Silvia Pettini

“Don’t Use This Word”

Racist Hate Speech in Online English Dictionaries for Learners

Abstract

This paper presents the preliminary findings of a wider research project which aims at exploring the relationship between racist hate speech and online English lexicography for learners. In particular, this pilot study focuses on the treatment of “ethnophaulisms” (Roback 1944), most commonly referred to as ethnic slurs, in the online editions of three major British English dictionaries for advanced learners: the Longman Dictionary of Contemporary English, the Macmillan English Dictionary for Advanced Learners and the Oxford Advanced Learner’s Dictionary. The rationale behind this paper lies in the connection between two phenomena of the present cultural moment and digital age: the impact of the Internet on dictionary consulting (Jackson 2017; Béjoint 2016) and the alarming increase of online racism and xenophobia at the global level (see Gagliardone et al. 2015). Accordingly, the research objective is to analyse whether and how the three dictionaries selected include, label, define or, more generally, present learners with data reflecting the taboo and offensive nature of these instances of racist hate speech.

Keywords: *online English lexicography, learner’s dictionary, hate speech, ethnic slur, ethnophaulism*

No one is born hating another person because of the colour of his skin, or his background, or his religion. People must learn to hate, and if they can learn to hate, they can be taught to love […].
(Nelson Mandela, *Long Walk to Freedom*)

The idea that language, culture, and society are inextricably linked is well-known to any language learner. We are taught and we more or less directly experience that society alters the world in which we live, and that social and cultural changes often lead to lexical creation or variation. Indeed, language tells us a good deal about our social and cultural beliefs, values, and norms, the way we think, feel, and judge aspects of reality. Language, however, not only reflects

but also shapes our perceptions of reality, that is the culture in which it exists and constantly evolves.

At present, according to Faloppa (2020a), we hear about 'hate speech' or 'hate words' almost every day even if we live in a 'politically correct' cultural climate, even if political correctness "prescribes and proscribes public language for ethnicity, race, gender, sexual preference, appearance, religion, (dis)ability and so on" (Allan and Burridge 2006, 105). In particular, as regards ethnicity, "ethnophaulisms" (Roback 1944), most commonly referred to as ethnic or racial slurs, often hit the headlines as the media continue to report these instances of racist hate speech in the many news concerning verbal and physical attacks against individuals or groups of people due to their nationality, colour or descent. In this regard, young learners of English as a foreign language often belong to ethnically diverse generations, they grew up in multicultural and multilingual environments, and they have become more sensitive toward identity issues than their elders or previous generations, also at the level of language use.

For this reason, when I was invited to give a lecture about learners' dictionaries within the EMLex[1] learner lexicography module, as usually happens when designing a talk, given the specific features of the target audience, that is, a multicultural and multilingual group of young students attending this Erasmus Mundus joint master's degree course and, thus, an ethnically diverse group of future lexicographers, this special learning environment could not but influence my research and I could not but think about how the strong link between language and society currently manifests itself in the global cultural phenomenon of hate speech.

In more detail, my interest in exploring this topic in online dictionaries arises from the interaction between two facets of this digital society. One is, as already mentioned, the alarming rise and impact of hate speech, especially online, that is being amplified by new communication technologies, and that mostly targets ethnic minorities (Gagliardone et al. 2015, 13). The second and more relevant feature of the present cultural moment is the undeniable influence the Internet is having on lexicography in general, and on dictionary consulting in particular. As concerns the latter, many authors have indeed openly highlighted that there is a clear and increasing tendency among dictionary users, including learners of English, to search for lexical information or to deal with language issues online in this digital age (see, among others, Jackson

[1] The European Master in Lexicography (EMLex) is an international Master's degree programme that was launched in 2009. Since 2016, it can be concluded with an Erasmus Mundus Joint Master's Degree, including a semester abroad for all students thanks to partnered universities in Italy, Germany, Portugal, Spain, France, Hungary, Poland, South Africa, among other countries.

2017; Béjoint 2016; Lew and de Schryver 2014; Müller-Spitzer and Koplenig 2014; Lorentzen and Theilgaard 2012).

Taking these aspects into account, the analysis presented in this paper is guided by the following research questions: what if an Internet user, particularly an English learner, consults an online learner's dictionary to search for hate words like ethnic slurs? What do learners of English find out about the taboo and offensive nature of ethnophaulisms when they look them up online, that is in the major online English dictionaries for learners? Do these specialist language-learning resources reflect their politically incorrect status? Do they warn learners against their use? Do they prescribe or even proscribe linguistic behaviour? And, if so, how do they provide learners with linguistic guidance?

Consequently, to address these questions, this paper presents a pilot study that examines the treatment of ethnophaulisms in the online editions of three major English dictionaries for advanced learners, namely, in alphabetical order, the Longman Dictionary of Contemporary English for advanced learners (LDOCE), the Macmillan English Dictionary for Advanced Learners (MEDAL)[2] and the Oxford Advanced Learner's Dictionary (OALD).

For these purposes, Section 1 explores the concept of hate speech, especially ethnic-based hate speech, and contextualizes it in the interrelationship between language, dictionaries, and society. Special attention will be paid to ethnophaulisms in the English-speaking world, as examples of the "language of ethnic conflict" (Allen 1983), in the sense of linguistic expressions of prejudice, racism, xenophobia, and discrimination against the many ethnicities with which Anglophone cultures have had contact over the centuries. Section 2 describes the working methodology used for the quantitative and qualitative analysis presented in Sections 3 and 4, which investigate whether and how a selected group of ethnophaulisms is treated in the three online dictionaries mentioned above. Lastly, Section 5 discusses preliminary findings to finally present some tentative conclusions and suggest future research directions.

1. Hate speech, dictionaries, and society

As Hughes explains (2006, 220), "hate speech is a significant new categorizing term, denoting the deliberate or concerted use of provocative slurs or offensive epithets" and reflecting "the power of language as the bearer of prejudice." However, according to Faloppa (2020a) and Sellars (2016), even if we are familiar with hate speech, the multifaceted nature of this alarming phenomenon makes its definition difficult, also due to the wide and complex range of feelings,

[2] As regards the MEDAL, it is important to specify that the website of this dictionary has been closed on June 30, 2023 (MacMillan Education, online).

moods, and reactions that hate words arouse in each of us. Faloppa (2020a) claims that, although many and diverse definitions exist, descriptions do not seem to be exhaustive, complete, and rigorous enough to be universally adopted. Moreover, many questions remain unanswered as regards the more properly linguistic manifestation of hate speech and the types of more or less explicit verbal and non-verbal expressions falling under this concept (Faloppa 2020a). Until quite recently, indeed, before the umbrella term 'hate speech' established itself, all these expressions of hate were simply labelled as linguistic racism, ethnic slurs, ethnophaulisms, etc. (Faloppa 2020b, 33). Due to space limitations, an in-depth discussion is beyond the scope of this manuscript, but, given the global scope of the phenomenon, this study adopts the following online definition provided by the United Nations on their website and extracted from the UN Strategy and Plan of Action on Hate Speech (UN 2019, 2), where hate speech is:

> Any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.

Such identity factors are of special relevance when examining the relationship between language and culture, and the entries for the words related to these identity issues must be examined to assess whether and how dictionaries, as "a mirror of society" (Iamartino 2020, 36), reflect that dynamic relationship.

