

CHAPTER 5

Homophobic hate speech in Italian tweets

Contextual cues of offensiveness

Stefano De Pascale

Vrije Universiteit Brussel, FWO & KU Leuven

Edoardo Cavarani

KU Leuven

Stefania Marzo

KU Leuven

Abstract

Social media channels have become omnipresent tools for communicating, sharing knowledge, and establishing communities, but also places where all sorts of hate speech comes to the surface. This chapter contributes to the body of work on online hate speech towards sexual orientation minorities and takes some initial steps towards a quantitative variationist sociolinguistic study of homophobic language in Italian social media. By exploring the different lexicalisations (i.e. near-synonyms) available in Italian

How to cite this book chapter:

De Pascale, Stefano, Edoardo Cavarani, and Stefania Marzo. 2024. 'Homophobic hate speech in Italian tweets: Contextual cues of offensiveness'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 159–191. Helsinki: Helsinki University Press. DOI: <https://doi.org/10.33134/HUP-27-5>.

for expressing or referring to the concept of HOMOSEXUAL MAN on X (formerly known as Twitter), this study provides both a descriptive and a methodological contribution. For this purpose, we created a dataset of 3000 manually annotated tweets in which at least one of the 33 lexicalisations of the concept HOMOSEXUAL MAN was recorded, and which we annotated for linguistic and stylistic factors that favour the perception of an expression as offensive. Our study shows that the interaction of lexical variation and explicit and implicit contextual cues, such as irony and dialect use, is significant in determining the degree of offensiveness of tweets. Furthermore, this study provides further evidence in favour of using social media as a laboratory for mapping language variation on a large scale, and for reflecting on the refinement of semi-automatic annotation of linguistic, stylistic, and social variables in written social media language.

Keywords: homophobia, lexical variation, contextual cues, conditional inference tree

5.1 Introduction

Since the massive rise of social media networks, online hate speech has been a widespread phenomenon, often targeting vulnerable minorities on the basis of their social, religious, or ethnic background, their gender identity, or their sexual orientation. This chapter focuses on hate speech towards the last of these groups and takes some initial steps towards a quantitative variationist sociolinguistic study of homophobic language in Italian social media. By exploring lexical variation in a large-scale corpus of terms referring to homosexuality on the social network X (formerly known as Twitter), this study provides both a descriptive and a methodological contribution. Given that there is increased evidence of the effects of hate speech on offline behaviour (see, among others, Soral, Bilewicz, and Winiewski 2018; Müller and Schwarz 2021), it is necessary to strive for a multidata approach to these issues. The current study, with its combination of corpus and attitudinal data, attempts to provide a contribution to such an endeavour.

On a descriptive level, the chapter offers an initial insight into lexical variation in the common terminology used on Italian social media to refer to homosexual people. For this purpose, a large-scale X corpus was compiled that includes expressions featuring the 33 most commonly used terms (henceforth called ‘near-synonyms’) referring to homosexual men. The aim of the study is not only to identify variation *per se*, but to examine to what extent the degree of offensiveness towards homosexuals correlates with the variation in terminology. To this end, we examine which linguistic and stylistic factors favour the perception of an expression as insulting.

In addition to providing a quantitative analysis of an X corpus, this chapter addresses a number of methodological issues. Although we are not the first to work with X data for the analysis of hate speech (Poletto et al. 2017; Sanguinetti et al. 2018; [Chapter 4](#) in this volume), our analysis aims to provide further support for the value of X data (and social media data in general) as a research tool for studying language variation (Bohmann 2016; Coats 2016; Grondelaers and Stuart-Smith 2021; Grondelaers and Marzo 2023). Specifically, this study provides further evidence in favour of using X as a laboratory for mapping language variation on a large scale, and for reflecting on the annotation of linguistic, stylistic, and social variables in written social media language.¹ In what follows, we will discuss the possibility of annotating stylistic aspects of spoken language that can also be found in social media language, and that seem to be highly relevant for explaining the construction of homophobic language.

In the remainder of this introduction, we will outline our research against the background of the public debate on hate speech in general; specifically, we will focus on the complexity of

1 Since the collection of the data in March 2021 and the drafting of the chapter in 2022, Twitter has changed name and owner (now being called X), and has severely limited the free functionalities of its API and the availability of its data. As a result of these restrictions, the microblogging service has partially lost its appeal as a provider of large amounts of informal written language.

pinpointing hate speech due to the multiplicity of contextual factors that play a role in the construction of offensive speech. In [Section 5.2](#), we describe our dataset and the methodological tools that we have used to study lexical variation in our data. We will also explain the annotation parameters for the detection of the stylistic and linguistic cues of offensiveness, as well as the automatic annotation of the sociodemographic characteristics of the X users. [Section 5.3](#) presents the results of our statistical analysis, which are organised in terms of the descriptive and methodological goals of the study. In the discussion and conclusion ([Section 5.4](#)), we flesh out the most striking results while reflecting on the limitations of this preliminary study but also on the possibilities that it has opened up for further research on the use of homophobic language in Italy.

In recent years, interest in the detection of online hate speech has steadily increased, due to the societal impact of the phenomenon since the rise of web content. In particular, the automation of hate speech detection has grown significantly. Natural language processing has become a primary method for detecting hate speech since it became clear that simple word queries did not provide sufficient insight into such a complex phenomenon. Indeed, hate messages appear to be determined not only by the use of explicit hate words and slurs but also by multiple contextual aspects such as ‘the domain of an utterance, its discourse context, as well as [...] co-occurring media objects (e.g. images, videos, audio), the exact time of posting and world events at this moment, identity of author and targeted recipient’ (Schmidt and Wiegand 2017: 1), in addition to more abstract semantic and discursive frames (for an example of those, see the discussion on heteronormative matrices which go beyond sexual identity in [Chapter 4](#) in this volume). Although it is controversial and generally condemned, hate speech is particularly multilayered and may not be directly observable at first sight. As Federico Faloppa puts it, observing and studying hate speech is therefore like ‘chasing the panther to which Dante, in *De vulgari eloquentia*, compares the illustrious vernacular: although you can smell it, it cannot be

grasped' (Faloppa 2020: 22, our translation). The boundaries of what can be called 'hate' are far from clear, and relatively subjective.

According to research since the mid-2010s (Bianchi 2017, 2021; Gao, Kuppersmith, and Huang 2017; Caselli et al. 2020; Faloppa 2020), there are two types of hate speech. The most obvious type involves utterances that contain explicitly derogatory terms—that is, terms referring to stereotypes, insults, or threats, or that clearly incite hatred or discrimination. In some cases, it contains explicitly the word 'hate' or other words that evoke hate (Orlando and Saab 2021). The degree of offensiveness of these words is widely recognised, and is almost objectively measurable, as confirmed by the extensive literature on swearing and slurs (see Jay and Janschewitz 2008, Beers Fägersten and Stapleton 2017).

Much more complex and less studied are the implicit verbal manifestations of hate, conveyed by words that are not necessarily derogatory in and of themselves (see chapters 4 and 7 in this volume). This type of hate speech often implicitly contains, suggests, or builds on offensive associations which are difficult to pinpoint (see, e.g., [Chapter 4](#) in this volume for the role played by a set of semantic frames in homotransphobic speech in an X corpus that does not contain explicit incitement of hatred). There is generally less consensus about the impact of this form of hate speech, and it is the context in which it occurs that seems to determine its degree of offensiveness.

As these implicit forms of hate speech are less evident, they are also much more complex to investigate. Natural language processing has been investigating the role of contextual factors in the construction of hate speech for years. Many aspects were found to play a role in the development of offensive or aggressive language use, for example linguistic features such as imperatives or specific syntactic constructions. However, a detailed study of the role of finer-grained linguistic features in hate speech has not yet been carried out (whereas the role that some of these features—specifically, 'deviations' from a canonical grammatical structure in phonology, morphology, and syntax—play in the encoding of

emotion/expressivity has been given some attention; see Corver 2016; Saab 2021, 2022; see also Trotzke and Villalba 2021).

