

Est.
1841

YORK
ST JOHN
UNIVERSITY

Zingaretti, Mattia, Garraffa, Maria and Sorace, Antonella (2024) The impact of bilingualism on hate speech perception and slur appropriation: An initial study of Italian UK residents. In: EditorsEmailORCIDCruschina, SilvioUNSPECIFIEDUNSPECIFIEDUNSPECIFIEDGianollo, ChiaraUNSPECIFIEDUNSPECIFIEDUNSPECIFIED, (eds.) An Investigation of Hate Speech in Italian: Use, Identification, and Perception. 1 ed. Helsinki, Helsinki University Press, pp. 193-230

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/10032/>

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version:

<https://hup.fi/site/chapters/e/10.33134/HUP-27-6/>

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. [Institutional Repositories Policy Statement](#)

RaY

Research at the University of York St John

For more information please contact RaY at
ray@yorks.ac.uk

CHAPTER 6

The impact of bilingualism on hate speech perception and slur appropriation

An initial study of Italian UK residents

Mattia Zingaretti
York St John University

Maria Garraffa
University of East Anglia

Antonella Sorace
University of Edinburgh

Abstract

The complex relationship between bilingualism and emotions has been extensively studied since the early 2000s, but the potential impact of bilingualism on speakers' perceptions and reactions to an emotionally loaded topic such as hate speech has been overlooked. This chapter reports the first investigation of this kind,

How to cite this book chapter:

Zingaretti, Mattia, Maria Garraffa, and Antonella Sorace. 2024. 'The impact of bilingualism on hate speech perception and slur appropriation: An initial study of Italian UK residents'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 193–230. Helsinki: Helsinki University Press. DOI: <https://doi.org/10.33134/HUP-27-6>.

examining whether hate speech perception differs for late bilinguals in their first language (L1) versus their second language (L2), and how bilingual experience factors such as length of residence in the L2 country and language dominance may predict these perceptions. This research also explores whether the same factors, along with identifying with a sexual or ethnic minority, may predict bilinguals' perception of appropriateness in using slurs to react to hate speech, and whether bilinguals would appropriate slurs themselves. The bilingual group surveyed consists of 43 highly proficient L1 Italian speakers of L2 English, who grew up in Italy until at least the age of 16 and have been in the UK for an average of 5 years. The results indicate that the participants perceive hate speech rather similarly in their L1 and L2. Importantly, despite the overall higher familiarity with L1 hate words, a longer period of residence in the UK is associated with L1 hate words becoming less accessible in terms of familiarity, use, and imageability, while L2 words become less offensive. Moreover, slur appropriation is not predicted by any of the bilingual experience variables, but only by whether participants identify as part of a minority. The findings are discussed with reference to bilingualism research on L1 attrition and emotion, and by highlighting the implications of considering bilinguals' unique perceptions of hate speech from both linguistic and interdisciplinary perspectives.

Keywords: hate speech perception, slur appropriation, bilingualism, second language acquisition, first language attrition

6.1 Introduction

The proliferation of content across various media has allowed hate speech to spread more widely, consequently exacerbating concerns as to how to best define and identify it (see MacAvaney et al. 2019; Kovács, Alonso, and Saini 2021; see also chapters [1](#), [4](#), [5](#), and [7](#) in this volume). Importantly, these concerns are not confined to the online realm, but affect society and individuals more broadly, highlighting the pervasive nature of hate speech and its impact on social interactions and public discourse (see [Chapter 1](#)

for an introduction to these general issues). In addition to this, more than half of the world's population today is bilingual (Grosjean 2010), and the questions of how bilingualism may impact the perception of hate speech, as well as bilingual speakers' reactions to hate speech, have been overlooked.¹ Interestingly, while the literature on how bilinguals express emotions is rich and outlines a complex interplay of factors behind bilingual speakers' language choice when expressing emotions (see, e.g., Dewaele 2010), there is a definite lack of research that explicitly examines the way that speakers of more than one language perceive the emotions conveyed by the use of hate words in their different languages, or how bilingual speakers may react in situations where they encounter, or are the target of, hate speech. The research contained in this chapter thus represents a first attempt at filling this gap.

We begin by highlighting the main research findings on bilingualism and emotions (Altarriba 2003; Ramírez-Esparza et al. 2006; Pavlenko 2006, 2008; Kim and Starks 2008; Wilson 2008, 2013; Dewaele 2010; Gawinkowska, Paradowski, and Bilewicz 2013; Costa et al. 2014), as well as hate speech and emotions (Calvert 1997; Boeckmann and Liew 2002; Gelber and McNamara 2016; Brown 2017a,b; Chiril et al. 2022), and then proceed to review some pragmatic accounts of slur appropriation employed as a way of reacting to hate speech (Hornsby 2003; Hom 2008; Bianchi 2014). After explaining the motivations for our investigation, we present our methodology and report hate speech perceptions and reactions of a group of late bilinguals (namely, 43 Italian people resident in the UK). In particular, we analyse their perception of hate words via ratings of word pairs as well as their reactions to hate speech scenarios with and without slur appropriation, in both their first language (i.e. Italian, henceforth 'L1') and their second language (i.e. English, henceforth 'L2'). After discussing the results gathered through our online study, we high-

1 Throughout this chapter, we use terms such as 'bilingualism' and 'bilingual' to refer to the use of more than one language, synonymously with 'multilingualism' and 'multilingual'.

light the important implications of the current and future research for other disciplines (such as philosophy and ethics, as well as law and politics), acknowledging the limitations of this study and providing suggestions for future research into bilingualism and hate speech.

6.1.1 Research questions and hypotheses

This study analyses the potential impact of bilingualism on hate speech perception and slur appropriation, by seeking to answer the following research questions (RQs):

- RQ1. Do (late) bilingual speakers perceive hate speech similarly in their two languages, and is their perception in the L1/L2 predicted by bilingual experience factors—specifically, by the length of residence (LoR) in the L2 country and their language dominance?
- RQ2. Do the same factors (i.e. L1/L2, LoR, and dominance), as well as identifying as part of a (sexual or ethnic) minority, predict the degree to which bilingual speakers: a) find it appropriate to use slurs to react in situations where someone is the target of hate speech, and b) would appropriate slurs themselves in such situations?

Our main hypotheses (Hs) are the following:

- H1. Firstly, bilingual speakers' perception of slurs may be predicted by increasing length of residence in the L2 country and a switch to L2 dominance, with hate speech being perceived more emotionally in the L2 and less in the L1.
- H2. Secondly, bilinguals may find it more appropriate to use slurs as a response to hate speech in the L2 and less appropriate in the L1, the longer they reside in the L2 country and the more L2 dominant they are. Identifying with a minority may predict the degree of appropriateness perceived, irrespective of other factors. Further, identifying with a minority may also predict the degree

to which bilingual speakers appropriate slurs themselves, and may interact with bilingual experience factors: sexual and ethnic minorities may appropriate slurs more in the L2 and less in the L1, the longer they reside in the L2 country and the more dominant they are in the L2.

The main findings of existing research on the links between bilingualism and emotions as well as the links between hate speech and emotions are reviewed below, laying the groundwork for the present investigation.

6.2 Background

6.2.1 *Bilingualism and emotions*

A large amount of the research into bilingualism and emotions since the turn of the century has highlighted the fact that bilingual speakers feel different when speaking different languages. For instance, Ramírez-Esparza et al. (2006) investigate whether bilinguals show different personalities with regard to the Big Five personality traits in English and in Spanish (i.e. extraversion, agreeableness, openness, conscientiousness, and neuroticism) and whether these differences are consistent with differences between English- and Spanish-speaking cultures. The bilinguals in their study are indeed found to be more extroverted, agreeable, and conscientious (but not more open or more neurotic) in English than in Spanish, and these differences are consistent with the personality displayed in each culture.

Similarly, Pavlenko (2006) explores whether multilingual speakers feel different when changing languages. After analysing 1039 responses to an open question about feeling different in a foreign language in her Bilingualism and Emotion Questionnaire (BEQ), Pavlenko finds that 65 per cent of her participants report feeling different when using another language compared to only a quarter of participants who reported not feeling different, with 10 per cent giving ambiguous responses. Pavlenko further observes that there are four main sources of perception of a

different self in a foreign language for her participants: linguistic and cultural differences, distinct learning contexts, different levels of language emotionality, and different levels of language proficiency. What is interesting to note here is that feeling different in a foreign language is not restricted to late or immigrant bilinguals, but, as Pavlenko (2006: 27) states, it ‘is a more general part of bi- and multilingual experience.’ Her participants also describe their bilingual experiences as enjoyable and unique and refer to the integration of their different identities.