As regards ethnicity, nationality, race, colour, and descent, since the development of political correctness in the late 20th century thanks to the civil rights movements (see Pinnavaia 2020), ethnic slurs have turned into a social taboo (Green 2005; Wachal 2002; Zgusta 1998), and racial abuse has become the most derogatory area of language (Allan and Burridge 2006, 105), meaning that among offensive words "ethnic slurs qualify unambiguously" as politically incorrect (Hughes 2010, 12). The degree of tolerance towards politically incorrect language differs across space and time, it depends on the values and belief systems of societies and these change over the centuries (Allan and Burridge 2006, 105). Accordingly, as Allan and Burridge remark (2006, 108), since the sociocultural dynamics behind "the evolving nature of taboo" are reflected by dictionaries and "revealed in changes to lexicographic conventions," since the late 20th century, in response to social pressure, "dictionaries makers have been much more regulative in their policy" and started to "clearly explain, label and exemplify offensive senses and uses in the dictionary's metalanguage."

Considerable academic attention has been paid to the treatment of ethnophaulisms in English monolingual lexicography and studies from this critical perspective have intensified exactly since the late 20th century. The literature is thus extremely rich, but also extremely heterogenous, because scholars have addressed this semantic field proper, or within wider phenomena like bad language, taboo words, and sensitive terms, with different approaches and foci. Differences can be found regarding the number of lemmas and dictionaries examined, the type, size, and target users of dictionaries, and the aspects of the dictionary entries analysed, among others. More importantly, online English lexicography, and online learners' dictionaries in particular, represent a neglected research line (Henderson 2003; Nissinen 2015; Pettini 2023; Žugić and Vuković-Stamatović 2021).

A clear example of the literature's heterogeneity is the wide range of names scholars have used to define the expressions under scrutiny since the late 20th century, including, among others, terms for racial abuse (Burchfield 1980), ethnonyms (Rader 1989), words offensive to groups (McCluskey 1989), racial labels (Murphy 1991, 1997, 1998), racist language (Hauptfleisch 1993; Krishnamurthy 1996), derogatory words for nationality and a racial or cultural group (Norri 2000), ethnocentrism (Benson 2001), racial slurs (Himma 2002), taboo words (Wachal 2002), offensive language (Coffey 2010; Schutz 2002), ethnic slurs or epithets (Pullum 2018; Croom 2015; Henderson 2003), bad language (Pinnavaia 2014), ethnocentricity (Moon 2014), insulting nationality words (Nissinen 2015), and ethnicity terms (Žugić and Vuković-Stamatović 2021). Nevertheless, as Filmer argues (2011, 21-25), "whichever term we use to denote ethnophaulisms," they "are the linguistic manifestation of one culture's attitudes to the other," and, as such, they evidence the language of ethnic conflict.

"The language of ethnic conflict," as seminally defined by Allen (1983), is a long-standing universal phenomenon. As Palmore (1962, 442) explains, "it seems to be universal for racial and ethnic groups to coin derogatory terms and sayings to refer to other ethnic groups," or, in the words of Allan and Burridge (2006, 83), "all human groups, it seems, have available in their language a derogatory term for at least one other group with which they have contact." As Filmer (2011, 18) observes, intolerance for ethnic diversity has manifested itself linguistically since humans began travelling and encountering peoples from other cultures and religions, and the first offensive terms appeared in English in the Middle Ages (Hughes 2006, 147).

In sum, thousands of "ethnophaulisms" exist across languages. According to the Oxford English Dictionary online (OED 2023), this noun was first introduced by Roback (1944) in his *Dictionary of International Slurs (Ethnophaulisms)* and denotes "a contemptuous expression for (a member

of) a people or ethnic group; an expression containing a disparaging allusion to another people or ethnic group."

In the English-speaking world, this semantic field shows stages of growth and decline linked to "periods of migration, religious conflict, war, territorial expansion, political and business rivalry, immigration, and colonialism" (Hughes 2006, 146). Nevertheless, the contacts Anglophone cultures have had with other ethnicities for centuries, and almost always from a dominant position, produced a very large number of ethnophaulisms in English, especially in comparison with other languages such as Italian (Filmer 2011, 2012).

In this light, ethnophaulisms give voice to intercultural or ethnic conflict. According to Reid and Anderson (2010, 100), they represent verbal expressions of ethnic stereotypes based on distinctive and concrete aspects of group practices or characteristics (food preference or habits, physical traits, personal and group names), which allow high-status groups to maintain hierarchies by treating low-status groups as a whole, and the lower the perceived status of a group, the higher the number of and the more negative the nature of ethnophaulisms for that group. As such, they are a prime example of linguistic racism, manifest forms of racial intolerance, they are "the most obvious linguistic manifestation of xenophobia and prejudice against out-groups […] based on the malicious, ironic, or humorous distortion of the target group's identity or 'otherness'" (Hughes 2006, 146). In van Dijk's words (2004, 427), ethnophaulisms are a form of "racist discourse," one of the major discriminatory practices reproducing racism "as a system of social domination and inequality," reproducing "racist prejudices and ideologies" which "in turn are the basis of discriminatory practices (including discourse)." In short, they evidence racist hate speech, as conceived in this paper.

2. Methodology

This section describes some important aspects of the working methodology used in this study to investigate whether and how learners' dictionaries mirror the racist nature of ethnophaulisms and prescribe an appropriate linguistic behaviour.

First, as concerns the selection of the expressions to be analysed, given the focus on online dictionary users, relevant online data have been extracted from Wikipedia, which is the largest online encyclopaedia and one of the top ten most visited websites in the world (Semrush 2023). In more detail, the selection was based on Wikipedia's "List of ethnic slurs," which is, "a list of ethnic slurs or **ethnophaulisms** or ethnic epithets that are, or have been, used as insinuations or allegations about members of a given ethnicity or racial group or to refer to them in a derogatory, pejorative, or otherwise insulting manner" (Wikipedia 2023, online, emphasis

added). The Wikipedia list includes more than 450 ethnic slurs, but the number of terms collected is 281 because further selection has been made according to more language-specific criteria and ethnic slurs belonging to languages other than English have been excluded from the analysis. Examples include *crucco*, *polentone*, and *terrone* in Italian, among many others.

The second element of this research methodology that deserves clarification relates to the selection of dictionaries. The LDOCE, the MEDAL, and the OALD belong to the group of the so-called "Big Five," also including the Cambridge Advanced Learner's Dictionary and the Collins COBUILD Advanced Learner's Dictionary. The reason for excluding the latter two in this initial stage of the research is that there is a significant difference between them and the three works selected as far as the Internet editions of learners' dictionaries are concerned, because their websites prove to be more complex dictionary platforms. While users can access the LDOCE, the MEDAL, and the OALD on their websites and consult the proper online editions of their print counterparts, in the sense that the entries users are presented with belong to those specific learners' dictionaries, the quantity and quality of accessible reference works and lexicographic data provided on Cambridge and Collins websites considerably vary. The websites dictionary.cambridge.org and collinsdictionary.com host several resources of very different types, including both monolingual and bilingual dictionaries, also of languages other than English, including many language combinations, both learner's and general-purpose dictionaries, etc. On Collins' website, moreover, as regards English resources only, search results are multiple on the webpages displayed by the dictionary: they contain a list of entries from different English dictionaries published by Collins, and users must carefully look at the bottom of these entries to find out whether those data are from the learner's dictionary or not. These features make users' experiences across these online editions of the "Big Five" remarkably different. For this reason, they will be examined in the next stage of this ongoing research.