This chapter explores the difference between explicit and implicit hate speech in reference to homosexual men (see [Chapter 1](#) for a comparison with the distinction between ‘hard’ and ‘soft’ hate speech, particularly in a legal context). It investigates the role of the linguistic context in the creation of homophobic hate speech and aims to better understand the grey area in which apparently neutral terms nevertheless lead to hate messages. We therefore use quantitative methods to investigate which semantic, structural, and stylistic elements operate together in order to construct hate speech towards homosexuals.

Our work can thus be seen as complementary to Safina’s study ([Chapter 4](#)) in this volume. Safina uses X data on homotransphobia and aims to go beyond lexical evidence, but, unlike our study, she carries out a qualitative study that tries to understand ‘which semantic spheres are involved in stereotypical representations of the queer community’ and to detect which word clusters possibly construct these semantic spheres. Crucially, she shows that the most significant clusters are not related to words expressing desire, sexuality, or erotic practices, but rather to Politics (i.e. law and criminality, freedom and fundamental rights, and immigration), Nature (i.e. health, diseases, and dehumanising references to animals), and Values and Customs (i.e. morality, family, and religion).

5.2 Materials, data, and methods

5.2.1 Data collection and corpus

In order to investigate online homophobic language, we decided to look at X data.² X represents a relevant type of digital social space, which people across the world use to communicate, share

2 We have made the instructions to annotate the tweets, the sample of annotated tweets, and the data for the perception survey publicly available through the OSF-project ‘Homophobic hate speech in Italian tweets: contextual cues of offensiveness’, retrievable via <https://osf.io/4ftq8/>.

knowledge, and establish communities (Fuchs 2014), but also where all sorts of anti-social behaviour comes to the surface relatively unfiltered. The verbal manifestation of this behaviour, hate speech, has been studied in detail by computational linguists (Schmidt and Wiegand 2017). One of the reasons for its popularity in this field is that X has for a long time made the data extraction process very easy through its API (Application Programming Interface; but see [note 1](#) for the current state of affairs), which we also take advantage of in our study. However, tweets are also popular in variationist research, not only because they are available in enormous quantities but also because they are an extremely rich source of informal language use with features mirroring orality through several compensation techniques, such as variant spellings or expressive compensation strategies (repetition of characters, excessive use of interpunctuations, etc.). This non-standard language use and these orthographic compensations are seen as a 'form of identity, signaling authenticity, solidarity of resistance to norms imposed from above' (Eisenstein 2013: 362, cited in Grondelaelers and Marzo 2023). For this reason, X has been found to be highly suitable for the investigation of language variation and the indexical meanings attached to it, including, as in this case, expressions of hate or offensiveness.

For this study, we first determined the lexical field consisting of all the near-synonymous lexicalisations of the concept `HOMOSEXUAL MAN`. For two main reasons, we have opted to focus exclusively on terms referring to homosexual men, and not on women. This is partially for reasons of feasibility, but we also have a theoretical-methodological justification for this choice. While it might be arguable, we found that homosexual men and women do not belong to the same conceptual category from a linguistic point of view. Methodologically, it is therefore not completely appropriate to bundle them together under one unique conceptual 'flag': it would have been odd to have a concept `HOMOSEXUAL PERSON` with the subordinate concepts of `HOMOSEXUAL MAN` and `HOMOSEXUAL WOMAN`. We have therefore opted to consider them as distinct concepts.

Table 5.1: Near-synonyms of the concept HOMOSEXUAL MAN.

Near-synonyms (Treccani)
<i>bardassa, buco, checca, cinedo, culattone, culorotto, cupio, finocchio, frocio, gay, invertito, omo, omofilo, omosessuale, omosex, paraculo, pederesta, recchione/ricchione, sodomita, uranista</i>
Near-synonyms (Wikipedia)
<i>bucaiolo, buggerone, buliccio, caghinero, ciucciacazzo, ciucciapisello, gar-rusu, matellu, piglianculo, pivellu, puppu, rottinculo, succhiacazzo</i>

In order to obtain the most exhaustive list possible, thereby adhering to the principle of accountability (a cornerstone of variationist sociolinguistics; see Tagliamonte 2006), we primarily consulted two types of resource. First, we skimmed through the online Treccani synonyms and antonyms dictionary.³ The procedure for finding all the near-synonyms was carried out recursively: we i) started out with the lemma *gay*, ii) recorded all the synonyms listed under that entry, and then iii) searched for each synonym lemma individually, so as to retrieve the synonyms of the synonyms of *gay*. This search continued until no more new synonyms could be retrieved. We then consulted the Italian Wikipedia (n.d.) entry ‘Lessico dell’omofobia’ (The lexicon of homophobia) as a complementary resource, in order to retrieve regionally marked words that were not included in the Treccani dictionary. In total, we collected 33 near-synonym nouns for this concept, which are listed in [Table 5.1](#).

A quick glance at the two lists shows, unsurprisingly, the different focus and coverage of the two resources. Whereas the Treccani list covers a diverse set of properties, the Wikipedia list provides near-synonyms that are either regional or with a very negative connotation. Based on the stylistic and sociolinguistic labels in the Treccani, it is possible to identify roughly four qualitatively

3 The dictionary can be accessed online at https://www.treccani.it/enciclopedia/elenco-opere/Sinonimi_e_Contrari.

different groups of near-synonyms: vulgar and derogatory terms; regional or dialectal terms; archaic and uncommon terms; literary terms. As some near-synonyms receive multiple labels, we have summarised the relationships in the full lexical field in [Figure 5.1](#). Given that the geosynonyms taken from Wikipedia are all consistently considered derogatory, they are shown as nested within the pool of vulgar terms. There is also wide variation in the origin of the regional terms: *matellu*, *garrusu*, and *puppu* are found in Sicily; *caghinero* and *pivellu* in Sardinia; *checca*, *bucaiolo*, and *finocchio* are considered to have originated in Tuscany; *buliccio* in Liguria; *frocio* in Rome; *ricchione* in Southern Italy; whereas *culattone* and *cupio* seem to come from northern Italian varieties. Some of these terms (e.g. *checca*, *finocchio*, and *frocio*) no longer have a specific regional distribution and might thus be considered pan-Italian.

Some terms, which are also the most frequently used, have no label and are classified in [Figure 5.1](#) as a superordinate category of ‘neutral’ terms: *gay*, *omosessuale*, *omosex*, *pederasta*, and *sodomita*. The term ‘neutral’ refers in the first place to those near-synonyms that do not receive any sociolinguistic or stylistic label in the given source (Treccani or Wikipedia). For the first three terms, we can indeed expect that the lack of such a label genuinely represents the unmarked nature of these near-synonyms. Furthermore, *gay*, *omosessuale*, and *omosex* will be particularly relevant for our upcoming analyses: as they are not inherently negatively connotated, in order to use them in a derogatory way, speakers have to resort to strategies involving the manipulation of their sentential context. However, with respect to the latter two words in this category, *pederasta* and *sodomita*, the lack of label is probably more due to specific lexicographic choices than to a truthful representation of the sociolinguistic status of these near-synonyms.

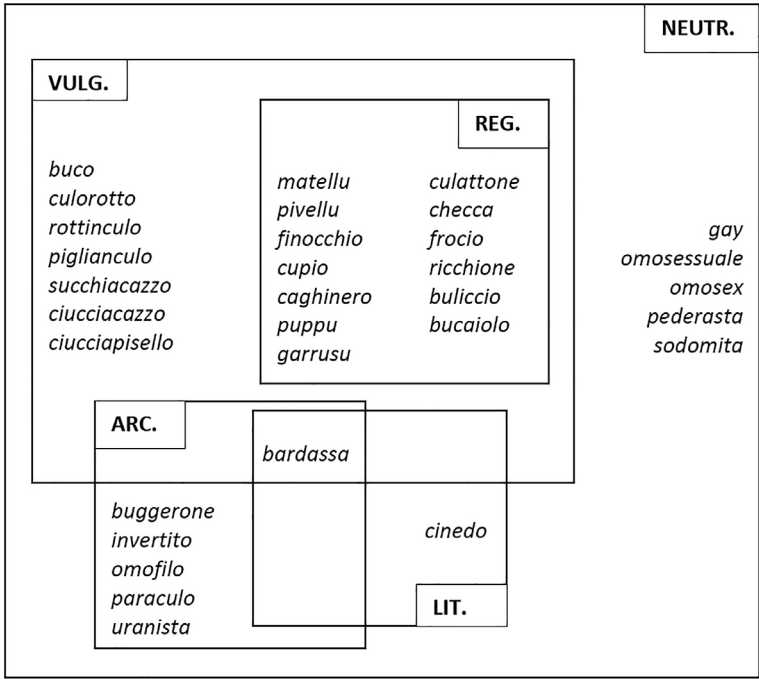


Figure 5.1: Qualitative classification of the near-synonyms, in terms of vulgarity (VULG.), regionality (REG.), archaism (ARC.), literary use (LIT.), and neutrality (NEUTR.).