Following Pavlenko, Wilson (2008) further investigates the issue of feeling different in a foreign language, and categorises and quantifies the responses of 1414 participants in the BEQ. The feelings reported by Wilson’s participants when interacting in their foreign language are overall positive, with most participants feeling as if it were somebody else speaking, hence reporting feeling more confident and more outgoing in their foreign language, as well as highlighting changes in body language, mannerisms, and voice, and deeper levels of disguise. Wilson (2013) later correlates the BEQ scores of 108 adult participants with scores on the Big Five personality traits: while gender and age have no effect, higher BEQ scores of feeling different in a foreign language are reported by introverts who rate their L2 proficiency at intermediate level or above, participants with lower educational levels, participants with higher levels of perceived L2 proficiency and who started learning the L2 at a younger age, as well as mixed and naturalistic learners as opposed to instructed learners.

The evidence reviewed so far shows that feeling different in an L2 is incredibly common among most bilinguals, and that the degree to which bilinguals feel different in their L2 can be affected by factors such as L2 learning contexts and age, as well as proficiency and educational levels. More recently, research into the factors influencing the use of one of the languages (LX) of a multilingual speaker to express emotions finds similar results. In particular, Dewaele (2010) notes that late learners tend to use the LX less frequently to communicate emotions, rating positive characteristics of the LX lower and reporting higher levels of anxiety

when communicating in the LX. Moreover, similarly to Pavlenko (2006) and Wilson (2008, 2013), Dewaele finds that languages learnt only through formal instruction are less frequently used to communicate emotions, and that naturalistic and mixed learners feel more competent and less anxious than instructed learners. Something not noted in previous research but reported in Dewaele is that frequency of use of the LX, LX socialisation, and networks of interlocutors, along with the number of languages spoken, are also all factors that significantly influence the use of an LX to express emotion: specifically, more frequent use of an LX, higher levels of socialisation in the LX, and larger networks of interlocutors in the language, as well as a higher number of LXs spoken, all correspond to more frequent use of an LX to express emotions. Lastly, as also reported in Pavlenko (2006), Dewaele observes that communicating emotions covers a range of speech acts that are often culture specific: for example, raising one's voice when angry may be acceptable in some parts of the Western world, such as Southern Europe, but it is considered taboo in Asia. Importantly, multilingual speakers are able to exploit their multicompetence to develop speech acts and emotion scripts that are entirely unique to them or their peers, thus revealing an incredibly dynamic aspect of language choice to express emotions, and showing a growing awareness of sociocultural and sociopragmatic LX norms is accompanied by an evolution of the LX user's repertoire to express emotions in the LX. While cultural and social norms need to be learnt in the L2, these kinds of norms are already acquired in the L1—and they are not the only aspects that a late bilingual speaker is already acquainted with in their L1.

Indeed, some research has demonstrated the existence of a greater emotional overtone of the L1 connected with first emotion experience (Altarriba 2003): in other words, because certain emotions may be experienced first in a bilingual's L1, the expression of those emotions may come more easily for them in their L1, as that context presumably carries more connotations and associations, at least for late bilinguals. The fact that words in the L1 may seem more natural for late bilinguals, while words in the L2 may

be considered ‘disembodied’, as some further research suggests (Pavlenko 2008), leads Gawinkowska, Paradowski, and Bilewicz (2013) to point out that two opposite conclusions could be drawn: either it may be easier for bilinguals to talk about emotional topics in their ‘more natural’ L1, or they may prefer to do so in their L2 when social and cultural norms of their L1 may be ‘too burdensome.’ In their research, Gawinkowska, Paradowski, and Bilewicz report evidence that supports the latter conclusion. Specifically, they analyse the offensiveness ratings given by 61 Polish–English bilingual students to two texts they are asked to translate in both of their languages (from the L1 to their L2 and from the L2 to their L1). Gawinkowska, Paradowski, and Bilewicz find an effect only when analysing the results of offensiveness in target ethnophaulisms (i.e. words targeting social groups, which, as such, are subject to greater norms of political correctness) as opposed to general swear words (which are subject to lower norms of political correctness). Specifically, ethnophaulisms in the L2 translations are found to be significantly more offensive than source words in the L1, and vice versa: ethnophaulisms in the L1 translations are significantly less offensive than source words in the L2. They thus conclude that the main factor triggering ‘emotion-related language choice’ (ERLC; Kim and Starks 2008), with students feeling freer to swear in their L2, are social and cultural norms. According to Gawinkowska, Paradowski, and Bilewicz (2003: 5), ‘the foreign language exempts us from our own socially imposed norms and limitations and makes us more prone to swearing and offending others.’ Although the authors note that these findings may be extended to all people who know a foreign language to a ‘communicative level’, the extent to which this is true remains to be seen empirically.

This is also true of the conclusions drawn by other research that highlights an increase in psychological distance inducing utilitarianism when bilingual speakers are asked to make moral judgements in their L2—for instance, when answering a question such as: ‘Would you sacrifice a man to save five?’ (Costa et al. 2014). Further research is required to empirically verify the extent

to which these conclusions can be applied, because bilingualism is a complex and dynamic process that sees both languages always being active and interacting in the bilingual mind (for an overview, see Bialystok, Craik, and Luk 2012). The interaction of the two languages results in significant linguistic changes for bilingual speakers which set them apart from monolingual speakers of either language—or, as Grosjean (1989) famously observed, bilingual speakers are not the sum of two monolinguals in one. Crucial to the present discussion are the changes experienced by late bilingual speakers in the understanding and use of their L1, usually referred to as L1 ‘attrition’ (for an overview of works in the field from different perspectives, see, among others, Schmid 2016; Sorace 2020; Gallo et al. 2021; Zingaretti 2022; Zingaretti et al. forthcoming). In particular, despite the general consensus on the importance of emotions in L1 attrition, the links between emotions and L1 attrition remain largely understudied, and neither Gawinkowska, Paradowski, and Bilewicz (2003) nor Costa et al. (2014) take L1 attrition into account in their investigations.

One of the few studies attempting to bridge this gap in the field is research by Kim and Starks (2008), which investigates emotions in L1 attrition and L2 acquisition in a group of 30 Korean-English L1-dominant late bilinguals in New Zealand. The results of a story-retelling task, a questionnaire, and a follow-up interview with their participants show a shift from the L1 to the L2 related to an increase in L2 fluency and a decrease in L1 accuracy. Indeed, despite the overall preference for ERLC to be in the L1 rather than the L2, there is a considerable amount of L2 use, particularly when anger-related emotion is involved, as well as correlations between most measures of ERLC with decreasing L1 accuracy and increasing L2 fluency. Ultimately, Kim and Starks note that these results call for an urgent need for dual language support for young Korean immigrants in New Zealand: for these immigrants, the increase in socialisation in the L2 may come with delayed or ‘primitive’ socialisation in the L1, which may ultimately result in the attrition (here intended as ‘loss’, due to disuse) of the language in later stages of their lives.

Overall, while there seem to be few differences in terms of the emotional resonance of L1 and L2 words for early and simultaneous bilinguals (see Harris 2004), the same cannot be said for late bilinguals, whose relationships with emotions in the different languages they speak are shaped by a variety of different factors, and ultimately give rise to highly individual preferences when expressing emotions in either language, as seen thus far. To better understand why bilingualism may also influence speakers' perception of hate speech, we now review some of the literature on the relationship between emotions and hate speech.

6.2.2 *Hate speech and emotions*

The relationship between hate speech and emotions is similarly complex, starting from the very emotion and feeling that is evoked by the terminology itself (i.e. hate) or, as some scholars somewhat drastically put it, 'the myth of hate' (Brown 2017a, 2017b). Brown (2017a) observes that hate speech, in its ordinary (rather than legal) meaning, does not correspond to a single monolithic phenomenon, but to a diverse set of expressive phenomena that may not necessarily involve hate in its most distinctive quality of intense or extreme dislike. Indeed, according to Brown, there are many occasions on which the terms 'hate speech' or 'hate crime' may be used where no hate or hatred is involved. One of the examples proposed is that of a fundamentalist or evangelical Christian directing the following words at people entering a gay club on the street: 'You homosexuals and lesbians are sinners in the eyes of God, you disobey the teachings of the Bible, and for this reason you will go to hell if you do not repent' (Brown 2017a: 450). Rather than expressing emotions, feelings or attitudes of hate or hatred, these words serve as an expression of religious belief or feelings of disgust or repulsion learnt from parents or community leaders; however, expressions like these, as well as others, can still be classified as hate speech on the basis that they are forms of speech that carry prejudiced messages (here, as homosexuals are portrayed as

morally inferior beings) or hate speech acts that rank, degrade, harass, or persecute someone (homosexuals in this case).