The third methodological element deserving a brief description relates to the sections examined in the dictionaries' entries. In all three dictionaries, indeed, entries present a multi-layered structure typically including, from the top to the bottom of the entries, the following lexicographic data: the headword, audio pronunciation, phonetic transcription, grammatical information (word class, word forms, spelling variants), usage labels, definitions, usage examples, collocations, word origin, phrases, and usage notes. Most of these sections are not scrutinized in this study because they proved to be irrelevant, since they do not contain any data concerning offensiveness, or because they were not provided at all. For instance, examples illustrating the usage of ethnophaulisms are almost always, and quite interestingly one might add, not included. The entry sections scrutinized in this study comprise (1) usage labels, (2)

definitions, and (3) usage notes. Although these data will be discussed more in-depth in Section 4, it is worth briefly mentioning here some of their characteristics in the dictionaries.

Usage labels are shown in italics or capitals, highlighted in colour, sometimes in brackets, and placed below grammatical information or at the beginning of the sense they describe, depending on whether the lemma is monosemous or polysemous. Definitions correspond to the list of senses and subsenses, each displayed on a new line and typically numbered if the lemma is polysemous. In this respect, only semantically relevant senses, that is, senses related to ethnicity or nationality, have been considered in the analysis. Lastly, usage notes, which were found only in the OALD, are placed in isolated coloured boxes below definitions and above word origin.

Another important methodological feature of this pilot study regards usage labels. Special attention is indeed paid to "effect labels" which "relate to the effect that a word or sense is intended by the speaker or writer to produce in the hearer or reader" (Jackson 2013, 113). Effect labels are generally *derogatory* and *offensive*, and the difference between the two typically reflects the effect intended and/or perceived by the people involved. Indeed, as Jackson explains (2013, 113), while *derogatory* means "intending to be disrespectful, *offensive* "may have intent on the part of the speaker or may be unconscious, but" it "could be taken by a hearer as offensive, **either racially** or in some other way" (emphasis added). In the three dictionaries selected, however, the following effect labels have been identified in the analysis: *taboo* and *not polite* in italics in the LDOCE, OFFENSIVE and IMPOLITE in capitals in the MEDAL, and *(offensive)* alone, *(taboo, offensive)* in combination and *(disapproving)*, all in italics and brackets in the OALD.

Concerning the meaning of these effect labels, since the dictionaries examined are the online editions of their printed counterparts, and since online dictionaries do not present 'prefatory' material, it is necessary to draw from the printed editions to look at their definitions in each respective 'labels' section. In the LDOCE (6th edition, emphasis added), *taboo* is "a word or phrase **that should not be used** because it is very rude or **offensive**," while *not polite* is "a word or phrase that is considered rude, or that might **offend** some people." In the MEDAL (2nd edition, emphasis added), OFFENSIVE means "extremely rude or likely to cause **offence**" and IMPOLITE means "likely to **offend** some people." In the OALD (10th edition, emphasis added), *offensive* labels expressions that "are used by some people to address or refer to people in a way that is **very insulting**, **especially in connection with their race**, religion, sex or disabilities," *taboo* labels the expressions that "are likely to be thought by many people to be **very offensive** or shocking" and, finally, *disapproving* labels the expressions that "show that

you feel disapproval or contempt." More importantly, the definitions of the *offensive* and *taboo* labels in the OALD end with a clear prescriptive statement: "You should not use these words." Lastly, going back to methodological aspects of the analysis, the 281 ethnic slurs derived from Wikipedia's list have been classified according to three major criteria: (1) inclusion, and if included, (2) semantic relevance, depending on whether lemmas or one of their senses are related to ethnicity or nationality, and (3) offensiveness. To assess whether and how the three learners' dictionaries signal the potentially discriminatory usage of these words, the analysis has focused on the lexicographic data contained in the entry sections mentioned above, namely effect labels, definitions, and usage notes.

3. Ethnophaulisms between inclusion, semantic relevance, and offensiveness

Before presenting the major findings of this pilot study, it seems worth briefly recapitulating the main objectives of the analysis discussed in the following paragraphs. As already mentioned, this study aims to, first, identify what ethnic slurs, among those listed in Wikipedia's entry, are recorded in the three learners' dictionaries; secondly, among those included, to find out whether these lemmas are relevant or irrelevant to the semantic field, that is, whether they are ethnicity-related or unrelated, and lastly, among those presenting ethnicity-related senses, to show what lemmas are treated as ethnophaulisms, and if so, how they are treated, thus observing how the dictionaries reflect and mark their offensive nature in pertinent lexicographic information.

As concerns the first criterion used to classify data, namely inclusion, as Figure 1 shows, the number of the terms found is comparable across the three works. In descending order, the OALD, the LDOCE, and the MEDAL record 115, 113, and 99 lemmas respectively, representing 41%, 40%, and 35% of the total (281). A slight difference can be observed in the MEDAL, the dictionary containing the smallest quantity of terms, but it is important to highlight in this respect that the number of 'not included' items in this dictionary also incorporates 14 lemmas (seven of which are ethnophaulisms) which were excluded from the classification because they belong to Macmillan's Open Dictionary, which is hosted on the same website, but which is a crowdsourced reference work, meaning that those 14 entries were added by online users.