For each of these terms we set up a search making use of X's API. Given the lack of elaborate wildcard possibilities in the API, or basic lemmatisation, we retrieved both plural and singular forms of the terms separately. We searched for tweets in which all of the 33 near-synonyms appeared from 2006 (when the site now known as X was launched) up to 2020, making a total of 14 years of tweets. This resulted in a database of more than 2.6 million tweets in which one of the lexicalisations of HOMOSEXUAL MAN was used. From this collection we randomly sampled 3044 tokens—occurrences of each of the near-synonyms—for further inspection and annotation, such that two conditions were met: i) the relative proportion (in terms of frequency) of the near-synonyms in the dataset sampled for annotation would mirror the

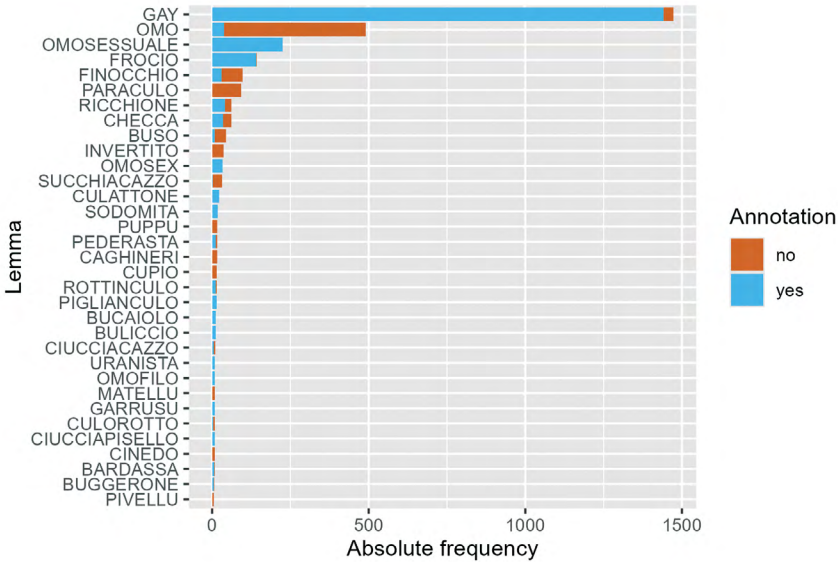


Figure 5.2: Absolute frequencies of near-synonyms for the concept HOMOSEXUAL MAN.

one of the larger database as much as possible, and ii) at least five tokens per near-synonym would be sampled. [Figure 5.2](#) plots the set of near-synonyms for which tokens were found, in descending order of frequency. Almost half of the sampled tokens are occurrences of *gay* (1442), which is therefore by far the most frequent lexicalisation of the concept. Not taking into account *omo* (which is mainly used in a different sense; see below), the second most frequent lexicalisation is also a connotationally neutral term, namely *omosessuale* (225). The third most frequent is the derogatory *frocio* (141), which appears more than three times as often as the next most frequent terms, *ricchione* (41) and *checca* (35). Unsurprisingly, the whole lexical field of HOMOSEXUAL MAN follows a Zipfian word frequency distribution.⁴

⁴ When generating frequency lists of word types in a corpus, one typically encounters a so-called Zipfian frequency distribution (named after George K. Zipf [1949]). The fundamental property of such a distribu-

5.2.2 *Neutral and offensive near-synonyms and the role of context*

Before we began our quantitative analysis of the impact of linguistic and stylistic factors on the degree of offensiveness, we explored the degree of offensiveness of these near-synonyms in a preliminary perception test. The aim of this pilot test was to better understand the difference between supposedly neutral terms and offensive near-synonyms and to assess the specific role of linguistic and stylistic context among native Italians. We therefore administered an online perceptual test (in Qualtrics) to 49 Italian respondents, all native speakers (with 20 men, 26 women, and 3 respondents who indicated no specific gender). Each was presented with a set of the five most common near-synonyms referring to homosexuals, those that occurred most frequently in the corpus: *gay*, *omosessuale*, *ricchione*, *culattone*, *frocio*. The first two near-synonyms were supposed to be less offensive (and more commonly used), whereas the other three were supposedly more directly and inherently offensive. The five near-synonyms were alternately presented separately (i.e. without context) and in a context containing linguistic (morphological, dialectal) or stylistic (ironic) elements that contributed to making the term more expressive and possibly offensive. For each tweet, respondents had to indicate the degree of offensiveness on a scale from 0 to 10.

[Figure 5.3](#) shows the relative (average) values given by respondents for each term, both in a context without linguistically or stylistically expressive cues and in one with those cues. The results of this perception experiment clearly showed that *gay* and *omosessuale*, the supposedly unmarked terms, can also be perceived

tion is its skewness: a few word types are extremely frequent (usually such lists are topped by closed class types such as prepositions, pronouns, and articles) while the vast majority of the remaining word types are very infrequent, mainly occurring just once or twice (mainly open class items). The lexical field under investigation here also follows this type of distribution, having one very frequent lexicalisation (*gay*) and a long tail of infrequent near-synonyms (occurring just a few times). [Figure 5.2](#) clearly displays the skewness in the frequency counts.

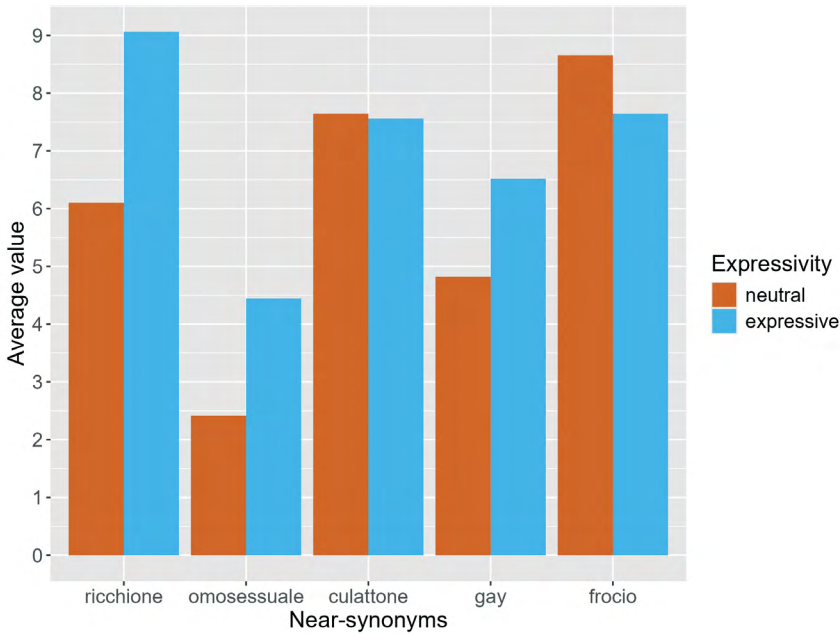


Figure 5.3: Average scores of offensiveness of five near-synonyms for homosexual uses in neutral or expressive contexts.

offensively in a linguistically and stylistically charged tweet: the averages for these terms were clearly different depending on the context. For the other three terms—the supposedly inherently offensive *ricchione*, *culattone*, and *frocio*—there was less noticeable difference between the two contexts, although there still was some difference noticed for *ricchione*. This means that these terms might be perceived as inherently offensive regardless of the contextual elements in which they are used. This pilot test served as preparatory work for the annotation and analyses we conducted on a larger scale with the X corpus (see following sections) and as a validation of our starting hypothesis that hate speech, and specifically homophobic terminology in this case, is strongly dependent on the linguistic context in which it occurs.

