.

   - Acconsento a partecipare, confermando di avere più 18 anni e che l'italiano sia la mia lingua nativa (o una delle mie lingue native). *Passa alla domanda 2.*
   - Non voglio o non posso partecipare a questo sondaggio.

PARTE I

Iniziamo!

#1

per riproclamarlo E BASTA. Se lo usano altri è solo continua oppressione. Per di più Harry non è un vostro amico, e non vi dovete permettere di dire certe cose. Se una mia amica mi chiama frocia/troia non me la prendo perché hanno il permesso di farlo ma lui non lo conoscete.

2. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

3. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

4. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

5. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#2

Le mie coinquiline mi hanno chiesto di parlare col padrone di casa il quale non sospetta minimamente che io sia ricchione sarà divertente spiegargli la rainbow flag in camera

6. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- ☐ in un contesto amicale
- ☐ in un contesto familiare
- ☐ in un contesto lavorativo
- ☐ in un contesto attivista
- ☐ Altro: _____

7. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

8. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

9. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#3

Un po' come quando tra gay si chiamano "finocchio", "frocio" o "ricchione". Lo dicessi io sarei un omofobo

10. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

11. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

12. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

13. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#4

Bro ma ti rendi conto che così pari solo un succhiacazzi per quell'altro frocio 😂

14. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

15. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

16. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

17. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#5

Fra siamo letteralmente gay, posso usare le parole frocio finocchio e ricchione, nessuno l3 usa più come insulti, dovresti aggiornarti un attimo che siamo 2022

18. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

19. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

20. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

21. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#6

Sto tornando a casa ora dal pride + serata frocia 💀

22. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

23. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

24. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

25. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#7

Allora posso dirti per la parte non cishet e io ho trovato un lavoro, anzi me ne sono stati offerti anche Coi capelli viola corti e visibilmente frocia. Come al solito hanno fatto speculazioni alle mie spalle ma il lavoro me lo hanno dato. Però conta che vivo a milano e+

26. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

27. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

28. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

29. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#8

Sei solo un invertito che non vuole ammettere che certi comportamenti sono schifosi. Vuoi fare il ricchione? Bene, mettiti il guanto.

30. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

31. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

32. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

33. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#9

RT Mi basta ascoltare 10 secondi DRAGHI per realizzare quanto Conte è una nullità (e checca rancorosa).

34. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

35. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

36. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

37. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

#10

Ama io parlo come mi pare essere checca è uno spasso giuro

38. In quale situazione comunicativa potrebbe apparire questa frase? *
[Puoi indicare più di una risposta]

Seleziona tutte le voci applicabili.

- [ ] in un contesto amicale
- [ ] in un contesto familiare
- [ ] in un contesto lavorativo
- [ ] in un contesto attivista
- [ ] Altro: _____

39. Chi scrive la frase fa parte della comunità LGBT+? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

40. Chi scrive ha un intento offensivo? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente offensivo

41. La frase veicola odio? *
[Valuta da 1 a 5]

Contrassegna solo un ovale.

	1	2	3	4	5	
Per	◯	◯	◯	◯	◯	Decisamente sì

PARTE II

Siamo circa a metà del questionario!

Ti chiediamo di rispondere alle prossime domande considerando che con **riappropriazione** (e **riappropriativo**) facciamo riferimento a un fenomeno linguistico per il quale *le persone rivendicano l'utilizzo di termini dispregiativi per esprimere la propria identità e/o l'appartenenza ad una certa comunità.*

#1
per riproclamarlo E BASTA. Se lo usano altri è solo continua oppressione. Per di più Harry non è un vostro amico, e non vi dovete permettere di dire certe cose. Se una mia amica mi chiama ***frocia****/troia non me la prendo perché hanno il permesso di farlo ma lui non lo conoscete.*

42. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

43. Per favore, spiega brevemente il perché:

44. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#2

*Le mie coinquiline mi hanno chiesto di parlare col padrone di casa il quale non sospetta minimamente che io sia **ricchione** sarà divertente spiegargli la rainbow flag in camera*

45. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

46. Per favore, spiega brevemente il perché:

47. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#3

*Un po' come quando tra gay si chiamano "finocchio", "**frocio**" o "ricchione". Lo dicessi io sarei un omofobo*

48. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

49. Per favore, spiega brevemente il perché:

50. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#4

Bro ma ti rendi conto che così pari solo un succhiacazzi per quell'altro ***frocio*** 😂

51. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

52. Per favore, spiega brevemente il perché:

53. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#5

*Fra siamo letteralmente gay, posso usare le parole **frocio** finocchio e ricchione, nessuno l3 usa più come insulti, dovresti aggiornarti un attimo che siamo 2022*

54. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

55. Per favore, spiega brevemente il perché:

56. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#6

*Sto tornando a casa ora dal pride + serata **frocia*** 💀

57. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

58. Per favore, spiega brevemente il perché:

59. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#7

*Allora posso dirti per la parte non cishet e io ho trovato un lavoro, anzi me ne sono stati offerti anche Coi capelli viola corti e visibilmente **frocia**. Come al solito hanno fatto speculazioni alle mie spalle ma il lavoro me lo hanno dato. Però conta che vivo a milano e+*

60. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

61. Per favore, spiega brevemente il perché:

62. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#8

*Sei solo un invertito che non vuole ammettere che certi comportamenti sono schifosi. Vuoi fare il **ricchione**? Bene, mettiti il guanto.*

63. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

64. Per favore, spiega brevemente il perché:

65. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#9

*RT Mi basta ascoltare 10 secondi DRAGHI per realizzare quanto Conte è una nullità (e **checca** rancorosa).*

66. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

67. Per favore, spiega brevemente il perché:

68. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

#10

*Ama io parlo come mi pare essere **checca** è uno spasso giuro*

69. In questo caso la persona che scrive si è riappropriata della parola evidenziata? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

70. Per favore, spiega brevemente il perché:

71. A te capita o è mai capitato di utilizzare la parola evidenziata in maniera simile? *

Contrassegna solo un ovale.