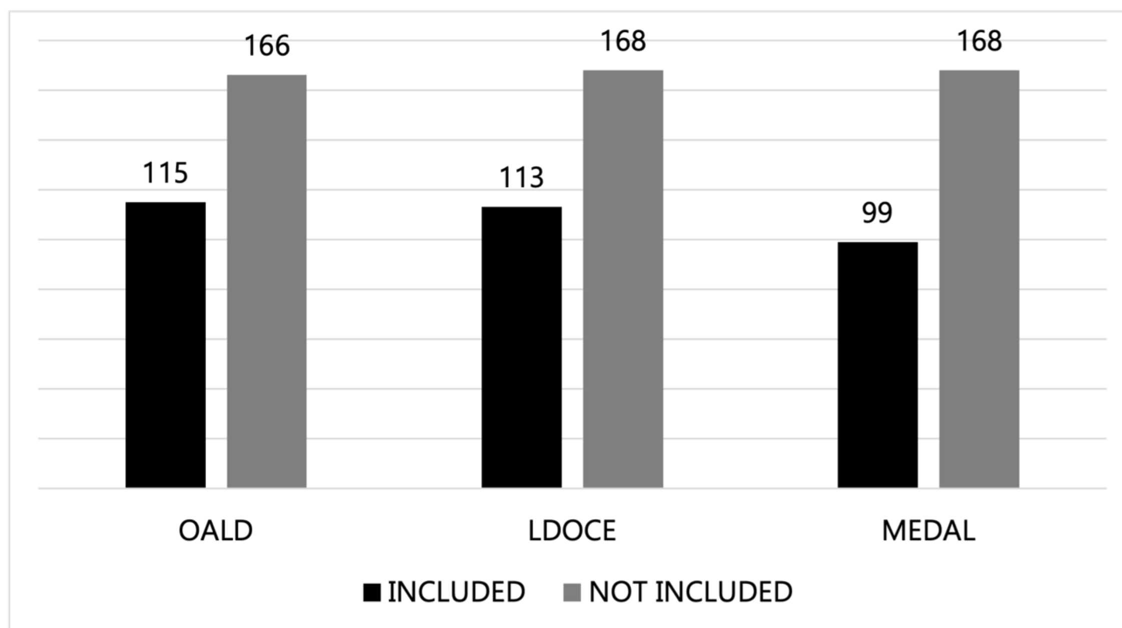

Fig. 1: Terms included in the three dictionaries

Regarding the second and the third criteria adopted to study the set of data, which are semantic relevance and offensiveness, as defined in the previous section, the analysis has shown the following three major groups of terms: (1) ethnicity-unrelated lemmas, which do not present any senses associated with the semantic field of nationality, ethnicity, colour, or descent; (2) ethnicity-related but not offensive lemmas, those entries which are semantically relevant, but they do not contain any data about offensiveness, in the sense that they are not ethnic slurs according to the dictionary; and (3) ethnophaulisms, or, ethnicity-related lemmas whose potential offensiveness is marked by the dictionary.

To go into more detail, an example of ethnicity-unrelated lemmas in all three dictionaries is CHARLIE. According to Wikipedia (2023), this word was used in the 1960s and 1970s as an ethnic slur, especially by black people, in US English to refer to white Americans, and it is used nowadays, still in US English, as an insulting epithet for Vietnamese people. It was introduced as a Vietnam War slang term, derived from NATO's phonetic alphabet letters Victor Charlie (VC) to be used as an abbreviation for Viet Cong, and later, by extension, to name a Vietnamese person. On the contrary, CHARLIE is, in all three learners' dictionaries examined in this manuscript, a *British English old-fashioned* and *informal* expression to simply mean a stupid or silly person.

A good example of ethnicity-related but not offensive lemmas in all three dictionaries is KIWI. Even though the number and the order of senses vary, all three works agree upon the fact that KIWI is an informal expression to refer to "someone" or "a person from New Zealand," not in an

offensive or insulting way. Figure 2 illustrates quantitative findings concerning these three groups in the three dictionaries.

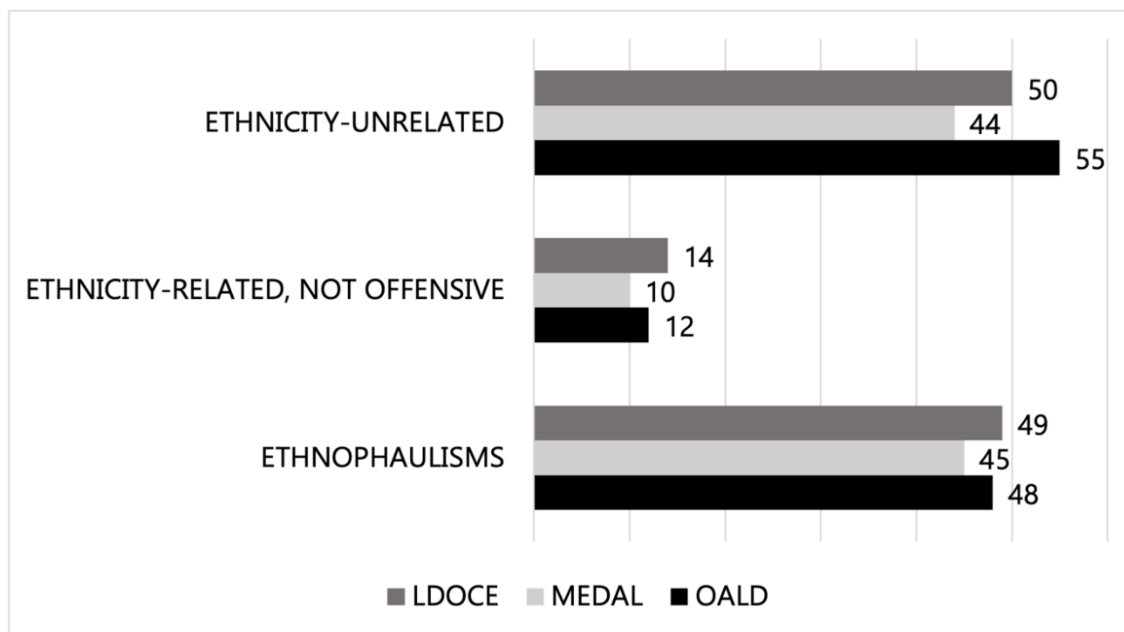

Fig. 2: Groups of lemmas included in the dictionaries

As shown in Figure 2, even if results are similar across the dictionaries, this does not mean that they necessarily correspond and match. Contrariwise, the three dictionaries do not record the same lemmas, and they do not label and/or define all lemmas they include in the same way. In short, symmetries and asymmetries can be observed within each group.

Focusing on similarities, in all three dictionaries: 38 terms are ethnicity-unrelated lemmas, 6 are ethnicity-related but not offensive, and, more importantly, 30 lemmas are ethnophaulisms. The following table is a sample of the classification made in the analysis in terms of inclusion, semantic relevance, and offensiveness. From left to right, the first column contains the terms examined, while the second column suggests their referents, that is the target ethnic group(s) according to Wikipedia. As to categorization, (0) means that the term is not recorded in the dictionary, (1) indicates ethnicity-unrelated lemmas, (2) means that the term is ethnicity-related but not offensive, and (3) is used to classify ethnophaulisms.

Term	Referent	LDOCE	MEDAL	OALD
ABO	aborigine	3	0	0
CHINK	Chinese	3	1	1
COON	black people	3	1	0
CRACKER	poor white people	1	1	3
DAGO	Spanish, Italian, and Portuguese	3	3	0
GOOK	Asian	0	3	0
GYPSY	Romani	2	3	3
HUN	German, WWI	3	0	3
JAP	Japanese	3	0	3
PICCANINNY	black child	0	0	3
POLACK	Polish	3	0	0

Tab. 1: Symmetries and asymmetries between the dictionaries

Based on the data shown in Table 1, for example, ABO is recorded and treated as an ethnophaulism only in the LDOCE. CHINK is included in all three dictionaries, but, while it represents an ethnophaulism in the LDOCE, it is ethnicity-unrelated in the MEDAL and the OALD. DAGO is included and treated as an ethnophaulism in the LDOCE and the MEDAL only. Similarly, information about the potential offensiveness of GIPSY is found in the MEDAL and the OALD, but not in the LDOCE, where it represents an inoffensive ethnicity-related lemma. In the next section, special attention will be paid to the group of ethnophaulisms in the three dictionaries, because their analysis clearly shows similarities and differences in the treatment of these hate words and allows for comparison between the three lexicographic approaches. However, to conclude this section, even if the lemmas of Macmillan's Open Dictionary were excluded, it seems interesting to specify that in this crowdsourced dictionary, among the lemmas listed in Table 2, HUN is an ethnicity-related but not offensive lemma, while POLACK is an ethnophaulism, according to the users who have authored these entries.