5.2.3 Annotation of the context

The next step involves the annotation of the tokens. Two native speakers of Italian were recruited for this task and both annotated the set of sampled tokens. Each tweet was annotated for linguistic and sociolinguistic parameters. The primary goal of the annotation was to provide grammatical and semantic disambiguation for the usage of the near-synonyms. This data-cleaning step principally involved three filters.

First, as we were only interested in uses of the words that refer to a person, which largely corresponds to nominal uses of the terms, we discarded all non-noun tokens, particularly adjectival uses that refer to the quality of being homosexual. Some of the lexemes in the list, such as the most frequent *gay* and *omosessuale* (e.g. *un atteggiamento gay/omosessuale*, ‘a gay/homosexual attitude’), can in fact be used both as nouns and as adjectives. However, the majority of the terms, especially the more derogatory ones, can only be used as nouns (e.g. *un atteggiamento *checca/*frocio/da checca/da frocio*, where asterisks indicate ill-formedness). The removal of adjectival use is particularly important for the term *invertito*, which is often found as a past participle of the verb *invertire*, ‘to turn, to exchange’. Similarly, we had to discard some hits of *checche*, as they turned out to be accent-less instances of the indefinite pronoun *checché*, ‘whatever’ (social media orthography is often characterised by such spelling). The second cleaning step involves cases of homonymy, such as the case of *finocchio*, which mostly refers to the ‘vegetable fennel’. A lexeme such as *paraculo* never refers to a homosexual man in our sample, but rather to ‘an opportunist, someone who can skilfully and casually turn a situation in their favour’. Therefore, this lexeme was also discarded. The majority of *omo* tokens are not instances of the meaning of the concept we are interested in here, but are occurrences of the Roman dialectal variant of *uomo* ‘man, mankind’. In this step we also removed instances of lexemes that refer to women rather than to men. A third class of discarded tokens are those in which it was impossible, within the context of the tweet, to ascertain what the

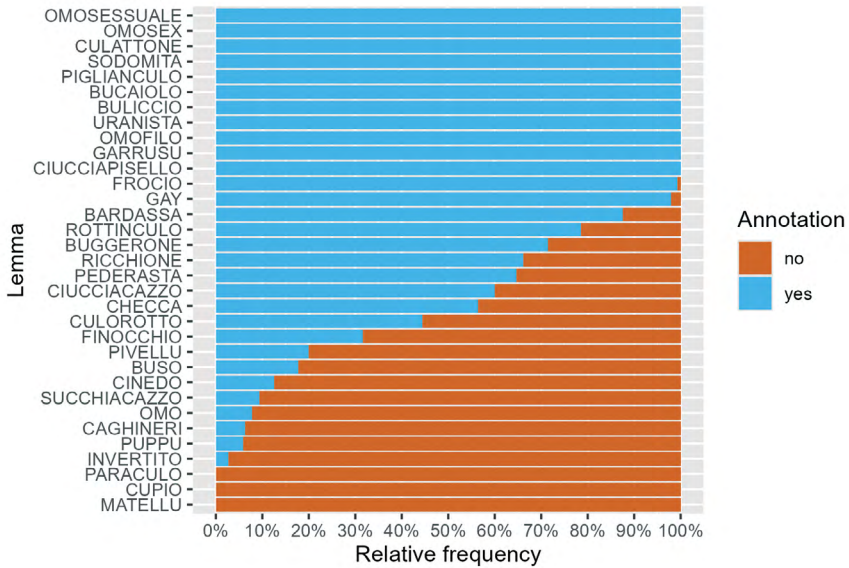


Figure 5.4: Proportions of ‘in-concept’ (blue) vs ‘out-of-concept’ (orange) tokens, ordered in decreasing order of ‘in-concept’ tokens.

word referred to. Given that we had not filtered out tweets shorter than a specific number of words, nor tweets that do not have well-formed syntactic structure, some tweets did not provide enough clues to allow us to disambiguate them.

At the end of this cleaning and disambiguation procedure, we were left with 2159 tokens (around 70 per cent of the dataset). [Figure 5.4](#) summarises the proportions of ‘out-of-concept’ occurrences (i.e. the tokens that did not refer to *HOMOSEXUAL MAN*) vs ‘in-concept’ occurrences for each near-synonym (i.e. all tokens that did refer to *HOMOSEXUAL MAN* and that were retained in the corpus). This plot shows the importance of a disambiguation and data-cleaning step for many of the lexemes in our dataset, without which the quantitative results would very likely be distorted.

The second part of the annotation consisted in the identification of other characteristics of the tweet that are considered relevant for the analysis of hate speech in social media. The annotation

scheme was modelled on that in Poletto et al. (2017), who also annotated a dataset of Italian tweets (for computational-linguistic purposes; see [Chapter 4](#) in this volume for a similar annotation scheme). In the following we describe the annotation scheme we provided to our two annotators. The first three parameters are taken from Poletto et al. (2017). The first is ‘degree of offensiveness’. In contrast to aggressiveness, this parameter focuses more on the potentially offensive effect of the content of the tweet. According to this parameter, a tweet is considered weakly offensive in most cases. These are cases in which, for example, the target is associated with a certain characteristic (biological, sociological, behavioural, etc.) that emphasises the target’s status as a disadvantaged or discriminated minority, or cases in which a description of the target is proposed that qualifies the target as an unpleasant person; on the other hand, if overtly offensive language is used, or if the target is addressed with outrageous or degrading expressions, the tweet is considered highly offensive. The coding of the degree of offensiveness consisted of three values: 0 (neutral content), 1 (mildly offensive content), and 2 (highly offensive content). For all the other features, the coding is binary: either the presence or the absence of the feature.

- (1) ma che cacchio dice?? vuol attirare l'attenzione con argomenti da invasato finocchio culattonne!! sparisci. [highly offensive]

‘what the heck is he saying?? he wants to attract attention with possessed gay faggot arguments!! get lost.’

The second parameter is ‘irony’. This parameter determines whether the tweet is ironic or sarcastic rather than being based on the literal meaning of words. The third is ‘stereotype’, which determines whether the tweet contains implicit or explicit references to (mostly false) beliefs about a given target.⁵

5 It is notoriously hard to generate reliable and consistent annotations for the abovementioned variables, due to the high degree of subjectivity involved and the difficulty in reaching an operational definition

- (2) Capisco gli uomini che diventano gay perché si rendono conto che è meglio prenderlo al culo da cialtroni frustrati come loro. Le riflessioni geniali della domenica sera. Buonanotte Twitter.

‘I understand men who turn gay because they realise it is better to take it up the ass from frustrated wafflers like them. Sunday night genius musings. Goodnight Twitter.’

The following parameters are original additions for our specific study. The fourth is ‘unconventional spelling’, and refers to the presence of unconventional punctuation or unexpected spelling arguably translating unexpected/emphatic pronunciations (e.g. vowel or consonant lengthening or use of capital letters).

- (3) GAY SI NASCE di solito forse boh chi lo sa ED IO NON LO NAQQQUIII.

‘ONE IS BORN GAY usually maybe eh who knows AND I WASN’T BORN LIKE THAT.’

The fifth concerns ‘dialectal words’, namely the use of words that are not part of (neo)standard Italian and can be considered dialectal or markedly regional, as in

- (4) @USERNAME ‘c’è la crisi e te stai a pensa’ ai matrimoni de’ li froci’

‘@USERNAME “there is crisis and you are thinking about gay weddings”’

The next three parameters are related to morphosyntax. The first concerns ‘dislocations’, namely the presence or absence of right- or left-dislocation of the nominal phrase containing the near-synonym, or the pronoun referring to it. The relevance of this parameter

of those notions (Sanguinetti et al. 2018). For degree of offensiveness, irony, and stereotype, we find inter-annotator agreement values, measured with weighted Cohen’s K, between 0.14 and 0.42, which indicate only fair agreement overall.

for the detection of hate speech builds on: i) the assumption that ‘dislocations’ are ‘deviations’ from the (pragmatically) unmarked constituent order that bring about a change in information structure and have a marked pragmatic function (Lambrecht 1994), and ii) the hypothesis that such a function might be that of signalling a negative emotive attitude on the part of the speaker (Fónagy 1995; Oliveira 2013). The second is ‘derivational morphology’, namely the presence or absence of derivational morphemes (diminutives, pejoratives, reduplication) on nouns, adjectives, and adverbs. The final is the presence or absence of deictics such as pronouns and demonstratives, which can be employed to emphasise social distinctions and, crucially, the attitude of a speaker towards a specific referent, such as a person or a group of people (Hart 2010, 2014; Da Milano 2016).