Brown (2017b: 562) goes further, saying that hate speech should not be interpreted as a compositional concept made up of the literal meanings of the words ‘hate’ and ‘speech’, but as an equivocal idiom, or rather ‘a family of meanings, for which there is no one overarching precise definition available.’ Although the aim of this chapter is not to discuss the multifarious ways in which hate speech can manifest, it seems important to emphasise that the phenomenon—or rather ‘phenomena’, following Brown’s logic—under investigation here entails a complexity of emotions and feelings that go beyond mere hate. Acknowledging that hate speech entails complex emotions and feelings is ultimately crucial in identifying it successfully, as more recent research in natural language processing shows that the emotions encoded in sentic computing sources and semantically structured hate lexicons help to detect forms of online hate speech more accurately (Chiril et al. 2022).

Importantly, emotions are not only part of the source—they are also generated in the listener as a result of being targeted by hate speech. In this respect, borrowing from Carey’s (1989) transmission model of communication, Calvert (1997) discusses the emotional and physical harms of hate speech:

The question of harm caused by hate speech, when considered from the perspective of the transmission model, boils down to this: Did communication of a particular message, X, cause a change, Y, in the attitude or behavior of the recipient of the message? Does a bigot’s calling the African-American standing next to him in line at the movie theater a ‘nigger’ cause the African-American’s *pulse rate to increase* or his stomach to *feel nauseated*? Does it cause him *mental pain and anguish* or make him feel *angry*? Does it cause him to strike the speaker? Each affirmative response is a direct physical or emotional change caused by a particular message. (Calvert 1997: 10; emphasis added).

The important point highlighted by Calvert (1997: 10) is that law courts deem emotional harms to be intangible unless accompanied

by proof of physical harms, and that the physical symptoms experienced by victims of hate speech (e.g. increased pulse rate or breathing difficulties) do not 'lend themselves to proof of harm at trial', while, in fact, these harms are very much real.

The ways in which hate speech impacts targets emotionally are significant and widely documented. For instance, Gelber and McNamara (2016) investigate the harms of hate speech as evidenced by the experiences of 101 members of indigenous and minority ethnic communities in Australia. Their research demonstrates multiple types of harms reported by the participants, among which are hurt and upset feelings that are sometimes deep enough to be perceived as 'existential' pain. Hate speech also has the power to silence targets, rendering them unable to respond directly, and at times causing them to withdraw from the situation altogether. Feelings of exclusion, dehumanisation, anger, and frustration, alongside the inability to identify with one's own ethnicity or religion, are also reported as consequences of being targeted by hate speech, and, at times, also as the consequence of the perpetuation of negative stereotypes by the media (Gelber and McNamara 2016: 334–335). Similar results are also reported in research analysing Asian American students' responses after reading second-hand accounts of hate speech: not only do the students suffer what Boeckmann and Liew (2002: 377) define as a '(presumably) temporary reduction in collective self-esteem', but, according to the researchers, being the direct target of similar speech in real-life scenarios would surely result in 'more extreme and enduring consequences'. Both Gelber and McNamara (2016) and Boeckmann and Liew (2002) agree that the significant emotional and psychological effects of hate speech constitute evidence in favour of sanctions against it due to its extremely harmful impact.

Clearly, then, the relationship between hate speech and emotions is complex, and the impact that hate speech has on its targets can result in significant emotional and psychological harms: feelings of exclusion, dehumanisation, anger, frustration, reduced identification with one's own ethnicity or identity, and reduced self-esteem are all reported in the literature, alongside silence and withdrawal from

the situation as common reactions among victims of hate speech. We will now explore some accounts that seek to explain another possible reaction to hate speech—namely, slur appropriation.

6.2.3 *Slur appropriation as a pragmatic response to hate speech*

From a linguistic perspective, specific pragmatic strategies related to hate speech reactions have been investigated. One of these strategies is ‘slur appropriation’—that is, the use of a slur usually targeting a specific group by that very same targeted group, for non-derogatory purposes within the group (Hom 2008). According to Potts (2007: 10): ‘When lesbian and gay activists use the word “queer”, its meaning (its expressive content) differs dramatically from when it is used on conservative talk radio’. According to Hom (2008) there are multiple reasons why in-group members may wish to appropriate a slur. Not only does slur appropriation allow in-group members to take back the instrument of discrimination from the discriminators, but in doing so the appropriators also soothe and neutralise the originally offensive effect of the slur. Importantly, slur appropriation allows members to demarcate their (in-)group and show solidarity among other members, while highlighting that they are still objects of discrimination. Hornsby (2003) further explains that in-groups critically position themselves against normal (i.e. derogatory) uses of the slur, and adds to Hom’s ‘soothing’ and ‘neutralising’ objectives that of subverting the old, non-descriptive meanings of the slur being appropriated.

Appropriation, reversal, and subversion are conceived by Bianchi (2014) as ‘echoic uses’. To elaborate further, in echoic uses speakers not only report the utterances or thoughts of others, but also convey their own attitude in regard to those utterances or thoughts: by appropriating slurs, then, in-groups echo derogatory uses in ways that make explicit the dissociation from the offensive content originally conveyed by the slur. The effect, Bianchi suggests, is ironic: slur appropriators mock those who make use of the slurs in a derogatory way, by only mentioning part or some of

the constituents of the attributed utterance or thought, with a tacit echo and dissociative attitude that has to be understood from the context, facial expression, tone of voice, or other paralinguistic cues. Lastly, Bianchi notes that over time, appropriated uses may extend from in-group use only to out-group use as well, which is what has happened to words such as ‘queer’ or ‘gay’ today, as they have lost their hints of echo or irony, and are openly used without connotation by people who do not identify as such.

The relevance of slur appropriation to the present study becomes clear when we consider that most of our knowledge about how slur appropriation works, and about the links between hate speech and emotions more generally, comes from monolingual research. Crucially, since bilingual speakers are part of two linguistic and cultural communities, they may be better able to manipulate the polarity of a term, particularly due to their reportedly enhanced ability to understand behaviours that differ from their own (see the work on Theory of Mind in adult bilinguals by Navarro and Conway 2021). Moreover, given the heavy emotional load associated with L1 slurs due to first emotion experience (see [Section 6.2.1](#) on bilingualism and emotions), it may be easier for bilinguals to appropriate slurs in the L2 than in the L1. The extent to which this is true requires empirical observation, which the research contained within the present chapter carries out.

6.2.4 Focus of the previous literature

In brief, the research presented so far demonstrates clear links and complex relationships between i) bilingualism and emotions, as well as ii) hate speech and emotions. On the one hand, bilinguals commonly report feeling different in the different languages they speak for a variety of reasons (e.g. L2 learning contexts and age, L2 proficiency and educational levels) and they often choose to express their emotions in a language due to a range of factors (e.g. medium of instruction in the language, socialisation, number of interlocutors, and cultural specificity). L1 attrition also contributes to shaping bilingual speakers’ language choice when

expressing emotions, although the extent of its role requires further scrutiny. On the other hand, the impact of hate speech on the individual can be extreme in terms of emotions, with feelings of exclusion, dehumanisation, anger, frustration, reduced identification with one's own ethnicity or identity, and reduced self-esteem all being reported in the literature, together with silence and withdrawal from hate speech situations.

However, the evidence we have regarding the emotional and psychological impact of hate speech, as well as what we know regarding the possible reactions to hate speech through slur appropriation, has only been gathered within monolingual contexts. Indeed, no study to date has sought to investigate the ways in which bilingual speakers perceive an emotionally loaded topic such as hate speech in their different languages, or the possibly different reactions that bilingual speakers may have when they are the target of hate speech in either of their languages. This research thus represents an initial attempt at addressing this gap.

6.3 Methodology of the present study

6.3.1 *Materials and design*

This study gathered data through an online questionnaire administered on the Qualtrics XM platform (<https://www.qualtrics.com>) after obtaining ethics approval from the School of Philosophy, Psychology and Language Sciences Research Ethics Committee at the University of Edinburgh (ref. 377-2021/7). The online questionnaire employed in the study required participants to obligatorily answer three main blocks of questions and it gathered participants' background information through optional final questions. The first part of the questionnaire collected qualitative data regarding participants' experiences with hate speech during their adolescence and in the preceding five years. The second and third parts constitute the experimental block of the study; they gathered participants' quantitative responses on ratings of word pairs and reactions to hate speech scenarios. The materials used and data

gathered in all three parts of the study can be found on the project page on the Open Science Framework website (<https://osf.io/fbtmy/>). In this chapter, we focus on the design, results, and discussion of the data gathered in the experimental block of the study.