4. The treatment of ethnophaulisms

In this section, the three online dictionaries selected will be comparatively examined to show similarities and differences in their approach to the language of racial abuse or racist hate

speech. Emphasis is placed on the treatment of ethnophaulisms, based on markers of offensiveness in the lexicographic data offered in the pertinent entry components described in Section 2, namely (1) effect labels, (2) definitions, and (3) usage notes. As regards the latter, however, it is important to underline that notes are found in the OALD only, so findings relate to this dictionary only. Figure 3 shows the locus of ethnophaulisms' taboo or offensive nature in the reference works examined.

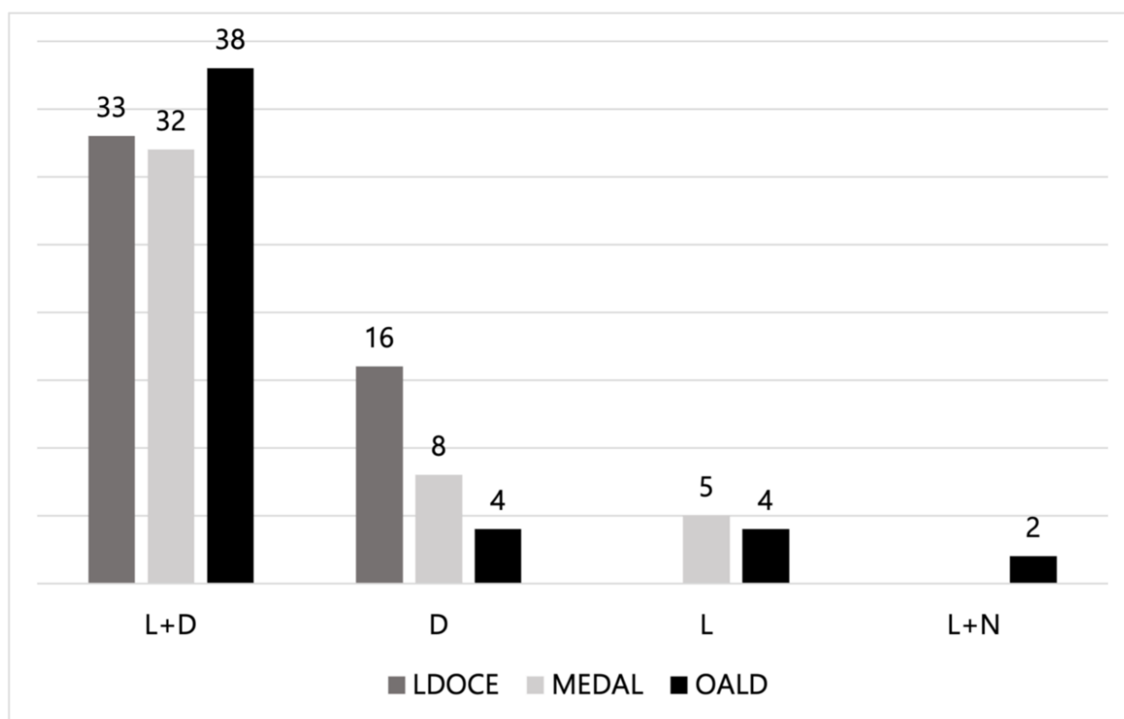

Fig. 3: Markers of offensiveness in the dictionaries

Findings show that all three dictionaries tend to indicate offensiveness in a combination between effect labels and definitions (L+D), meaning that ethnophaulisms are both labelled and defined as such. This feature characterizes 67% of lemmas (33/49) in the LDOCE, 71% of lemmas (32/45) in the MEDAL, and 79% of lemmas (38/48) in the OALD. Moreover, even though to a lesser extent, data concerning offensiveness can be also found in single sections of the entries, namely definitions (D) only (LDOCE 33%, MEDAL 18%, and OALD 9%) and effect labels (L) only (MEDAL 11% and OALD 8%). In addition, and this is specific to the OALD, there are a couple of entries whose potential offensiveness is expressed in effect labels and usage notes (L+N, 4%). As already mentioned, this depends on the fact that notes are included only in the OALD. The following subsections discuss the research findings based on the entry sections mentioned above.

4.1 Effect Labels

As testified to by data shown in Figure 3, labels of effect, thus, either alone or in combination with other entry sections (almost always definitions), are the first and most important lexicographic information users find in respect of the offensiveness of ethnophaulisms. Indeed, overall, effect labels characterize the use of 67% of ethnophaulisms in the LDOCE, always co-occurring with definitions, 82% of ethnophaulisms in the MEDAL, mostly combining with definitions, and 89% in the OALD, again, mostly associated with definitions. With respect to previous studies on the labelling of derogatory words in learners' dictionaries, findings confirm the quite uniform treatment Norri (2000, 91) already observed more than twenty years ago as to "terms for nationality, race/culture," which "are consistently accompanied by a 'negative' indication or comment" because "such words are felt to be particularly insulting."

Like all usage labels, in the three dictionaries scrutinized in this paper, they are highlighted in italics or capitals and with different colours, and placed immediately after the headword or at the beginning of the sense they describe, depending on whether the term is monosemic or polysemic. In particular, as mentioned earlier, effect labels are *taboo* and *not polite* in the LDOCE, *offensive* and *impolite* in the MEDAL, and *offensive*, *taboo* plus *offensive*, and *disapproving* in the OALD. However, as in Figure 4, even though effect labels can vary, *taboo* and *offensive* are respectively used to label almost all lemmas in the LDOCE (31/33) and the MEDAL (36/37). As regards the OALD, on the contrary, *offensive* alone (22/43) and *taboo, offensive* as a combination (19/43) are almost equally used.

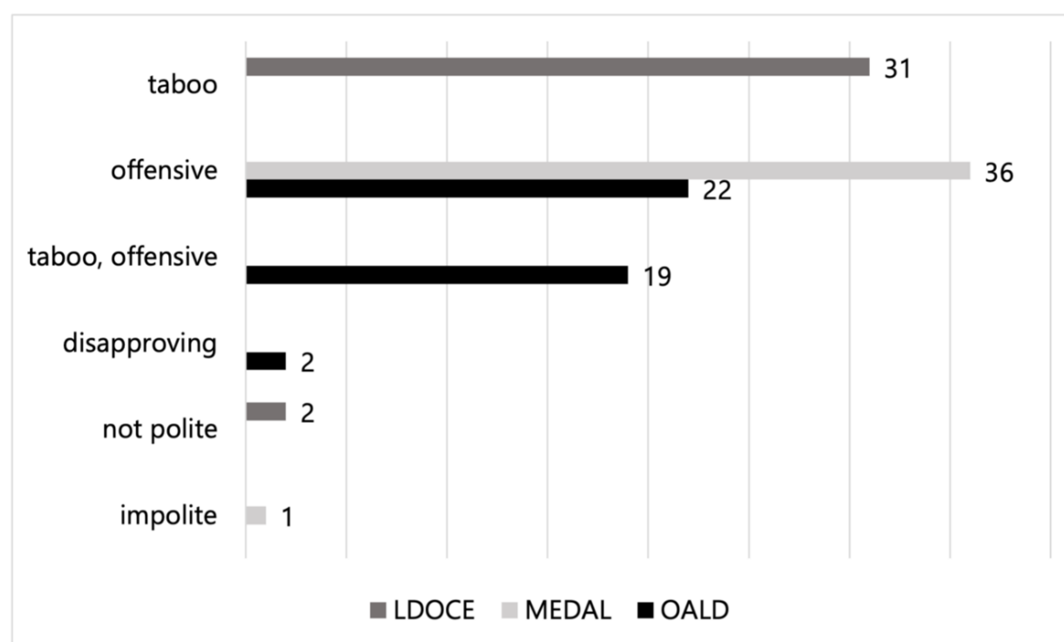

Fig. 4: Effect labels in the dictionaries

For example, PAKI is labelled as *British English* and *taboo* in the LDOCE, *British* and *offensive* in the MEDAL, *British English, taboo, offensive* in the OALD. Similarly, KRAUT is labelled as *taboo* in the LDOCE, *offensive* in the MEDAL and *taboo, offensive* in the OALD.