- (5) Bene è l'ora della buona notte, etero gay trans che voi siate vi amo quasi tutt*! #noeterosessualità ahaahah

‘Well it’s night time, straight gay trans whatever you are I love almost all of you! #noheterosexuality hahaha’

Finally, two demographic parameters are reliant on the information provided by the X API. One is the gender of the user, which is not given by X users as such in their profile but can be inferred from their username or user screen name. The assumption was that the username would be a good indicator of the gender of the user, which was coded as either male, female, or ‘not available’ (when the username did not allow the recognition of a gender). The other demographic parameter is ‘geographical origin’, which was based on the information available in the ‘place_country’, ‘place_name’, or ‘user_location’ tags of the output. If a real, concrete place was mentioned, the tweet was coded with the Italian region the place was located in. Together with the semantic and grammatical disambiguation, a total of 12 properties were annotated.

Based on this dataset, we carried out two sets of analyses. In the first analysis, we explored the impact of the annotated parameters—linguistic, stylistic, and demographic—of the tweets on the

distribution of our near-synonyms ([Section 5.3.1](#)). The second analysis adopted the opposite perspective and investigated which of the properties of the tweet create ‘offensive’ tweets ([Section 5.3.2](#)). In variationist and statistical parlance, in the first analysis, ‘degree of offensiveness’ corresponds to the ‘predictor variable’ of the analysis, while the near-synonyms are the ‘response variable’. Both the predictor and response variables are categorical, with the response variable following a multinomial distribution. In the second analysis, ‘degree of offensiveness’ corresponds to the response variable, and the distribution of near-synonyms is part of the set of predictors.

In both cases, the statistical technique we adopt is *conditional inference tree*. This technique is part of the family of classification tree methods, and offers a few advantages over similar inferential methods: first, its output is mostly visual (in the form of a tree structure), which makes an analysis of the patterns easier to interpret for users with a more limited knowledge of statistics; second, it is especially useful for reporting on how multiple features cooperate in the selection of the near-synonyms (or, in statistical parlance, how the predictors interact); and third, it is in principle more robust against the expected correlation between the features and the unbalanced distribution of both our response variable and the predictors. As our near-synonyms follow a skewed, Zipfian distribution, and the majority of the tweets received a ‘0’ or ‘no’ coding for many features, this is a particularly relevant property of conditional inference trees.

In brief terms, the technique works as follows. Firstly, independence tests between each predictor and the response variable are performed and the predictor that covaries most strongly with the response variable gets selected. In the following step, the technique splits the data into two subsets according to the levels of this selected predictor and then once again tests all predictors as before on these separate partitions of the data. This procedure is repeated until no further splits are justified by the independence tests (in other words, until no statistically significant patterns remain). The result is a flowchart-like decision tree, with

the strongest discriminative predictor at the top of the tree, and recursive splitting by other predictors generates a hierarchy of interacting predictors. The leaves at the end of the branches of the tree are barplots that show the distribution of the response variable values for that subset of predictor variables selected by means of the interactions (see Strobl, Malley, and Tutz 2009; Tagliamonte and Baayen 2012; and Gries 2020 for an introduction, but also for the drawbacks of this type of technique). For the analysis of our dataset, we made use of the implementation of inference trees available in the `party` and `partykit` packages in R.

5.3 Analyses

5.3.1 Analysis with ‘near-synonyms’ as response variable

We start with an analysis of the conditional inference tree that has the set of near-synonyms as the response variable. This should help us to understand which properties of the tweet (i.e. the predictor variables) drive the distribution of those synonyms. This analysis has two components: the first focuses on the 9 most frequent near-synonyms, while the second focuses on the 3 most frequent ones: *gay*, *omosessuale*, and *frocio*. There are two reasons why it is reasonable not to model the full set of near-synonyms. The first is methodological, and comes as a consequence of the Zipfian distribution of the lexical field: with most of the words having very low frequencies, it is impossible to carry out a statistical analysis that involves some type of interaction. Recall that the conditional inference tree algorithm functions with increasingly smaller subsets. As a consequence, the infrequent items would quickly fall out of the picture. We maintain that looking at 9 near-synonyms instead of 32 strikes a good balance between variation and feasibility.

For the analysis of the three most frequent items, the analysis becomes statistically even more robust. Moreover, we consider each of the three items to have a specific ‘sociolinguistic’ profile. *Gay*, the most frequent item, is a loanword; *omosessuale* is the

standard Italian counterpart with the same neutral connotational value as *gay*; and *frocio* is the most frequent regionally marked and derogatory item. In a sense, *frocio* can be seen as the exponent of the range of regional offensive terms and can be argued to cover the other terms in the dataset.

We start with the conditional inference tree of the nine most frequent items: *gay*, *omosessuale*, *frocio*, *ricchione*, *omo*, *checca*, *omosex*, *finocchio*, and *culattone* (Figure 5.5). The sociolinguistic predictors ‘gender’ and ‘region’ do not seem to have a significant impact, nor do the presence of stereotypes, dialectal terms, unconventional orthography, derivational morphology, or the use of dislocations. All these predictors are excluded from the plotted conditional inference tree.

Unsurprisingly, the most determinant predictor is ‘degree of offensiveness’, which sets apart non-offensive tweets from offensive tweets (regardless of the magnitude). The second most important predictor is ‘irony’, which only has an impact on the non-offensive tweets. And within the group of ironic tweets, the use of deictics also has an influence on the distribution of the near-synonyms. When looking at the four barplot figures (i.e. the ‘leaves’ of the tree) from left to right, a natural cline can be observed from the most neutral and least expressive tweets (i.e. the inoffensive and unironic tweets) to the most charged and expressive tweets (i.e. the most offensive ones and those with an ironic undertone and/or deictic forms). With this cline of ‘expressivity’ comes a marked change in the distribution of the nine near-synonyms: neutral tweets follow the expected overall distribution, with a large predominance of *gay*. The more a tweet becomes expressive (and here ‘offensive’), the smaller the proportion of *gay* tweets (i.e. the bars become smaller) and the higher the lexical diversity.

This relation between affect and lexical variation has been attested in previous lexicological research (see, among others, Franco et al. 2019), showing that negative connotations boost lexical variation and the use of less frequent and more marked alternative words. However, in the absence of a clear negative connotation (in the left branches of inoffensive tweets), X users resort

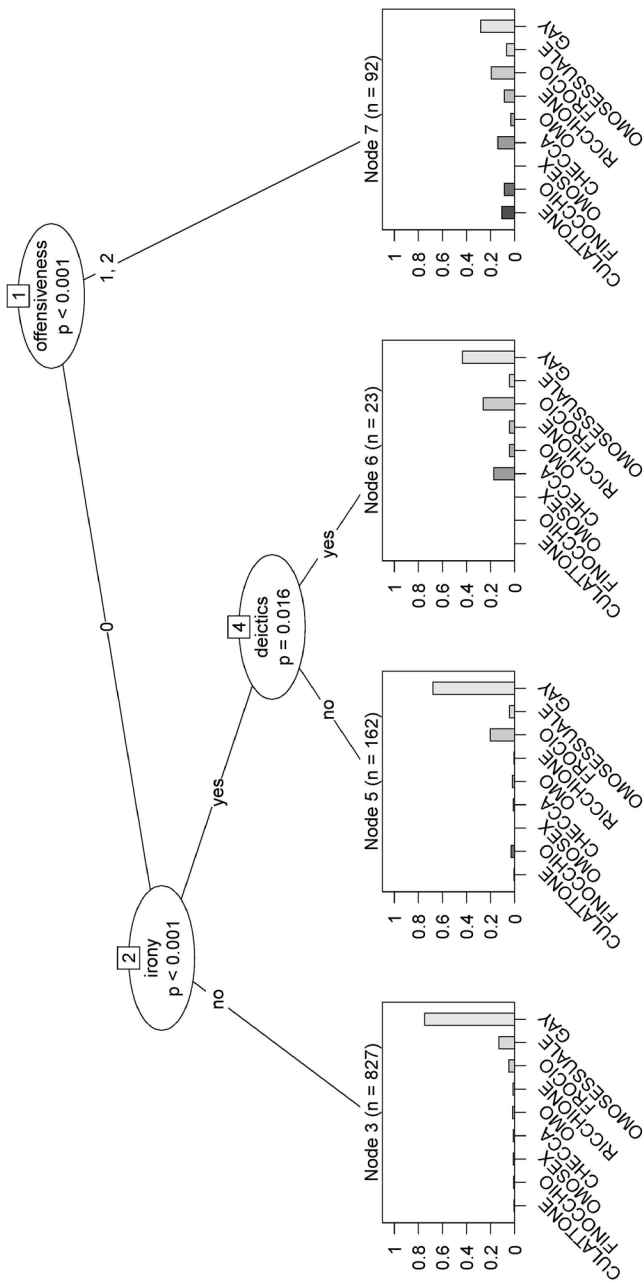


Figure 5.5: Conditional inference tree plot of the impact of the features on the distribution of the nine most frequent near-synonyms.

to other stylistic strategies that trigger the use of marked lexicalisations, such as irony and deictics, which tend to reinforce one another (in the interaction, Node 6).