The word pair ratings included the ten word pairs shown in [Table 6.1](#), which were presented to our participants first in Italian and then in English in randomised order. Following what has already been carried out by other studies on monolingual speakers (see, for English, Janschewitz 2008; for Italian, Sulpizio et al. 2020), we asked participants to evaluate each of the words on a Likert scale from 1 to 9, for the following seven criteria in each language: *familiarity* (i.e. how familiar they were with the term presented, where 1 = never having encountered the word, and 9 = encountering it all the time); *personal use* (i.e. how often they used the term, where 1 = never having used the word, and 9 = using it all the time); *imageability* (i.e. how easily the term evoked an image for the participants, where 1 = no image being evoked at all, and 9 = image being evoked easily); *arousal* (i.e. how much the term stimulated their attention, where 1 = not finding the term very stimulating, and 9 = finding it very stimulating); *offensiveness* (i.e. how offensive the term was perceived to be, where 1 = not finding the word offensive, and 9 = finding it extremely offensive); *tabooness* (i.e. how taboo the term was perceived to be, where 1 = not recognising the word as taboo, and 9 = recognising it as a major taboo); and finally *valence* (i.e. how positive or negative they found the term, where 1 = finding the word very unpleasant, and 9 = finding it very pleasant).

With regard to the reactions elicited by reading scenarios containing hate speech, inspired by similar work in the field with monolingual speakers (see Boeckmann and Liew 2002), we designed four different scenarios in each language, for a total of eight scenarios presented to our participants. Further, in each language, two of the scenarios contained the metalinguistic use of the same slur used to target the victim by the very same victim in response to hate speech (in a similar though not identical way to how slur appropriation works, as described in [Section 6.2.3](#)).

Table 6.1: Italian–English word pairs rated by the participants of the study.

Italian	English
<i>Testa di cazzo</i>	Dickhead
<i>Stronzo/a</i>	Asshole
<i>Spastico/a</i>	Spastic
<i>Ritardato/a</i>	Retard
<i>Troia</i>	Whore
<i>Puttana</i>	Slut
<i>Cagna</i>	Bitch
<i>Negro/a</i>	Nigger
<i>Frocio</i>	Faggot
<i>Lella</i>	Dyke

Note: The last word pair (*lella*, dyke) was removed from the analysis because many respondents reported never having heard the Italian word (probably due it being mainly used within the Rome area).

This is illustrated, for instance, in (1), where the female victim of hate speech responds to the slur uttered by the person sitting next to her reclaiming the same slur. This affects the person who insulted her, who leaves, embarrassed:

- (1) An African American lady is talking on the phone with her son on the tram. When the tram stops, an old woman gets on board and sits down next to her while she keeps on speaking to her son on the phone. As the old woman starts to get irritated, she looks at the lady and says: ‘You niggers just know how to shout, don’t you?’ As soon as the lady hears the woman’s comment she looks away from her phone, staring straight into the woman’s eyes, and shouts: ‘*Of course us niggers know how to shout. We’ve got to make sure people hear us when there’s some racist around us!*’ The old woman then stands up and leaves, feeling completely embarrassed. (Emphasis added)

The remaining two scenarios did not include slur appropriation—as in (2), where a young woman subject to hate speech keeps quiet as a result:

- (2) A young woman is heading to her best friend's wedding on the train. She is wearing a low-cut dress which shows her cleavage. When she gets up to go to the bathroom, she sees two women staring at her. One of them says to the other: 'Have you seen her?! With that dress, she looks like such a slut.' The young woman notices the two. *She pretends not to have overheard anything and proceeds to walk towards the bathroom, feeling humiliated.* (Emphasis added)

After reading each scenario, participants were prompted to answer the following question: 'Do you find it appropriate to use the word X to respond to an insult (as in the example above),² and would you ever use a slur which refers to sexuality or ethnicity to respond to an insult yourself?' As we explain in [Section 6.4](#), we analysed the responses that deemed (in)appropriate the use of an insult in response to an insult (i.e. the dependent variable we named 'appropriateness') separately from the responses relating to the reappropriation of the slur by the participants themselves (i.e. the dependent variable named 'appropriation').

6.3.2 Participants

For the purposes of this study, we aimed to recruit a group of late bilingual speakers in the UK—specifically, Italian speakers of English with high intermediate/advanced L2 proficiency, who had to have grown up in Italy and emigrated to the UK after 16 years of age. The participants were recruited online on social media platforms (e.g. on Facebook pages for Italian expatriates in the UK, Twitter, and Instagram) and by word of mouth. After discarding

2 'As in the example above' only appeared in questions relating to scenarios with slur appropriation.

Table 6.2: Participants' background information variables with either Mean (M) and Standard Deviation (SD) or percentage (%) for each variable.

Background variable	M and SD or %
Age <i>Years</i>	M = 31.2, SD = 7.94
Length of residence <i>Years</i>	M = 5.05, SD = 2.86
Language dominance <i>Italian</i> <i>English</i>	83.72% 16.28%
Gender <i>Woman</i> <i>Man</i> <i>Non-binary</i>	67.44% 27.91% 4.65%
Sexual orientation <i>Heterosexual</i> <i>Homosexual</i> <i>Bisexual</i> <i>Pansexual</i> <i>Other</i>	69.77% 11.63% 11.63% 2.33% 4.65%
Ethnicity <i>White</i> <i>Hispanic/Latinx</i> <i>Asian</i> <i>Middle Eastern</i>	90.70% 4.65% 2.33% 2.33%
Disability <i>No</i> <i>Yes</i> <i>Rather not say</i>	83.72% 13.95% 2.33%
Educational level <i>University</i> <i>Secondary school</i>	79.06% 20.93%

6 participants who did not fit our initial requirements, our final sample consisted of 43 bilingual speakers, whose background information is summarised in [Table 6.2](#).

Participants were instructed to complete the questionnaire in their own time as no time limitations were set, although they were not allowed to go back and change answers that they had

previously given. Upon completion, they were entered into a prize draw with a chance to win one of three different monetary prizes.

6.4 Results

Below we report our results for the word pair ratings and reactions to hate speech scenarios. All of our analyses were carried out using jamovi (The jamovi project 2021), an open-source statistical software based on the R programming language (R Core Team 2021). For optimal understanding, our results are presented in two complementary formats. First, we provide an accessible explanation, facilitating comprehension for a broad audience irrespective of their statistical proficiency. The initial, non-technical explanation is then followed by detailed statistical results, catering to those keen on methodological specifics. Corresponding figures and tables are provided to visually aid the interpretation of our results.

6.4.1 *Word pair ratings*

To begin, our investigation into the various word characteristics rated by our participants revealed that participants were significantly more familiar with the Italian words than they were with the corresponding words in English. In terms of other characteristics (personal use, imageability, arousal, offensiveness, tabooeness, and valence), there were no notable differences in ratings between languages.

Precisely, paired samples Student's *t*-tests (or Wilcoxon signed-ranks tests where data failed to meet normality assumptions) were run to compare ratings for the same pairs of criteria between languages (e.g. familiarity in Italian and familiarity in English, personal use in Italian and personal use in English). The results are reported in [Table 6.3](#) for familiarity (Fam.), personal use (P.use), imageability (Ima.), arousal (Aro.), offensiveness (Off.), tabooeness (Tab.) and valence (Val.) in Italian and English with Mean (M), Median (Mdn), Standard Deviation (SD), degrees of freedom (df) and *p* values.

Table 6.3: Student's t-tests and Wilcoxon signed-ranks tests for word pair ratings.

	Italian			English			Test type	Stat	df	<i>p</i>
	M	Mdn	SD	M	Mdn	SD				
Fam.	5.58	—	1.73	4.67	—	1.52	Stud.'s <i>t</i>	3.89	42	<.001
P. use	2.49	—	0.85	2.28	—	0.66	Stud.'s <i>t</i>	1.54	35	0.132
Ima.	4.82	—	1.99	4.68	—	2.12	Stud.'s <i>t</i>	0.49	42	0.621
Aro.	—	5.11	—	—	4.56	—	Wilcox. W	473	—	0.249
Off.	—	6.44	—	—	6.22	—	Wilcox. W	580	—	0.198
Tab.	—	4.67	—	—	4.89	—	Wilcox. W	339	—	0.648
Val.	—	2.78	—	—	3.11	—	Wilcox. W	395	—	0.950

As can be seen, the only between-language rating comparison that showed a statistically significant difference was familiarity: that is, participants were significantly more familiar with words in Italian ($M = 5.58$, $SD = 1.73$) than in English ($M = 4.67$, $SD = 1.52$), $t(42) = 3.89$, $p < .001$. The remaining comparisons—personal use, imageability, arousal, offensiveness, tabooeness, and valence—were not significant between languages ($p > .05$).

In addition to comparing the ratings in both languages, we also examined how each criterion related to the other between languages. In other words, we checked whether individuals who rate a word highly for a particular characteristic in one language, such as familiarity, also rate it highly for the same characteristic in the other language. We found that this is generally the case. For instance, if a participant was familiar with a word in Italian, they also tended to be familiar with the corresponding word in English. The same was true for all the other characteristics examined: personal use, imageability, arousal, offensiveness, tabooeness, and valence.