4.2 Definitions

According to findings shown in Figure 3, definitions, either alone or combined with effect labels, are the second most important marker of offensiveness in the three dictionaries. They represent another important lexicographic component that clearly warns the user against the discriminatory potential of these words. Overall, definitions describe the total number of ethnophaulisms found in the LDOCE (49/49), mostly co-occurring with effect labels (67%) and alone (33%). In the MEDAL, they are used to mark the offensiveness of 89% of terms (40/45), they mostly co-occur with effect labels (71%), but they can also be used as single markers (18%). In the OALD, 88% of ethnophaulisms are defined as offensive words (42/48), mostly in combination with effect labels (79%), but also alone (9%).

Moreover, recurrent patterns in the phrasing of definitions can be observed in all three dictionaries. To illustrate this point, the entries for PAKI and KRAUT represent the tendencies found in the wording of descriptions, as exemplified in the following tables.

PAKI	
LDOCE	A very offensive word for someone from Pakistan or India. Do not use this word.
MEDAL	An extremely offensive word for a South Asian person.
OALD	A very offensive word for a person from Pakistan, sometimes also used in an equally offensive way for people from India or Bangladesh.

Tab. 2: Definitions of PAKI in the dictionaries

KRAUT	
LDOCE	A very offensive word for someone from Germany. Do not use this word.
MEDAL	An insulting word for a German.
OALD	An offensive word for a person from Germany.

Tab. 3: Definitions of KRAUT in the dictionaries

In the LDOCE, 43 definitions (88%) present the following patterns: A/An + (very/extremely) + offensive/insulting word for + (a) someone from + COUNTRY, (b) someone who is + ETHNIC

ADJECTIVE, or (c) a/an + ETHNIC ADJECTIVE + person. More importantly, 70% of definitions typically end with the following prescriptive statement: "Do not use this word," as in the entries for PAKI and KRAUT above.

In the MEDAL, two different patterns have been observed, presenting different orders. The most frequently used phrasing, accounting for 64% of definitions, is A/An + (very/extremely) + offensive/insulting word for + (a) someone from, or (b) a person from + COUNTRY, or (c) a/an + ETHNIC ADJECTIVE + person. The second pattern in the wording of MEDAL's definitions (36%), starts with A/An + (a) ETHNIC ADJECTIVE + person or (b) ETHNIC NOUN, to be later followed in a few instances (4%) by prescriptive statements like the one contained in the definition shown in the following table. The latter contains the definitions of the sense of GIPSY examined in this research and exemplifies how MEDAL's prescriptive suggestion contrasts with the usage preferences explained by the LDOCE and the OALD.

GIPSY	
LDOCE	A member of a group of people originally from India, who traditionally live and travel around in caravans, and who now live all over the world. **Most gypsies prefer to be called Romanies**.
MEDAL	A Romany. **This word is sometimes considered offensive, so you should be careful about using it.**
OALD	A member of an ethnic group (= a group of people with a shared cultural background, language, etc.), originally from Asia, who traditionally travel around and live in caravans. **Many people prefer to use the name Roma or Romani**.

Tab. 4: Definitions of GIPSY in the three dictionaries

Lastly, in the OALD the most frequently used pattern to word definitions (81%) includes: A/An (very) offensive word/name for + (a) a/an ETHNIC ADJECTIVE + person, or (b) a person from + COUNTRY.

Another relevant example is the entry for WOG in the three dictionaries, which relates to racial abuse due to colour, beyond the typical geographical borders found in the above-mentioned examples.

WOG	
LDOCE	*British English taboo* A very offensive word for a black person. Do not use this word.
MEDAL	BRITISH OFFENSIVE An extremely offensive word for a person with black or brown skin.
OALD	*(British English, taboo, offensive, slang)* A very offensive word for a person who does not have white skin.

Tab. 5: Definitions of WOG in the dictionaries

4.3 Usage Notes

As Section 2 explains, the third entry section examined in this paper because it is used in the OALD to warn learners about ethnophaulisms' taboo and offensive nature is the usage note. According to the findings illustrated in Figure 3, however, usage notes play a very minor role if compared to effect labels and definitions. They always co-occur with effect labels and characterize two entries only, namely BLACK and ORIENTAL, representing 4% of the total number of ethnophaulisms in the OALD (48). BLACK is a polysemous word belonging to different word classes, but here it is examined as a noun. When referring to people, according to the dictionary, BLACK is a countable, usually plural, *offensive* noun used to denote "a member of a group of people who have dark skin, especially people who come from or whose ancestors came from Africa." Immediately below the definition of this second sense of BLACK, there is a note in a pale orange isolated box, which states that "Using the noun **black** to refer a person is usually considered offensive, so it is better to use the adjective: *black people • a Black man/woman"* (original emphasis). In this regard, in the entry for BLACK as an adjective, which is not labelled as offensive, this word is defined as "belonging to a group of people who have dark skin, especially people who come from or whose ancestors came from Africa; connected with black people" and the note users are provided in this adjectival entry, under this sense of BLACK, explains that (original emphasis):

> **Black** is the word most widely used and generally accepted in Britain. In the US the currently accepted terms are **African American** or **Black American**. However, the term **person of colour/color** is now often the preferred way of talking in general about people who are not white. Do not confuse this term with *coloured* person, which is now considered offensive.

The second instance of notes regards ORIENTAL, which, like BLACK, can be both adjective and noun, but these two entries differ in some respect. As a noun, ORIENTAL is just labelled as *old-fashioned, offensive,* and defined as "an offensive word for a person from China, Japan or another

country in East Asia." As an adjective, ORIENTAL is labelled as *old-fashioned, sometimes offensive,* and defined as "connected with or typical of the eastern part of the world, especially China and Japan, and the people who live there." Moreover, in the note placed under the only usage example, that is "oriental languages," the OALD further explains that (original emphasis):

> This term is now old-fashioned and can be offensive, as it suggests all people from East Asia are the same and is sometimes used in connection with stereotypes (= fixed but not accurate ideas about a group of people). You can use the term **East Asian** instead, but it is better to be more specific if you can, for example by saying **Chinese** or **Japanese**.