After the nine near-synonyms, we focus on the three most frequent ones (*gay*, *omosessuale*, and *frocio*) in the conditional inference tree shown in [Figure 5.6](#). Restricting the lexical variation brings about a reordering of the importance of predictors, and the addition of new ones. While the presence of irony and offensive language remains influential, we see an increased relevance of the gender assigned to the X user and the presence of dialectal terms. In the left branches of the tree, where we find the unironic tweets, we observe a similar pattern as in [Figure 5.5](#): in neutral tweets, speakers largely favour the use of *gay*, while in more offensive but unironic tweets, there is a decrease in *gay* which is coupled with an expected increase in the derogatory term *frocio*. The right branching of the tree reveals new information. The three rightmost barplot groups (nodes 6, 8, and 9) show a similar cline as that which was attested in the previous analysis, but this time structured along completely different predictors. Irony tweets written by female users (Node 6) follow the general pattern of stark preference for *gay*, although there is a noticeable increase in the derogatory *frocio*. On the other hand, ironic tweets written by male users largely deviate from the overall distribution. When those tweets do not contain dialectal terms, *gay* is still the majority term, but *frocio* accounts for a third of occurrences. However, when male X users insert dialectal terms into their tweets, *gay* becomes the minority lexicalisation, and instances of *frocio* and *omosessuale* are greater. This is also the first time *omosessuale*, which has a similar neutral connotational value as *gay*, is more frequently attested than the loanword. The majority lexicalisation is yet another term, *frocio*. In sum, ironic tweets containing dialectal features written by men, as opposed to women, use the derogatory term twice as much as the overall most frequent near-synonym *gay*. Two observations should be made with regard to these results. First, the preponderance of the indigenous Italian words, and especially *omosessuale*, with respect to the loanword

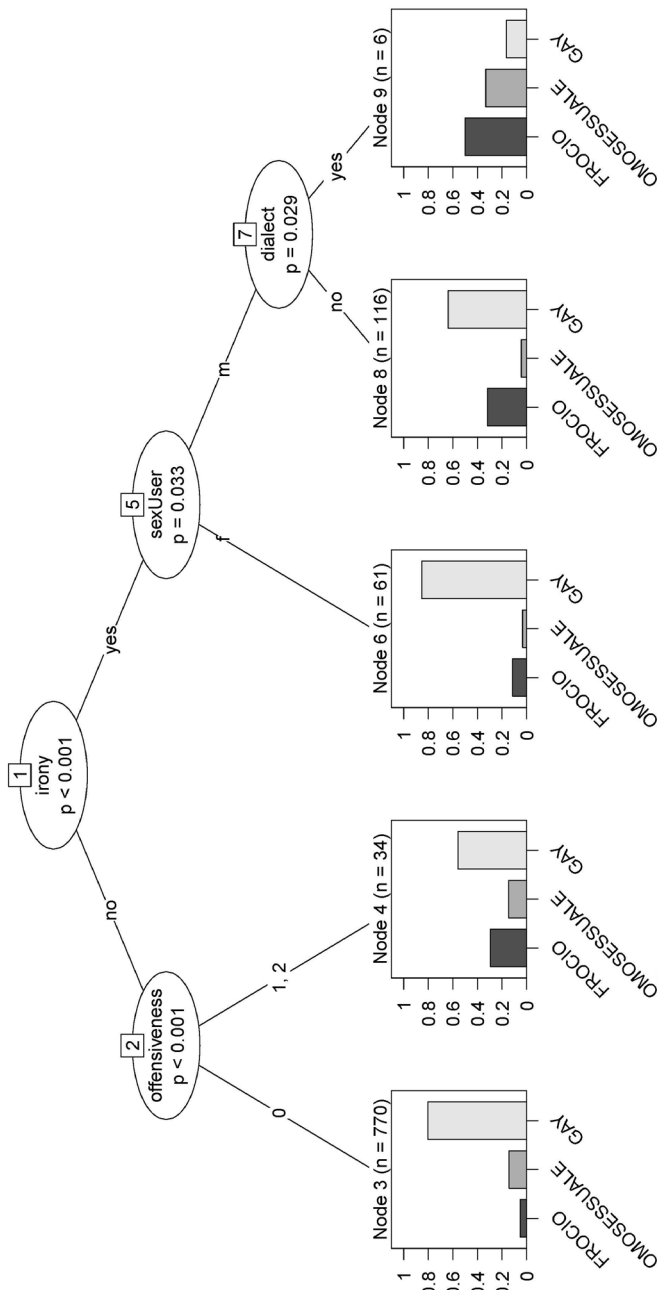


Figure 5.6: Conditional inference tree plot of the impact of the features on the distribution of the three most frequent near-synonyms.

gay, in tweets that contain some dialectal material, might indicate that users tend to be linguistically coherent in their production. *Gay* might be perceived as the more modern, or perhaps more written-language alternative, whereas *omosessuale* could be more easily integrated into an utterance which is either closer to the regional or dialectal speech of the user, or is embedded in a more informal conversation or expression on X. Further research into the nuances between those two terms will hopefully shed more light on this division of labour. Second, in the right branching of the tree (i.e. in all the ironic tweets) ‘degree of offensiveness’ no longer plays a major role. This means that among the abovementioned tweets one can find both offensive and inoffensive tweets, and that the dialectal tweets written by men are not necessarily offensive. This leads to the observation that the relative preference for the Italian terms over the loanword *gay* is not only a matter of contextual and communicative expressivity, but rather a stable property of the subcode of this sociolinguistic group of writers/speakers—that is, men writing in dialect. We have to stress at this point that the number of tweets being ironic, written by men, and with the inclusion of dialectal words, is very small: only six (i.e. Node 9 on the very right of the tree only contains six tweets). These last observations therefore cannot be generalised with the same confidence as those made on the other nodes of the conditional inference tree.

5.3.2 Analysis with ‘degree of offensiveness’ as response variable

The second analysis adopts the opposite perspective and looks at the factors that influence the ‘degree of offensiveness’ of a tweet, thus taking offensiveness as the response variable. This time the three-way distinction in ‘not offensive’, ‘mildly offensive’, and ‘very offensive’ (see [Section 5.2.3](#)) has been reduced to a binary distinction—offensive (labelled ‘yes’ in the conditional inference tree) vs inoffensive (labelled ‘no’)—in order to determine what contributes