Elaborating further, Pearson correlation analyses (or Spearman rank-order correlations where data failed to meet normality

assumptions) were run to explore the relationship of each criterion between languages (e.g. the relationship between familiarity in Italian and familiarity in English, personal use in Italian and personal use in English), as well as across criteria within languages (e.g. familiarity in Italian and personal use in Italian, familiarity in English and personal use in English). The results are reported in [Table 6.4](#). As can be seen, we found positive moderate-to-strong correlations for all criteria between languages, indicating that high ratings for each criterion in one language corresponded to high ratings for the same criterion in the other language (e.g. high ratings for familiarity in Italian corresponded to high ratings for familiarity in English, high ratings for personal use in Italian corresponded to high ratings for personal use in English, and so on for each pair of criteria).

Moreover, we examined how different criteria related to each other within a single language, either Italian or English, discovering interesting patterns. For instance, in both languages, the words people were more familiar with were also the ones they used more often, and the words deemed more taboo were also viewed as more offensive. Additionally, again in both languages, words that grabbed more attention (arousal) were seen as more offensive and were also more likely to bring a clear mental picture (imageability). Specific to English, participants tended to use offensive words less frequently, and words that grabbed attention (arousal) were more often labelled as taboo. On the other hand, specific to Italian, words that easily brought a mental picture (imageability) were seen as more familiar and more offensive.

To unpack the specifics, we found correlations of different strengths across criteria within a language. Among the within-language correlations reported in [Table 6.4](#), we found: positive correlations between familiarity and personal use in both Italian, $r_s(41) = .435$, $p = .004$, and English, $r_s(41) = .415$, $p = .006$, meaning that in both languages words rated as more familiar were also rated as being more used; positive correlations between taboo-ness and offensiveness in both Italian, $r_s(41) = .431$, $p = .004$, and English, $r(41) = .507$, $p < .001$, meaning that words rated as being

Table 6.4: Correlation matrix for all variables in Italian (I) and English (E) with Pearson's r (r) and Spearman's rho (ρ) values.

	Fam. (I)	Fam. (E)	P. use (I)	P. use (E)	Ima. (I)	Ima. (E)	Aro. (I)	Aro. (E)	Off. (I)	Off. (E)	Tab. (I)	Tab. (E)	Val. (I)	Val. (E)
Familiarity (I)	r	—												
	ρ	—												
Familiarity (E)	r	.553***												
	ρ	—												
Personal use (I)	r	.435***	.192	—										
	ρ	—	—	—										
Personal use (E)	r	.383*	.415*	.465**	—									
	ρ	—	—	—	—									
Imageability (I)	r	.512***	.308*	—	.149	.041	—							
	ρ	—	—	—	—	—	—							
Imageability (E)	r	.208	.189	—	.644***	—	—							
	ρ	—	—	.128	-.021	—	—							
Arousal (I)	r	.352*	.134	—	.495***	.479**	—							
	ρ	—	—	.039	-.153	—	—							
Arousal (E)	r	.109	-.049	—	.467**	.611***	.641***	—						
	ρ	—	—	.064	-.238	—	—	—						
Offensiveness (I)	r	.019	-.072	—	.346*	.27	.363*	.336*	—					
	ρ	—	—	-.053	-.107	-.346*	.32*	.384*	—					
Offensiveness (E)	r	.028	-.142	—	.306*	.251	.32*	.384*	—					
	ρ	—	—	-.028	-.34*	-.028	-.028	.639***	—					
Tabooness (I)	r	-.001	-.182	—	.047	.021	.192	.036	—	.466**	—			
	ρ	—	—	-.06	-.216	—	—	—	.431**	—	—			
Tabooness (E)	r	.072	-.119	—	.034	.129	.358*	.361*	—	.507***	.584***	—		
	ρ	—	—	.55	-.221	—	—	—	.319*	—	—	—		
Valence (I)	r	.068	-.106	—	.121	-.008	.133	.161	—	-.046	-.094	-.026	—	
	ρ	—	—	.251	.009	-.042	.048	.06	-.133	-.183	-.087	.085	-.626***	—
Valence (E)	r	.084	-.000	—	.207	-.286	.042	.06	-.133	-.183	-.087	.085	-.626***	—
	ρ	—	—	.207	.017	-.286	.042	.06	-.133	-.183	-.087	.085	-.626***	—

Note: * p <.05, ** p <.01, *** p <.001

more taboo were also rated as being more offensive in both languages; positive correlations between arousal and offensiveness in both Italian, $r_s(41) = .363$, $p = .017$, and English, $r(41) = .384$, $p = .011$, and also between arousal and imageability in both Italian, $r(41) = .495$, $p < .001$, and English, $r(41) = .611$, $p < .001$, indicating that words that stimulated our participants' attention more were also rated as being more offensive and more easily evoked in their minds in both languages; in English only, a negative correlation between offensiveness and use in English, $r_s(41) = -.340$, $p = .026$, indicating that high ratings for offensiveness in our participants' L2 corresponded to low ratings for personal use of the same words in the same language, but not in their L1; also only in English, a positive correlation between arousal and tabooeness, $r(41) = .361$, $p = .018$, meaning that more stimulating words were also rated as being more taboo in the L2; finally, in Italian only, positive correlations between imageability and familiarity, $r(41) = .512$, $p < .001$, and imageability and offensiveness, $r_s(41) = .346$, $p = .023$, indicating that words that more easily evoked images in our participants' minds were also rated as being more familiar to them and more offensive in their L1.

Lastly, we wanted to understand if LoR in the UK and language dominance could predict the ratings for each criterion in each language. To summarise, we found that the length of time participants had lived in the UK—not their language dominance—predicted their familiarity, their personal use, and how vividly they could picture the meanings of Italian words. Specifically, for Italian, a longer LoR predicted lower familiarity, lower personal use, and lower imageability. On the other hand, for English, neither LoR nor language dominance could predict familiarity, personal use, or imageability. However, the LoR in the UK did predict how offensive participants found English words, with longer LoR predicting lower perceived offensiveness. None of the factors we tested could significantly predict arousal, tabooeness, and valence in either language.

Going into further detail, multiple linear regressions were run to investigate whether each of the seven criteria rated by our

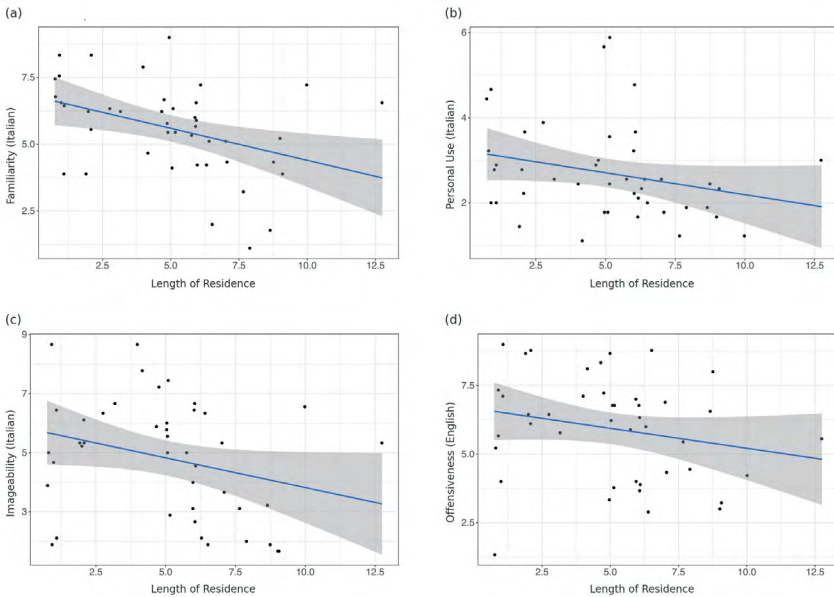


Figure 6.1: Scatter plots with regression lines and shaded standard errors, showing how length of residence predicts familiarity in Italian (a), personal use in Italian (b), imageability in Italian (c) and offensiveness in English (d).

participants in each language (e.g. familiarity in Italian, familiarity in English, personal use in Italian, personal use in English) could be predicted by LoR, language dominance (i.e. Italian/English), and their interaction. The effects of significant predictors are shown in [Figure 6.1](#).