In addition to the notes described above, some extra instances deserve to be discussed even if they do not contain information explicitly relating to offensiveness. Indeed, out of a total of 48 ethnophaulisms examined in the OALD, usage notes appear in eight extra entries, and they vary in the quantity and quality of information offered: some focus on current usage only, while other notes, titled 'Culture,' are more detailed and provide information also relating to the word origin or history, which help the user understand why those lemmas are or can be ethnophaulisms even though offensiveness is not expressed in the usage note.

A prime example of an elaborate usage note is found in the entry for LIMEY. Labelled as *North American English, Australian English, old-fashioned, informal*, this noun represents "a word for a British person that can be offensive." Furthermore, according to the 'Culture' box below this definition:

> The word 'Limey' was used especially by US military forces during the Second World War, often as a way of showing a lack of respect, to mean a British sailor or soldier. It refers to the old practice in the British navy of drinking the juice of limes (= green fruit like lemons) to avoid getting the disease of scurvy which is caused by the lack of vitamin C.

Clear examples of notes which, on the contrary, focus on current usage and suggest appropriate linguistic behaviour, are those found in the entries for HALF-BREED, HALF-CASTE, and MULATTO, all representing hate words due to a person's descent.

HALF-BREED is labelled as *taboo, offensive*, and defined as "an offensive word for a person whose parents are from different races, especially when one is white and the other is a native North American." HALF-CASTE is also labelled *taboo, offensive,* and more briefly defined as "an offensive word for a person whose parents are from different races." MULATTO is a noun labelled as *old-fashioned, offensive* meaning "an offensive word for a person with one black parent and

one white parent." In all three entries, the OALD directly addresses users with the same usage note to suggest that "If you need to refer to a person's background you can use a term such as mixed-race."

5. Conclusions

In light of the increasingly politically incorrect status of these instances of racist hate speech, given the alarming increase of online xenophobia and ethnic-based intolerance around the world, and given the impact of the Internet on dictionary consulting in the current cultural moment, this paper has scrutinized the treatment of ethnophaulisms in the online editions of three major British learners' dictionaries. The main aim of this study has been to investigate whether and how the dictionaries record, label, and describe a sample of terms users might find online, or, at least, on one of the top ten most visited websites in the world, namely Wikipedia. In other words, it has tried to bring to light some possible answers to the following question: if a learner of English, or also an Internet user, looks an ethnophaulism up in an online pedagogical reference work like the dictionaries selected in this paper, is s/he warned against their discriminatory power?

To recapitulate the main points, despite the limitations and delimitations of a pilot study, with respect to the research questions mentioned in the Introduction, the analysis shows that online learners' dictionaries quite clearly reflect the politically incorrect nature of ethnophaulisms and quite consistently tend to present users with relevant data to make them aware that these words are likely to offend. Pertinent usage data like effect labels abound, they appear immediately below the headword or before the sense they describe, and they openly mark ethnophaulisms as *offensive* and/or *taboo*, meaning that learners 'should not' use these epithets. Other important sections of the dictionary entries also contribute to the overall pedagogical function, although to a lesser extent if compared to effect labels, either alone or combined with other sections, namely definitions and usage notes, and they all evidence a quite prescriptive approach of the dictionaries to ethnophaulisms, which can be interpreted as symptomatic of the greater public awareness or sensitivity to offensive racial references Anglophone cultures have developed since the late 20th century, also due to the influence political correctness has exerted on public debate. This is particularly evident in the LDOCE which typically characterizes the ending of the definitions of ethnophaulisms with the imperative comment "Do not use this word," which has been purposefully used in the title of this manuscript. As Pinnavaia observes (2014, 13), given the fundamental role learners' dictionaries play in providing users with "linguistic guidance," and given their "power to significantly influence learners and condition their linguistic

behaviour", users of these online specialist tools can learn not to use ethnophaulisms, to avoid these derogatory and offensive expressions belonging to the English language of ethnic conflict, thanks to the educational strategy adopted by lexicographers to mirror political (in)correctness for non-native and non-expert users. As Pinnavaia remarks (2014, 12), indeed, "the lexicographical team's theoretical plans, the readership's needs, and the social and cultural setting of the moment" are factors to be considered when examining the treatment of offensive words and interpreting lexicographers' prescriptive and proscriptive attitude.

However, further research needs to be carried out to achieve the aim of my ongoing research project on the treatment of ethnophaulisms in online learners' lexicography, of which this pilot study is part. The analysis will be indeed extended to other online dictionaries, such as the Cambridge Advanced Learner's Dictionary and the Collins-COBUILD Advanced Learner's Dictionary, as regards the Big Five, and to similar online dictionaries published by American publishers. Moreover, future research will also explore, if any, the typical semantically organized 'word lists' users find in these language-learning tools, to investigate whether and how these dictionaries classify English vocabulary when related to identity factors like ethnicity, nationality, race, colour, and descent, or, more generally, people in society.

Silvia Pettini *is a Research Fellow in English Language and Translation Studies at Roma Tre University (Italy). Her main research interests include Online Lexicography, Contrastive Linguistics and Audiovisual Translation. Among her most recent publications are* The Translation of Realia and Irrealia in Game Localization *(Routledge, 2022), and papers in journals like* Lexikos, JosTrans, MediAzioni, Lingue e Linguaggi*, and* Status Quaestionis.

Online dictionaries

Longman Dictionary of Contemporary English for Advanced Learners. http://www.ldoceonline.com. All websites last visited on 29/04/2023.

Macmillan English Dictionary for Advanced Learners. https://www.macmillandictionary.com.

Oxford Advanced Learner's Dictionary. https://www.oxfordlearnersdictionaries.com.

Works cited

Allan, Keith and Kate Burridge. *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press, 2006.

Allen, Irving. *The Language of Ethnic Conflict: Social Organization and Lexical Culture*. New York: Columbia University Press, 1983.

Béjoint, Henri. "Dictionaries for General Users: History and Development; Current Issues." *The Oxford Handbook of Lexicography*. Edited by Philip Durkin. Oxford: Oxford University Press, 2016. 7-24.

Benson, Phil. *Ethnocentrism and the English Dictionary*. London: Routledge, 2001.

Burchfield, Robert. "Dictionaries and Ethnic Sensibilities." *The State of the Language*. Edited by Leonard Michaels and Christopher Ricks. Berkeley: University of California Press, 1980. 15-23.

Coffey, Stephen. "'Offensive' Items, and Less Offensive Alternatives." *English Monolingual Learners' Dictionaries. Proceedings of the XIV EURALEX International Congress*. Edited by Anne Dykstra and Tanneke Schoonheim. Leeuwarden: Fryske Akademy Afuk, 2010. 1270-1281.

Croom, Adam. "The Semantics of Slurs: A Refutation of Coreferentialism." *Ampersand* 2 (2015): 30-38.

Faloppa, Federico. "L'hate speech, questo sconosciuto." *Lingua Italiana Treccani* 2020a. https://www.treccani.it/magazine/lingua_italiana/speciali/Hate_speech/01_Faloppa.html.

---. *#ODIO. Manuale di resistenza alla violenza delle parole*. Torino: UTET, 2020b.

Filmer, Denise. *Translating Racial Slurs: A Comparative Analysis of Gran Torino Assessing Transfer of Offensive Language between English and Italian*. Published MA Thesis, Durham University, 2011. http://etheses.dur.ac.uk/3337.