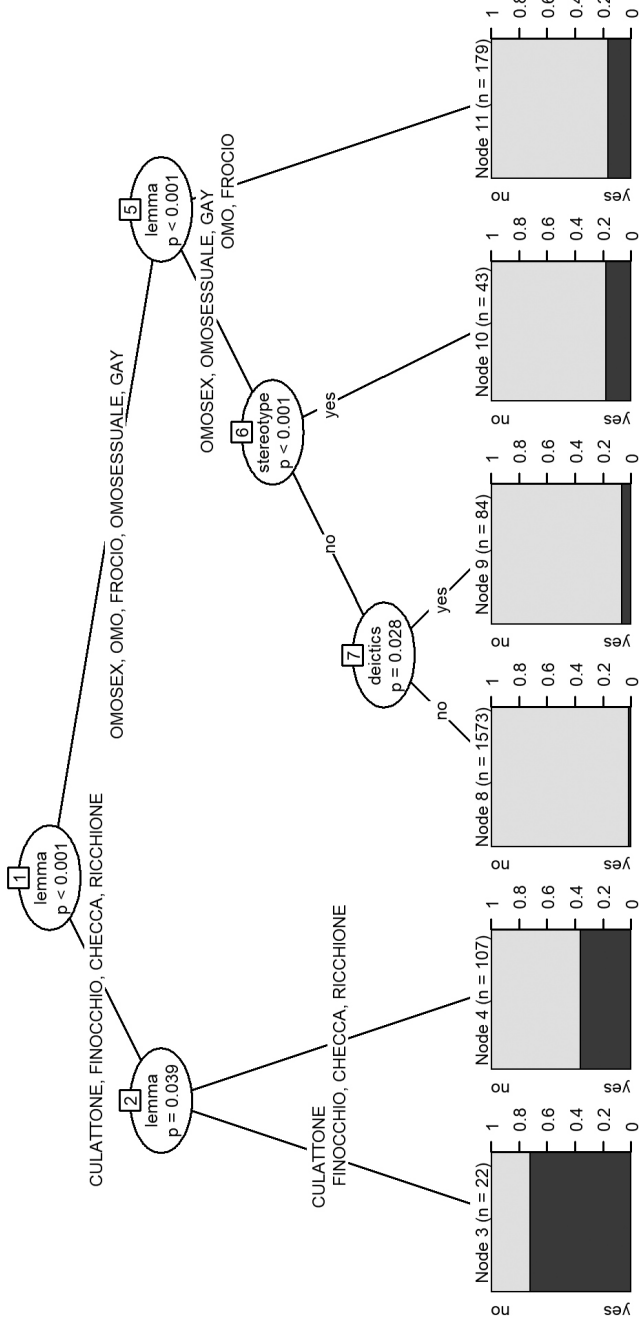


Figure 5.7: Conditional inference tree plot with offensiveness as a response variable.

to the offensive character of a tweet about a HOMOSEXUAL MAN or in which a HOMOSEXUAL MAN is mentioned.

In this case, we constructed a conditional inference tree with the near-synonym set as a predictor variable and offensiveness as a response variable. The same series of predictors as used in the previous analysis is included. What could we learn from turning a response variable into a predictor variable, and vice versa, with all the other predictors being retained? An interesting feature of conditional inference trees is that they can split predictors with many levels into groups of levels that behave homogeneously with respect to their impact on the response variable. In our case, ‘near-synonym set’ is such a multilevel predictor (with nine levels). So, while in the first analysis (in which the near-synonym set was a response variable) every level of that response was treated individually and mutually independently, we can now allow levels to be grouped together to assess their commonalities. In other words, we will be able to identify which near-synonyms are used in a similar way in the dataset and which of them lead to offensive tweets and under which contextual conditions.

In [Figure 5.7](#), we plotted the conditional inference tree with ‘degree of offensiveness’ as the response variable. This tree clearly shows that the nine different near-synonyms form several groups, as ‘lemma’ is now also the most important predictor. At the top of the tree, the first split takes place between the tweets containing *culattone*, *finocchio*, *checca*, and *ricchione* on the one hand, and *omosex*, *omo*, *frocio*, *omosessuale*, and *gay* on other hand. Before examining the interaction structure of the predictors in the branches, we can observe that in the stacked barplots of the leaves there is a non-trivial amount of variation between offensive and inoffensive tweets. This means that not all inherently derogatory terms in the first group automatically lead to or belong to offensive tweets, and not all neutral near-synonyms in the second group appear in inoffensive tweets. The context in which those words function will likely employ other compensatory or boosting strategies to either decrease or increase the expressive (and negative) effect of the tweet.

The barplots reveal the following overall pattern: *culattone*, *finocchio*, *checca*, and *ricchione* are associated with a relative majority of offensive tweets (nodes 3 and 4), while *omosex*, *omo*, *frocio*, *omosessuale*, and *gay* (nodes 8 to 11) are associated with far less offensive tweets. As the conditional inference tree further splits and diversifies, we can observe further granular distinctions within these two higher-order groups. In the most offensive group, *culattone* splits from *finocchio*, *checca*, and *ricchione*. The vast majority (73 per cent) of the tweets in which *culattone* appears are offensive, versus a sizeable minority (36 per cent) of the tweets with *finocchio*, *checca*, or *ricchione*. This means that the remaining 64 per cent of tweets containing the negatively connotated words *finocchio*, *checca*, or *ricchione* were not considered offensive, leaving open the question of what makes this possible. It is also remarkable that a word like *frocio* clusters with the more neutral group. Just as there was a split between extreme and very negative terms in the group of derogatory terms, there is also a split within the neutral group in Node 5: on the one hand, *omosex*, *omosessuale*, and *gay* (only 3 per cent of the tweets containing these words were considered offensive), and on the other hand, *omo* and *frocio* (18 per cent of the tweets containing these words were considered offensive). In summary, we again find a cline in ‘degree of offensiveness’, with at one extreme the tweets with *culattone*, followed by the group including *finocchio*, *checca*, and *ricchione*, then *omo* and *frocio*, and then the least offensive ones: *omosex*, *omosessuale*, and *gay*.

How does this interact with the other linguistic predictors in our tree? As expected, within the group of least offensive words (*omosex*, *omosessuale*, and *gay*), stylistic and indirect strategies play a substantial role and reach statistical significance: it is only when these terms are accompanied by deictics and stereotypes that the percentage of offensive tweets goes up (2 per cent for offensive tweets when no deictic or stereotypes are used, 19 per cent when a stereotype is mentioned, and 8 per cent when a deictic is used). Conversely, for the group of derogatory terms, these strategies do not seem to have an effect: the terms carry enough expressive

power themselves to communicate hate and do not need to be boosted. Compensatory strategies, which would help to decrease the expressive load of a tweet, were not included in the annotation scheme.

5.4 General discussion and conclusions

In this section, we discuss the main results of our study and explore future applications of this methodology in studies on hate speech, and on social media in general.

This preliminary analysis of the dataset allowed us to show that homophobic terminology in social media is heavily dependent on the linguistic context in which it appears. Linguistic and stylistic cues such as ironic and dialectal language can endow harmless terms with an offensive connotation. This was clearly shown by the second set of inference trees ([Figure 5.7](#)). The other analyses showed, among other things, that these same cues were used more often by male X users and, when combined, also lead to: i) an increased use of certain derogatory terms, including *frocio* and *culattone*, and ii) a decrease in the use of the more recent term *gay*. When the tweet was perceived as globally offensive, without contextual cues, more lexical variation in usage occurred. These results confirm that the role of the linguistic context in which terms are used is fundamental to the construction of homophobic language and possibly hate speech in general. This also confirms that natural language processing research should pay significant attention to these microlinguistic aspects in further refining automatic recognition of hate language, but also to other, more stylistic cues, such as ironic or dialectal language use.

On a methodological level, we demonstrate the usefulness of X and social media in general for the large-scale investigation of language variation and change (Grondelaers and Stuart-Smith 2021). We were able to show that X data presents a promising research tool for exploring contextual cues in the construction of hate speech. Not only does it open up a gigantic data source for lexical variation research but it also allows for microlinguistic analyses

of the context in which the relevant words or variables occur. The tweets feature not only morphological and syntactic cues, but also other complex and layered cues, such as ironic and dialectal usage.

Of course, this initial exploration does have some limitations. One is that, at this stage, this fairly large dataset was annotated by only two people. This is too limited to arrive at a broad and valid inter-annotator agreement. For annotations without an obvious measure, such as the degree of offensiveness, this poses a number of problems related to the subjectivity of the annotators. Thus, in future research, more annotators will be needed; they should vary in gender, age, and possibly regional origin. Furthermore, we could cluster among the predictors a number of (perhaps correlating) variables to create larger parameters that could tell us something about the language use in question. Following the example of Grondelaers and Marzo (2023), non-standard spelling and repetition, along with other parameters, could be a manifestation of reinforcement or expressive language use. These parameters could then be used as broader stylistic cues (stylistation) in the regression analyses. Finally, the study should be extended to hate speech towards homosexual women, with a similar dataset of near-synonymous lexicalisations of the concept of *HOMOSEXUAL WOMAN*. This expansion would not only offer further insight into how homophobic language is used on social media but would also provide a new parameter in the research. It would allow us to explore not only lexical differences in the conceptualisation of *HOMOSEXUAL WOMAN* (compared to men) but also variation in the use of conceptual cues in the construction of homophobic language towards men and women.