First, familiarity in Italian was significantly predicted by our model, $F(2, 40) = 3.84, p = .030, R^2 = .161$. While LoR ($\beta = -.379, p = .015$) added significantly to the prediction—see [Figure 6.1\(a\)](#)—dominance ($\beta = .183, p = .650$) was not a significant predictor in the model. In English, instead, familiarity was not significantly predicted by our model and none of our predictors added significantly to the prediction ($p > .05$). Second, personal use in Italian was not significantly predicted by our model, $F(2, 33) = 3.06, p = .060, R^2 = .157$; however, LoR ($\beta = -.392, p = .021$), unlike

dominance ($\beta = -.348, p = .455$), did add significantly to the prediction—see [Figure 6.1\(b\)](#). On the other hand, personal use in English was not significantly predicted by our model and none of our predictors added significantly to the prediction ($p > .05$). Third, imageability in Italian was significantly predicted by our model, $F(3, 30) = 2.90, p = .047, R^2 = .182$. LoR ($\beta = -.392, p = .021$) added significantly to the prediction—see [Figure 6.1\(c\)](#); however, neither dominance ($\beta = -.180, p = .142$) nor LoR * dominance ($\beta = 1.20, p = .071$) were significant predictors in the model. While imageability in English was also significantly predicted by our model, $F(3, 39) = 3.00, p = .042, R^2 = .188$, none of the variables added significantly to the prediction ($p > .05$). Fourth, offensiveness in Italian was not significantly predicted by our model and none of the variables added significantly to the prediction ($p > .05$). On the other hand, although offensiveness in English was not significantly predicted by our model either, $F(3, 39) = 1.92, p = .143, R^2 = .128$, LoR ($\beta = -1.41, p = .036$) did add significantly to the prediction, while dominance ($\beta = -.619, p = .080$) and LoR * dominance ($\beta = 1.26, p = .066$) were not significant predictors in the model. None of our remaining models for arousal, tabooeness, and valence were significant in Italian or English, with no variable adding significantly to the predictions ($p > .05$).

6.4.2 *Reactions to hate speech scenarios*

Looking at reactions to hate speech scenarios, we wished to investigate whether different factors could predict whether bilinguals would consider the appropriation of slurs as an appropriate or inappropriate choice. These factors included the language used, LoR, language dominance, ethnicity, sexuality, and a combination of these factors. Overall, our model showed that only sexuality played a significant role, with individuals identifying as a non-heterosexual being more likely to deem the appropriation of slurs as acceptable.

To delve deeper into the analysis, we ran a first binomial logistic regression to predict the likelihood that bilinguals would perceive

slur appropriation as an appropriate rather than inappropriate choice (i.e. ‘appropriateness’, yes/no) by language (i.e. L1/L2), LoR, dominance (i.e. Italian/English), ethnicity (i.e. white/non-white), sexuality (i.e. heterosexual/non-heterosexual), and the interaction of language with any of these variables.³ The overall model was significant, $\chi^2(9) = 17.65$, $p = .039$, with between 9.7 per cent and 13.8 per cent of the variance in the odds of appropriateness explained by the predictor set. Across both outcome categories, 72.7 per cent of cases were accurately classified, with sensitivity lower than specificity. Answers deeming slur appropriation an appropriate choice were correctly predicted in 25.5 per cent of cases compared to 92.6 per cent of answers deeming it inappropriate. Sexuality ($p = .025$) was a significant predictor in the model, as mentioned above, with participants identifying with a sexual minority (i.e. non-heterosexual) being more likely to perceive slur appropriation as an appropriate compared to those not identifying with a sexual minority (OR = 3.44, 95% CI [1.16, 10.18])—see [Figure 6.2](#). None of the other variables or interactions added significantly to the prediction ($p > .05$).

Lastly, we also wanted to determine the chances of our bilingual participants appropriating slurs themselves when responding to insults on the basis of the same factors used in the previous model—namely, language used, LoR, language dominance, ethnicity, sexuality, and a combination of these. Our findings suggest that, out of these factors, only participants’ sexuality and ethnicity played a crucial role, as those who identified as part of a sexual or ethnic minority were more likely to appropriate slurs in their response to insults.

Expanding on this, we ran a second binomial logistic regression to predict the likelihood that bilinguals would appropriate slurs in response to an insult (i.e. ‘appropriation’, yes/no) by language (i.e. L1/L2), LoR, dominance (i.e. Italian/English), ethnicity

3 To reduce the number of levels of the predictor variables, we decided to group all ethnic minorities as ‘non-white’ and sexual minorities as ‘non-heterosexual’.

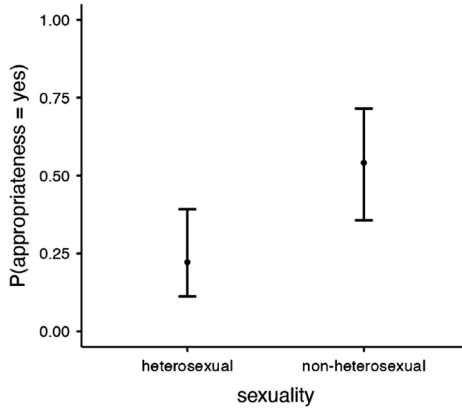


Figure 6.2: Effects plots with 95 per cent confidence interval showing how sexuality significantly predicts the likelihood of perceiving slur appropriation as appropriate.

(i.e. white/non-white), sexuality (i.e. heterosexual/non-heterosexual), and the interaction of language with any of these variables. The overall model was significant, $\chi^2(9) = 32.69$, $p < .001$, with between 17.3 per cent and 34.5 per cent of the variance in the odds of appropriateness explained by the predictor set. Across both outcome categories, 92.4 per cent of cases were accurately classified, with sensitivity lower than specificity. Answers deeming slur appropriation an appropriate choice were correctly predicted in 36.8 per cent of cases compared to 99.3 per cent of answers deeming it inappropriate. Sexuality ($p = .008$) and ethnicity ($p = .044$) were significant predictors in the model, with participants identifying with a sexual minority being more likely to appropriate slurs in response to hate speech compared to those not identifying with a sexual minority (OR = 25.71, 95% CI [2.32, 284.73]), and participants identifying with an ethnic minority being more likely to appropriate slurs in response to hate speech compared to those not identifying with an ethnic minority (OR = 9.36, 95% CI [1.05, 82.99])—see [Figure 6.3](#). None of the other variables or interactions added significantly to the prediction ($p > .05$).

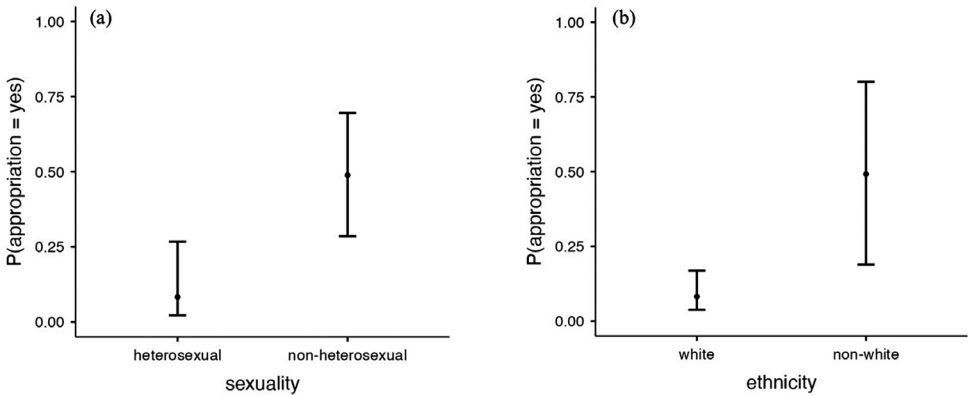


Figure 6.3: Effects plots with 95 per cent confidence interval showing how (a) sexuality and (b) ethnicity significantly predict the likelihood of slur appropriation in response to hate speech.

6.5 Discussion

Our initial analyses for word pair ratings reveal that the only significantly different criterion in the two languages is familiarity, with participants reporting greater familiarity with hate words in L1 Italian than L2 English despite their relatively long period of residence in the UK and high level of proficiency in L2 English. The remaining comparisons returned non-significant results, meaning that our participants' ratings for personal use, imageability, arousal, offensiveness, tabooeness, and valence are similar between languages. These findings seem to indicate that, despite the higher familiarity with L1 words, hate words in the different languages are perceived rather similarly for our participants. This is further corroborated by the positive correlations found between all of the same criteria between languages, demonstrating that high ratings in one language for one criterion correspond to equally high ratings for the same criterion in the other language. Different criteria were also related within each language: while some highlight easily interpretable relationships (e.g. the negative correlation between perceived offensiveness and personal use in English, with

highly offensive terms being used less, and vice versa) some reveal an intricate relationship between criteria, pointing to potential difficulties for participants—thus potential confounds for similar experiments—in teasing some of these aspects apart (specifically, due to the correlations of arousal with familiarity, offensiveness, tabooeness, and imageability, and of imageability with familiarity and offensiveness).