---. "The 'Gook' Goes 'Gay': Cultural Interference in Translating Offensive Language." *inTRAlinea* 14 (2012). http://www.intralinea.org/archive/article/1829.

Gagliardone, Iginio, et al. *Countering Online Hate Speech*. Paris: UNESCO Publications, 2015.

Green, Jonathon. "Language: Wash Your Mouth Out." *Critical Quarterly* 46.1 (2004): 107-111.

Hauptfleisch, D.C. "Racist Language in Society and in Dictionaries: A Pragmatic Perspective." *Lexikos* 3 (1993): 83-139.

Henderson, Anita. "What's in a Slur?". *American Speech* 78.1 (2003): 52-74.

Himma, Kenneth. "On the Definition of Unconscionable Racial and Sexual Slurs." *Journal of Social Philosophy* 33.3 (2002): 512-522.

Hughes, Geoffrey. *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, And Ethnic Slurs in The English-Speaking World*. Armonk: M.E. Sharpe, 2006.

---. *Political Correctness: A History of Semantics and Culture*. Malden: Wiley Blackwell, 2010.

Iamartino, Giovanni. "Lexicography as a Mirror of Society: Women in John Kersey's Dictionaries of the English Language." *Textus* 33.1 (2020): 35-67.

Jackson, Howard. "English Lexicography in the Internet Era." *The Routledge Handbook of Lexicography*. Edited by Pedro Fuertes-Olivera. London: Routledge, 2017. 540-553.

---. *Lexicography: An Introduction*. London: Routledge, 2013.

Krishnamurthy, Ramesh. "Ethnic, Racial and Tribal: The Language of Racism?". *Texts and Practices: Readings in Critical Discourse Analysis*. Edited by Carmen Caldas-Coulthard and Malcolm Coulthard. London: Routledge, 1996. 129-149.

Lew, Robert and Gilles-Maurice de Schryver. "Dictionary Users in the Digital Revolution." *International Journal of Lexicography* 27.4 (2014): 341-359.

Lorentzen, Henrik and Liisa Theilgaard. "Online Dictionaries – How Do Users Find Them and What Do They Do Once They Have?". *Proceedings of the XV EURALEX International Congress*. Edited by Ruth Vatvedt Fjeld and Julie Matilde Torjusen. Oslo: University of Oslo, 2012. 654-660.

McCluskey, John. "Dictionaries and Labeling of Words Offensive to Groups, with Particular Attention to the Second Edition of the OED." *Dictionaries* 11 (1989): 111-123.

MacMillan Education. "Macmillan Dictionary to Close After 14 Years of Online Excellence." 2023. https://macmillaneducation.secure.force.com/help/bg_FAQArticle?id=kA04H0000005kwWSAQ&language=en%3Cb.

Moon, Rosamund. "Meanings, Ideologies, and Learners' Dictionaries." *Proceedings of the XVI EURALEX International Congress: The User in Focus*. Edited by Andrea Abel, Chiara Vettori and Natascia Ralli. Bolzano: Institute for Specialised Communication and Multilingualism, 2014. 85-105.

Müller-Spitzer, Caroline and Alexander Koplenig. "Online Dictionaries: Expectations and Demands." *Using Online Dictionaries*. Edited by Caroline Müller-Spitzer. Berlin: de Gruyter, 2014. 143-188.

Murphy, M. Lynne. "Afrikaans, American, and British Models for South African English Lexicography: Racial Label Usage." *Lexikos* 7 (1997): 153-164.

---. "Defining People: Race and Ethnicity in South African English Dictionaries." *International Journal of Lexicography* 11.1 (1998): 1-33.

---. "Defining Racial Labels: Problems and Promise in American Dictionaries." *Dictionaries* 13.1 (1991): 43-64.

Nissinen, Suvi. *Insulting Nationality Words in Some British and American Dictionaries and in the BNC*. MA Thesis, University of Tampere, 2015.

Norri, Juhani. "Labelling of Derogatory Words in Some British and American Dictionaries." *International Journal of Lexicography* 13.2 (2000): 71-106.

OED (Oxford English Dictionary Online). "Ethnophaulism." https://www.oed.com.

Palmore, Erdman. "Ethnophaulisms and Ethnocentrism." *American Journal of Sociology* 67.4 (1962): 442-445.

Pettini, Silvia. "The Language of Ethnic Conflict in English Online Lexicography: Ethnophaulisms in 'powered by Oxford' Lexico.com." *Lexikos* 33 (2023): 299-317.

Pinnavaia, Laura. "Defining and Proscribing Bad Language Words in English Learner's Dictionaries." *Enforcing and Eluding Censorship: British and Anglo-Italian Perspectives*. Edited by Giuliana Iannaccaro and Giovanni Iamartino. Cambridge: Cambridge Scholars, 2014. 217-230.

---. "Tracing Political Correctness in Bilingual English-Italian Dictionaries." *Textus* 33.1 (2020): 87-106.

Pullum, Geoffrey. "Slurs and Obscenities: Lexicography, Semantics, and Philosophy." *Bad Words: Philosophical Perspectives on Slurs*. Edited by David Sosa. Oxford: Oxford University Press, 2018. 168-192.

Rader, James. "People and Language Names in Anglo-American Dictionaries." *Dictionaries* 11 (1989): 125-138.

Reid, Scott and Grace Anderson. "Language, Social Identity, and Stereotyping." *The Dynamics of Intergroup Communication*. Edited by Howard Giles, Scott Reid, and Jake Harwood. New York: Peter Lang, 2010. 91-104.

Roback, Abraham. *A Dictionary of International Slurs*. Cambridge: Sci-Art Publishers, 1994.

Schutz, Rik. "Indirect Offensive Language in Dictionaries." *Proceedings of the X EURALEX International Congress*. Edited by Anna Braasch and Claus Povlsen. Copenhagen: CST, 2002. 637-641.

Sellars, Andrew. "Defining Hate Speech." *Berkman Klein Center Research Publication* 20 (2016). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2882244.

Semrush. Top Websites. 2023. https://www.semrush.com/website/top.

UN (United Nations). *UN Strategy and Plan of Action on Hate Speech*. 2019. https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf.

van Dijk, Teun. "Racism, Discourse and Textbooks: The Coverage of Immigration in Spanish Textbooks." *Paper for the Symposium on Human Rights in Textbooks,* organized by the History Foundation, Istanbul, April 2004. https://discourses.org/wp-content/uploads/2022/07/Teun-A.-van-Dijk-2010-Racism-discourse-and-textbooks.pdf.

Wachal, Robert. "Taboo or Not Taboo: That is the Question." *American Speech* 77.2 (2002): 195-206.

Wikipedia. "List of ethnic slurs." 2023. http://en.wikipedia.org/wiki/List_of_ethnic_slurs.

Zgusta, Ladislav. "Some Developments in Lexicography, Past and Present." *Symposium on Lexicography IX.* Edited by Jens Eric Mogensen, Viggo Hjornager Pedersen and Arne Zettersten. Halle: Max Niemer Verlag, 1998/2000. 11-26.

Žugić, Dragica and Milika Vuković-Stamatović. "Problems in Defining Ethnicity Terms in Dictionaries." *Lexikos* 31 (2021): 177-194.