- Sì
- No
- Non saprei

PARTE III

Non manca molto, siamo alla penultima sezione!

72. Nella tua esperienza quotidiana ti capita o ti è capitato di sentire usi di linguaggio riappropriativo simile a quello proposto negli esempi sopra? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

73. Ritieni legittimo l'utilizzo di tali parole o espressioni in termini riappropriativi da parte di:
[Puoi indicare più di una risposta] *

Seleziona tutte le voci applicabili.

- ☐ Chiunque
- ☐ Persone LGBT+
- ☐ Persone o gruppi che fanno attivismo LGBT+
- ☐ Rete amicale di persone LGBT+
- ☐ Familiari di persone LGBT+
- ☐ Persone del mondo dello spettacolo e dell'intrattenimento (TV, spettacoli comici, teatro etc.)
- ☐ Persone del mondo dell'arte e della letteratura (libri, cinema, serie tv, etc.)
- ☐ Nessuna persona
- ☐ Altro: _____

74. In generale, a te capita o è capitato di utilizzare parole o espressioni con intento riappropriativo? *

Contrassegna solo un ovale.

- ◯ Sì
- ◯ No
- ◯ Non saprei

75. Oltre a quelle citate, conosci altre parole che vengono utilizzate in modo riappropriativo? *
[Se sì, indica quali]

76. Se ti va, commenta brevemente le tue risposte:

PARTE IV

Il questionario è quasi finito!

Ti chiediamo di compilare quest'ultima sezione, ricordandoti ancora una volta che i dati verranno trattati in maniera anonima.

77. In quale fascia d'età rientri? *

Contrassegna solo un ovale.

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- >65

78. Quale di queste definizioni ti rappresenta meglio? *

Contrassegna solo un ovale.

- Donna cis (mi riconosco nel genere assegnato alla nascita)
- Donna trans (non mi riconosco nel genere assegnato alla nascita)
- Uomo cis (mi riconosco nel genere assegnato alla nascita)
- Uomo trans (non mi riconosco nel genere assegnato alla nascita)
- Persona non binaria / Genderqueer / Agender / Genderfluid
- Preferisco non rispondere
- Altro: _____

79. Quale orientamento sessuale e/o romantico ti definisce meglio? *

Contrassegna solo un ovale.

- Asessuale/Aromantico
- Bisessuale
- Eterosessuale
- Gay
- Lesbica/Lesbicx
- Pansessuale
- Preferisco non rispondere
- Altro: _____

80. In quale regione hai vissuto la maggior parte del tempo prima dei tuoi 18 anni? * Dropdown

Contrassegna solo un ovale.

- Abruzzo
- Basilicata
- Calabria
- Campania
- Emilia-Romagna
- Friuli-Venezia Giulia
- Lazio
- Liguria
- Lombardia
- Marche
- Molise
- Piemonte
- Puglia
- Sardegna
- Sicilia
- Toscana
- Trentino-Alto Adige
- Umbria
- Val d'Aosta
- Veneto
- ESTERO

81. Considerando il numero di abitanti, indica in quale fascia rientra il comune in cui hai vissuto la maggior parte del tempo prima dei tuoi 18 anni:

Contrassegna solo un ovale.

- Meno di 500
- Tra 500 e 2000
- Tra 2000 e 20.000
- Tra 20.000 e 100.000
- Oltre 100.000

82. In quale regione attualmente vivi in maniera più stabile? * Dropdown

Contrassegna solo un ovale.

- Abruzzo
- Basilicata
- Calabria
- Campania
- Emilia-Romagna
- Friuli-Venezia Giulia
- Lazio
- Liguria
- Lombardia
- Marche
- Molise
- Piemonte
- Puglia
- Sardegna
- Sicilia
- Toscana
- Trentino-Alto Adige
- Umbria
- Val d'Aosta
- Veneto
- ESTERO

83. Considerando il numero di abitanti, indica in quale fascia rientra il comune in cui attualmente vivi in maniera più stabile:

Contrassegna solo un ovale.

- Meno di 500
- Tra 500 e 2000
- Tra 2000 e 20.000
- Tra 20.000 e 100.000
- Oltre 100.000

84. Qual è il tuo livello di istruzione? *

Contrassegna solo un ovale.

- Licenza elementare
- Licenza media
- Diploma di scuola superiore
- Laurea triennale (o equivalente)
- Laurea magistrale (o equivalente)
- Master
- Dottorato di ricerca

85. Quanto senti vicinanza alla comunità LGBT+? [Valuta da 1 a 5] *

Contrassegna solo un ovale.

	1	2	3	4	5	
poca	◯	◯	◯	◯	◯	molta vicinanza