Importantly, the results of our multiple regression models indicate that LoR—but not dominance, contrary to our predictions—has a significant effect on familiarity, personal use, and imageability in Italian, as well as on offensiveness in English (see [Figure 6.1](#)). Specifically, participants residing longer in the L2 country reported lower ratings for familiarity, personal use, and imageability in the L1, suggesting that moving away from the L1 country may have an impact not only on the amount of L1 use but also on the levels of familiarity with L1 words and the ease with which L1 words evoke an image in bilingual speakers' minds. This is in line with the potential linguistic changes that happen in bilingual speakers' L1 as a result of 'the co-activation of language, crosslinguistic transfer or disuse' (Schmid and Köpke 2017: 637; see also the works on L1 attrition previously cited). At the same time, an increase in LoR also corresponds to a significant decrease in the degree of offensiveness perceived in the L2—a finding that seems to suggest that the longer speakers reside in the L2 country (and possibly, the more accustomed they become to L2 sociocultural norms), the less offensive L2 words become for them. This expands on related findings in Gawinkowska, Paradowski, and Bilewicz (2013) and Costa et al. (2014), by demonstrating that the supposedly reduced emotional response in the L2 varies as a function of LoR (and, arguably, proficiency in the L2). The overall non-significant models for both personal use in Italian and perceived offensiveness in English, however, call for further research into the impact of LoR and related variables on hate speech perception in both the L1 and the L2.

On the other hand, contrary to our predictions regarding the effects of bilingualism on slur appropriation, our analyses reveal

that, among our participants, the factors related to the bilingual experience (i.e. language, LoR, dominance, and the interaction of language with each of these factors) do not predict the likelihood of deeming slur appropriation as an appropriate choice, nor the likelihood of them appropriating those slurs (for why this may be the case, see [Section 6.5.2](#)). Instead, identifying as part of a minority has a significant effect on both of these: interestingly, while it is only non-heterosexual participants who deem slur appropriation as a significantly more appropriate choice than heterosexual participants (see [Figure 6.2](#)), participants identifying with sexual and ethnic minorities are significantly more likely to appropriate slurs themselves as a reaction in situations where they are the target of hate speech (see [Figure 6.3](#)). These findings are overall in line with our predictions regarding the effect of identifying with minorities, and the mechanism of slur appropriation (see Bianchi 2014, among other works previously cited).

6.5.1 Implications

The present study has important implications for research in both linguistics and other disciplines. From a linguistic perspective, the research within this chapter represents a first attempt at understanding how bilingualism can influence the way hate speech is perceived, and the way bilingual speakers react when targeted by it. Specifically, this study shows the effects of LoR on familiarity, use, and imageability of hate words in the L1, and on perceived offensiveness in the L2. It also reports the absence of bilingualism-related effects on slur appropriation, at least when analysing behavioural data (but see [Section 6.5.2](#) below). From an interdisciplinary point of view, this chapter highlights the importance of considering bilingualism (and, more generally, psycholinguistic) research findings for disciplines dealing with hate speech from other angles—for instance, philosophy and ethics, as well as law and politics. Slagle (2009: 246) states that one of the difficulties reported in punishing hate speech is the fact that, with regards to emotional distress, ‘such damage can never be quantified or even

proven. Studies such as ours, as well as the evidence reported in our background section (see [Section 6.2](#)), suggest otherwise: the impact of hate speech on a psychological and emotional level is very much real (see Calvert 1997), has several consequences (see Gelber and McNamara 2016), some of which may be long-term (see Boeckmann and Liew 2002: 377), and may actually be influenced by the bilingual experience. Further, the fact that bilingual speakers may perceive hate speech as more or less emotional in their different languages has important implications for law and politics. Imagine two different scenarios: one where a bilingual speaker who no longer lives in their L1 country is insulted by a speaker of their L1 during a holiday in their home country; and another where a bilingual is summoned for jury duty in their L2 country and is asked to evaluate how offensive a hate speech scenario is. In both situations, different factors relating to the bilingual experience of the speakers will impact the perceived degree of offensiveness (and related dimensions) in each language. It is clear, then, that awareness of the bilingual experience factors that influence hate speech perceptions and reactions is crucial, as is further research into these factors.

6.5.2 Limitations and recommendations for future research

Before concluding, it is important to address some of the limitations of the present study that can inform future research on bilingualism and hate speech. Firstly, the number and type of word pairs employed here included both more ‘embodied’ (e.g. dick-head, asshole) and less embodied (e.g. nigger, faggot) insults; as some research points to greater recall of more embodied insults (Wellsby et al. 2010), this may have an impact on some of the variables rated by our participants. Future research could avoid potential confounds by either investigating ratings for a single term, or by employing equally large numbers of more/less embodied insults, and adding this factor into the analysis. Similarly, the number and type of scenarios employed here could be extended

to include more scenarios of different types. In particular, slur appropriation was presented here as a reaction to hate speech (i.e. in scenarios where participants were the direct target of slurs). It would be interesting to investigate whether similar ratings arise in contexts where speakers are appropriating slurs without being targeted in that specific scenario (e.g. a man referring to himself as a 'faggot' when talking with a friend).

The number of participants in the study also represents a limitation, particularly given the relatively unbalanced proportion of participants identifying with ethnic and sexual minorities compared with those who did not (see [Table 6.2](#)). Future research could include more participants, which would also allow for a larger number of possible explanatory variables in the models, as well as increasing power in the analyses. Indeed, with specific regard to the predictor variables used in this study, it must be noted that, although it did not contribute significantly to any of the models, language dominance was collected as a self-reported measure. A better measure of language dominance in the future may consider different predictors that include personal background factors and exposure, use, and attitudes, relating to both the L1 and L2 of the participants (as suggested, for instance, in Schmid and Yilmaz 2018). Importantly, given the significant effect of LoR as a predictor in both L1 familiarity, use, and imageability, and in L2 perceived offensiveness, it would be informative to gather linguistic information on participants' L1 and L2 (i.e. by collecting accurate proficiency measures in the two languages, and investigating attrition widely in more than one L1 domain).

Finally, it is important to consider the nature of the study itself: in this research, behavioural data was collected via an online questionnaire; while participants were not able to change an answer once it had been provided, they could complete the questionnaire in their own time. It would thus be interesting to see whether research collecting more spontaneous data would yield similar results—that is, through the use of technologies such as electroencephalogram (EEG), eye-tracking or pupillometry. This way,

future research could probe more deeply into the subtle changes that occur in the bilingual mind.

6.6 Conclusions

This chapter has outlined the evident links between bilingualism and the emotionally loaded topic of hate speech, and has reported an initial investigation into the understudied impact of bilingualism on hate speech perception and slur appropriation. The analyses of responses to an online questionnaire by 43 Italian residents in the UK demonstrate that despite higher degrees of familiarity with Italian hate words, the same familiarity together with personal use and imageability in Italian tend to decrease with longer LoR in the L2 country, in line with our first hypothesis. At the same time, longer LoR in the L2 country is associated with lower degrees of perceived offensiveness in the L2, which adds to findings in the literature that show lower degrees of emotionality in an L2, by suggesting that the degree of emotionality in the L2 can change according to LoR (and possibly other related variables). With regards to slur appropriation, contrary to our second hypothesis, bilingual experience factors do not seem to predict slur appropriation; on the other hand, participants identifying with sexual minorities are more likely to deem slur appropriation as an appropriate choice, and those identifying with both sexual and ethnic minorities are also more likely to appropriate slurs themselves in situations where they are targeted by hate speech. We have discussed the implications of our study as being central not only to linguistics research, which has long overlooked bilinguals' perception of hate speech specifically, but also as important for other disciplines. We have noted, for instance, what are now well-documented psychological harms of hate speech in response to ethical and philosophical accounts, as well as the need to be aware of the different hate speech perceptions and reactions of bilinguals, which are dependent on a variety of factors and may prove to be fundamental in courts of law and jurisdiction. Finally, we have acknowledged the limitations of our study, and suggested

that future research should not only increase and diversify the number of hate words, scenarios, and participants, but should also collect more linguistic data on both the L1 and L2. Ultimately, the implementation of technologies such as EEG, eye-tracking, or pupillometry may shed further light on the field, by looking at the cognitive and physiological processing of hate speech in late bilingual speakers.