In other words, much work remains to be done and despite the limitations of the present study, we hope that these initial analytical steps have paved the way for further research into the rich variation in the lexical field of homophobic language use in Italy.

References

- Beers Fägersten, Kristy, and Karyn Stapleton, eds. 2017. *Advances in Swearing Research: New Languages and New Contexts*. Amsterdam: John Benjamins.
- Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.
- Bianchi, Claudia. 2017. 'Linguaggio d'odio, autorità e ingiustizia discorsiva'. *Rivista di estetica* 64: 18–34. <https://doi.org/10.4000/estetica.2059>.
- Bohmann, Axel. 2016. 'Language change because Twitter? Factors motivating innovative uses of because across the English-speaking Twittersphere'. In Squires, *English in Computer-Mediated Communication*, 149–178. <https://doi.org/10.1515/9783110490817>.
- Caselli, Tommaso, Valerio Basile, Jelena Mitrović, Inga Kartoziya, and Michael Granitzer. 2020. 'I feel offended, don't be abusive! Implicit/explicit messages in offensive and abusive language'. Paper presented at *LREC 2020 – 12th International Conference on Language Resources and Evaluation*, Conference Proceedings, 6193–6202. Marseille: European Language Resources Association.
- Coats, Steven. 2016. 'Grammatical feature frequencies of English on Twitter in Finland'. In Squires, *English in Computer-Mediated Communication*, 179–210. <https://doi.org/10.1515/9783110490817-009>.
- Corver, Norbert. 2016. 'Emotion in the build of Dutch: deviation, augmentation and duplication'. *Tijdschrift voor Nederlandse Taal-en Letterkunde* 132 (4): 232–275.
- Da Milano, Federica. 2016. 'Deictic strategies as expression of identity'. In *Beyond Language Boundaries: Multimodal Use in Multilingual Contexts*, edited by Marta Fernández-Villanueva and Konstanze Jungbluth, 153–161. Berlin and Boston, MA: De Gruyter.
- Eisenstein, Jacob. 2013. 'What to do about bad language on the internet'. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, edited by Lucy Vanderwende, Hal Daumé III, and Katrin Kirchhoff, 359–369. Atlanta, GA: Association for Computational Linguistics.
- Faloppa, Federico. 2020. *#ODIO: manuale di resistenza alla violenza delle parole*. Turin. UTET.
- Fónagy, Ivan. 1995 'Iconicity of expressive syntactic transformations'. In *Syntactic Iconicity and Linguistic Freezes*, edited by Marge E. Landsberg, 285–304. Berlin and New York: De Gruyter Mouton.
- Franco, Karlien, Dirk Geeraerts, Dirk Speelman, and Roeland Van Hout. 2019. 'Concept characteristics and variation in lexical diversity in two Dutch dialect areas'. *Cognitive Linguistics* 30 (1): 205–242. <https://doi.org/10.1515/cog-2017-0136>.
- Fuchs, Christian. 2014. *Social Media: A Critical Introduction*. London: Sage.

- Gao, Lei, Alexis Kuppersmith, and Ruihong Huang. 2017. 'Recognizing explicit and implicit hate speech using a weakly supervised two-path bootstrapping.' *Approach Computation and Language* (cs.CL). <https://doi.org/10.48550/arXiv.1710.07394>.
- Gries, Stefan Th. 2020 'On classification trees and random forests in corpus linguistics: some words of caution and suggestions for improvement.' *Corpus Linguistics and Linguistic Theory* 16 (3): 617–647. <https://doi.org/10.1515/cllt-2018-0078>.
- Grondelaers, Stefan, and Stefania Marzo. 2023. 'Why does the *shtyle* spread? Street prestige boosts the diffusion of urban vernacular features.' *Language in Society*: 52(2): 295–320. <https://doi.org/10.1017/S0047404521001202>.
- Grondelaers, Stefan, and Jane Stuart-Smith. 2021. 'Twitter as a laboratory for language variation and change. New opportunities for social-media sociolinguistic research.' Panel organised at *New Ways of Analyzing Variation* (NWAY) 49. Austin, TX, 19–24 October 2021.
- Hart, Christopher. 2014. *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. London: Bloomsbury.
- Hart, Christopher. 2010. *Critical Discourse Analysis and Cognitive Science: New Perspectives on Immigration Discourse*. Basingstoke: Palgrave.
- Jay, Timothy, and Kristin Janschewitz. 2008. 'The pragmatics of swearing.' *Journal of Politeness Research* 4 (2): 267–288. <https://doi.org/10.1515/JPLR.2008.013>.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Müller, Karsten, and Carlo Schwarz. 2021. 'Fanning the flames of hate: social media and hate crime.' *Journal of the European Economic Association* 19 (4): 2131–2167. <https://doi.org/10.1093/jeea/jvaa045>.
- Oliveira, Ruth de. 2013. 'Manifestations émotionnelles de la dislocation: le cas de l'agacement.' In *Cartographie des émotions: propositions linguistiques et sociolinguistiques*, edited by Fabienne Baidier and Georgeta Cislaru, 211–222. Paris: Presses Sorbonne Nouvelle.
- Orlando, Eleonora, and Andrés Saab, eds. 2021. *Slurs and Expressivity: Semantics and Beyond*. Lanham, MD: Lexington.
- Poletto, Fabio, Marco Stranisci, Manuela Sanguinetti, Viviana Patti, and Cristina Bosco. 2017. 'Hate speech annotation: analysis of an Italian Twitter corpus.' In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-It 2017)*, edited by Roberto Basili, Malvina Nissim, and Giorgio Satta, 263–268. Aachen: CEUR-WS. <https://doi.org/10.4000/books.aaccademia.2448>.
- Saab, Andrés. 2022. 'Introducing expressives through equations: implications for the theory of nominal predication in romance.' In *Proceeding of SALT 32*, 356–383. <https://doi.org/10.3765/salt.v1i0.5330>.

- Saab, Andrés. 2021. 'On the locus of expressivity. deriving parallel meaning dimensions from architectural considerations'. In Orlando and Saab, *Slurs and Expressivity*, 17–44.
- Sanguinetti, Manuela, Fabio Poletto, Cristina Bosco, Viviana Patti, and Marco Stranisci. 2018. 'An Italian Twitter corpus of hate speech against immigrants'. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2798–2805. Miyazaki: European Language Resources Association.
- Schmidt, Anna, and Michael Wiegand. 2017. 'A survey on hate speech detection using natural language processing'. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. Valencia: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-1101>.
- Soral, Wiktor, Bilewicz Michał, and Winiewski Mikołaj. 2018. 'Exposure to hate speech increases prejudice through desensitization'. *Aggressive Behavior* 44 (2): 136–146. <https://doi.org/10.1002/ab.21737>.
- Squires, Lauren, ed. 2016. *English in Computer-Mediated Communication: Variation, Representation, and Change*. Berlin and Boston, MA: De Gruyter Mouton.
- Strobl, Carolin, James Malley, and Gerhard Tutz. 2009. 'An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests'. *Psychological Methods* 14 (4): 323–348. <https://doi.org/10.1037/a0016973>.
- Tagliamonte, Sali A. 2006. *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A., and Harald R. Baayen. 2012. 'Models, forests and trees of York English: was/were variation as a case study for statistical practice'. *Language Variation and Change* 24 (2): 135–178. <https://doi.org/10.1017/S0954394512000129>.
- Trotzke, Andreas, and Xavier Villalba. 2021. *Expressive Meaning Across Linguistic Levels and Frameworks*. Oxford: Oxford University Press.
- Wikipedia. n.d. 'Lessico dell'omofobia'. Accessed 23 November 2022. https://it.wikipedia.org/wiki/Lessico_dell%27omofobia.
- Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. New York: Hafner.