References

- Altarriba, Jeanette. 2003. 'Does cariño equal "liking"? A theoretical approach to conceptual nonequivalence between languages.' *International Journal of Bilingualism* 7 (3): 305–322. <https://doi.org/10.1177/13670069030070030501>.
- Bialystok, Ellen, Fergus I. M. Craik, and Gigi Luk. 2012. 'Bilingualism: consequences for mind and brain.' *Trends in Cognitive Sciences* 16 (4): 240–250. <https://doi.org/10.1016/j.tics.2012.03.001>.
- Bianchi, Claudia. 2014. 'Slurs and appropriation: an echoic account.' *Journal of Pragmatics* 66 (May): 35–44. <https://doi.org/10.1016/j.pragma.2014.02.009>.
- Boeckmann, Robert J., and Jeffrey Liew. 2002. 'Hate speech: Asian American students' justice judgments and psychological responses.' *Journal of Social Issues* 58 (2): 363–381. <https://doi.org/10.1111/1540-4560.00265>.
- Brown, Alexander. 2017a. 'What is hate speech? Part 1: The myth of hate.' *Law and Philosophy* 36 (4): 419–468. <https://doi.org/10.1007/s10982-017-9297-1>.
- Brown, Alexander. 2017b. 'What is hate speech? Part 2: Family resemblances.' *Law and Philosophy* 36 (5): 561–613. <https://doi.org/10.1007/s10982-017-9300-x>.
- Calvert, Clay. 1997. 'Hate speech and its harms: a communication theory perspective.' *Journal of Communication* 47 (1): 4–19. <https://doi.org/10.1111/j.1460-2466.1997.tb02690.x>.
- Carey, James W. 1989. *Communication as Culture: Essays on Media and Society*. Boston, MA: Unwin Hyman.
- Chiril, Patricia, Endang Wahyu Pamungkas, Farah Benamara, Véronique Moriceau, and Viviana Patti. 2022. 'Emotionally informed hate speech detection: a multi-target perspective.' *Cognitive Computation* 14 (1): 322–352. <https://doi.org/10.1007/s12559-021-09862-5>.
- Costa, Albert, Alice Foucart, Sayuri Hayakawa, Melina Aparici, Jose Apesteguia, Joy Heafner, and Boaz Keysar. 2014. 'Your morals depend on language.' Edited by Mariano Sigman. *PLoS ONE* 9 (4): e94842. <https://doi.org/10.1371/journal.pone.0094842>.

- Dewaele, Jean-Marc. 2010. *Emotions in Multiple Languages*. London: Palgrave Macmillan. <https://doi.org/10.1057/9780230289505>.
- Gallo, Federico, Beatriz Bermudez-Margaretto, Yury Shtyrov, Jubin Abutalebi, Hamutal Kreiner, Tamara Chitaya, Anna Petrova, and Andriy Myachykov. 2021. 'First language attrition: what it is, what it isn't, and what it can be'. *Frontiers in Human Neuroscience* 15 (September): 686388. <https://doi.org/10.3389/fnhum.2021.686388>.
- Gawinkowska, Marta, Michał B. Paradowski, and Michał Bilewicz. 2013. 'Second language as an exemptor from sociocultural norms. Emotion-related language choice revisited'. Edited by Mark Aronoff. *PLoS ONE* 8 (12): e81225. <https://doi.org/10.1371/journal.pone.0081225>.
- Gelber, Katharine, and Luke McNamara. 2016. 'Evidencing the harms of hate speech'. *Social Identities* 22 (3): 324–341. <https://doi.org/10.1080/13504630.2015.1128810>.
- Grosjean, François. 2010. *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.
- Grosjean, François. 1989. 'Neurolinguists, beware! The bilingual is not two monolinguals in one person'. *Brain and Language* 36 (1): 3–15. [https://doi.org/10.1016/0093-934X\(89\)90048-5](https://doi.org/10.1016/0093-934X(89)90048-5).
- Harris, Catherine L. 2004. 'Bilingual speakers in the lab: psychophysiological measures of emotional reactivity'. *Journal of Multilingual and Multicultural Development* 25 (2–3): 223–247. <https://doi.org/10.1080/01434630408666530>.
- Hom, Christopher. 2008. 'The semantics of racial epithets'. *Journal of Philosophy* 105 (8): 416–440. <https://doi.org/10.5840/jphil2008105834>.
- Hornsby, Jennifer. 2003. 'Free speech and hate speech: language and rights'. In *Normatività, fatti, valori*, edited by Rosaria Egidi, Massimo Dell'Utri, and Mario De Caro, 393. *Analisi filosofiche* 2. Macerata: Quodlibet.
- The jamovi project. 2021. *jamovi* (version 1.6) [computer software]. <https://www.jamovi.org>.
- Janschewitz, Kristin. 2008. 'Taboo, emotionally valenced, and emotionally neutral word norms'. *Behavior Research Methods* 40 (4): 1065–1074. <https://doi.org/10.3758/BRM.40.4.1065>.
- Kim, Sun Hee Ok, and Donna Starks. 2008. 'The role of emotions in L1 attrition: the case of Korean-English late bilinguals in New Zealand'. *International Journal of Bilingualism* 12 (4): 303–319. <https://doi.org/10.1177/1367006908098573>.
- Kovács, György, Pedro Alonso, and Rajkumar Saini. 2021. 'Challenges of hate speech detection in social media: data scarcity, and leveraging external resources'. *SN Computer Science* 2 (2): 95. <https://doi.org/10.1007/s42979-021-00457-3>.
- MacAvaney, Sean, Hao-Ren Yao, Eugene Yang, Katina Russell, Nazli Goharian, and Ophir Frieder. 2019. 'Hate speech detection: challenges and

- solutions'. Edited by Minlie Huang. *PLOS ONE* 14 (8): e0221152. <https://doi.org/10.1371/journal.pone.0221152>.
- Navarro, Ester, and Andrew R. A. Conway. 2021. 'Adult bilinguals outperform monolinguals in Theory of Mind'. *Quarterly Journal of Experimental Psychology* 74 (11): 1841–1851. <https://doi.org/10.1177/17470218211009159>.
- Pavlenko, Aneta. 2008. 'Emotion and emotion-laden words in the bilingual lexicon'. *Bilingualism: Language and Cognition* 11 (2): 147–164. <https://doi.org/10.1017/S1366728908003283>.
- Pavlenko, Aneta. 2006. 'Bilingual selves'. In *Bilingual Minds: Emotional Experience, Expression, and Representation*, edited by Aneta Pavlenko, 1–33. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781853598746-003>.
- Potts, Christopher. 2007. 'The centrality of expressive indices'. *Theoretical Linguistics* 33 (2): 255–268. <https://doi.org/10.1515/TL.2007.019>.
- Ramírez-Esparza, Nairán, Samuel D. Gosling, Verónica Benet-Martínez, Jeffrey P. Potter, and James W. Pennebaker. 2006. 'Do bilinguals have two personalities? A special case of cultural frame switching'. *Journal of Research in Personality* 40 (2): 99–120. <https://doi.org/10.1016/j.jrp.2004.09.001>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing* (version 4.0). Vienna: R Foundation for Statistical Computing. <https://cran.r-project.org>.
- Schmid, Monika S. 2016. 'First language attrition'. *Language Teaching* 49 (2): 186–212. <https://doi.org/10.1017/S0261444815000476>.
- Schmid, Monika S., and Barbara Köpke. 2017. 'The relevance of first language attrition to theories of bilingual development'. *Linguistic Approaches to Bilingualism* 7 (6): 637–667. <https://doi.org/10.1075/lab.17058.sch>.
- Schmid, Monika S., and Gülsen Yilmaz. 2018. 'Predictors of language dominance: an integrated analysis of first language attrition and second language acquisition in late bilinguals'. *Frontiers in Psychology* 9 (August): 1306. <https://doi.org/10.3389/fpsyg.2018.01306>.
- Slagle, Mark. 2009. 'An ethical exploration of free expression and the problem of hate speech'. *Journal of Mass Media Ethics* 24 (4): 238–250. <https://doi.org/10.1080/08900520903320894>.
- Sorace, Antonella. 2020. 'L1 attrition in a wider perspective'. *Second Language Research* 36 (2): 203–206. <https://doi.org/10.1177/0267658319895571>.
- Sulpizio, Simone, Elisa Vassallo, Remo Job, and Jubin Abutalebi. 2020. 'ITABÙ: Dati preliminari di un database delle parole tabù per l'italiano'. *Giornale italiano di psicologia* 2: 599–614. <https://doi.org/10.1421/97887>.
- Wellsby, Michele, Paul D. Siakaluk, Penny M. Pexman, and William J. Owen. 2010. 'Some insults are easier to detect: the embodied insult detection

- effect'. *Frontiers in Psychology* 1 (November). <https://doi.org/10.3389/fpsyg.2010.00198>.
- Wilson, Rosemary. 2008. 'Foreign languages: a guide for the inhibited'. *Language Learning Journal* 36 (1): 111–115. <https://doi.org/10.1080/09571730801988538>.
- Wilson, Rosemary. 2013. 'Another language is another soul'. *Language and Intercultural Communication* 13 (3): 298–309. <https://doi.org/10.1080/14708477.2013.804534>.
- Zingaretti, Mattia. 2022. 'First Language Attrition in Late Bilingualism: Lexical, Syntactic and Prosodic Changes in English-Italian Bilinguals'. PhD thesis, University of Edinburgh. *Edinburgh Research Archive*. <http://dx.doi.org/10.7488/era/3187>.
- Zingaretti, Mattia, Vasiliki Chondrogianni, D. Robert Ladd, and Antonella Sorace. Forthcoming. 'L1 Attrition *vis-à-vis* L2 Acquisition: Lexicon, Syntax-Pragmatics Interface and Prosody in L1-English L2-Italian Late Bilinguals'. Special issue, *Languages*.