

# “View and Hide Definitions” of Racist Hate Speech: Ethnophaulisms in Google’s English Dictionary

Silvia Pettini<sup>1</sup>

<sup>1</sup> Department of Foreign Languages, Literatures and Cultures, Roma Tre University, Rome, Italy

Correspondence: Silvia Pettini, Department of Foreign Languages, Literatures and Cultures, Roma Tre University, Via del Valco di San Paolo 19, 00146 Rome, Italy.

Received: March 2, 2024      Accepted: May 2, 2024      Online Published: May 20, 2024

doi:10.5539/ijel.v14n3p1      URL: <https://doi.org/10.5539/ijel.v14n3p1>

## Abstract

This paper aims to foster debate about the language of racist hate speech in online English lexicography. For this purpose, it presents a study on the treatment of ethnophaulisms, or ethnic slurs, in “powered by Oxford Languages” Google’s English dictionary. The focus is indeed on the perspective of the general user of the Internet, in light of the connection between two facets of this digital age. The first one is the strong and growing tendency among Internet users to ‘google’ their language issues. The second one is the alarming increase in cases of hate speech online, most of which are based on ethnicity and nationality, according to reports by the United Nations. Consequently, the free and pervasive content of Google’s English dictionary represents a case in point to investigate whether and how online users are warned against the power of these hate words. A selected sample of 285 English ethnic slurs have been looked up in the dictionary and, if recorded, their entries have been scrutinised to identify lexicographic data regarding their semantic relevance and offensiveness. Findings show that the majority are included, they mostly present ethnicity-related senses, but less than half of the total are treated as ethnophaulisms. In this respect, the major dictionary markers indicating offensiveness are effect labels, predominantly alone or combined with definitions. Relative to their size, thus, ethnophaulisms in Google’s English dictionary are clearly described as offensive or derogatory expressions, thus making online users aware of their hurtful nature.

**Keywords:** English lexicography, online dictionaries, hate speech, linguistic racism, ethnic slur, ethnophaulism, “powered by Oxford Languages” dictionary content, Google’s English dictionary

## 1. Introduction

In order to contribute to the debate about the language of racist hate speech in online English lexicography focusing on the perspective of the general user of the Internet, this paper presents the findings of the second stage of a research project (Pettini, 2023a) which examines the treatment of ethnic slurs, or “ethnophaulisms” (Roback, 1944), in the so-called “powered by Oxford Languages” content, a prime example of which is Google’s English dictionary, the topic of the present article. As explained by Ferrett and Dollinger (2021) and by Pettini (2021, 2023a, 2023b), “powered by Oxford Languages” is the content licensed for use by Oxford University Press (OUP hereafter) to the preinstalled dictionaries of market-leading operating systems like Microsoft and Apple and, more importantly, to the most popular search engine in the world, that is Google (Note 1). As concerns the latter, the use of search operators like ‘define ...’, ‘... definition’, and similar wording in Google’s search bar in fact displays and explicitly mentions Oxford data, because “Google’s English dictionary is provided by Oxford Languages” (Oxford Languages, n.d., online). As a result, thanks to the partnership OUP is in with this technology titan, the free content they license is remarkably widespread and the influence it may have on online users is profound, especially as far as linguistic manifestations of what the United Nations (UN) regard as the global alarming phenomenon of hate speech are concerned (see UN, 2019).

This study was indeed prompted by an interest to investigate the relationship between the dominant market position of “powered by Oxford Languages” dictionary content and two facets of this information age. The first one is the strong and growing tendency among Internet users to resort to general search engines like Google for answers to their language issues, a tendency presenting a threat to more traditional and specialised dictionaries, as observed by many scholars (see, for example, Béjoint, 2016; Jackson, 2017; Lew, 2011; Lew & De Schryver, 2014; Lorentzen & Theilgaard, 2012; Müller-Spitzer & Kopleinig, 2014). The second relevant feature of the present

cultural moment is, as previously mentioned, the escalation in cases of online hate speech, the majority of which attack or discriminate against individuals or minority groups on the grounds of ethnicity, nationality, race, colour, and descent, according to the UN (Gagliardone, Gal, Alves, & Martinez, 2015; UN, 2019).

Based on these premises, the research questions this study aims to answer are the following: what do general users of the Internet learn about ethnic slurs when they google these hate words? What do users find out about the discriminatory power of ethnophaulisms on the most visited website in the world? Does Google's English dictionary describe their offensive nature? Does it, and if so, how does it warn users against their use? Does it offer prescriptive or even proscriptive information? To reply to these questions, after exploring the concept of ethnophaulisms as instances of racist hate speech in the light of the dynamic relation linking language, dictionaries, and society, the materials and method of this study are described in Section 2 to contextualise the analysis of a selected group of ethnophaulisms presented in Section 3. Lastly, some conclusions and future research directions are suggested in Section 4.

### 1.1 Theoretical Background

As Palmore (1962, p. 442) seminally observed, "it seems to be universal for racial and ethnic groups to coin derogatory terms and sayings to refer to other ethnic groups", that are "ethnophaulisms". Similarly, in the words of Allan and Burrige (2006, p. 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". Intolerance towards ethnic diversity has manifested itself linguistically since humans began travelling and encountering other cultures (Filmer, 2011, p. 18), with the first offensive ethnic epithets appearing in English in the Middle Ages (Hughes, 2006, p. 147). Ethnophaulisms, thus, represent an ancient and cross-linguistic tendency. First proposed as a neologism by Abraham Aaron Roback (1944) in his *Dictionary of International Slurs (Ethnophaulisms)*, this term derives from a combination of Greek words meaning to disparage an ethnic group, it refers to group insults and is "a word used to deprecate a group of people, in other words, an ethnic slur" (Nuessel, 2008, p. 29). As defined by the Oxford English Dictionary online (2024, online), an ethnophaulism is indeed "a contemptuous expression for (a member of) a people or ethnic group; an expression containing a disparaging allusion to another people or ethnic group". Even though, as Hughes explains (2006, p. 146), this semantic field in English is characterised by phases of growth and decline, resulting from "periods of migration, religious conflict, war, territorial expansion, political and business rivalry, immigration, and colonialism", in the English-speaking world, "the practice of stigmatizing foreigners (...) has been established and *de rigueur*" for centuries (Hughes, 2006, p. 220, original emphasis), and the contact Anglophone cultures have had with other ethnicities, mainly from a dominant position, has given rise to a vast array of ethnophaulisms, apparently higher than in other languages (Filmer, 2011, 2012).

At present, in the politically correct cultural climate which "prescribes and proscribes public language for ethnicity, race, gender, sexual preference, appearance, religion, (dis)ability" (Allan & Burrige, 2006, p. 105), "ethnophaulisms constitute what has more recently been labelled as 'hate speech'" (Nuessel, 2008, p. 30), an alarming phenomenon which is attracting considerable and interdisciplinary interest inside and outside academic circles, with international and national organisations and governments actively committed to counter this global challenge, of which, however, there is no universally agreed-on definition. According to Bianchi (2022), hate speech is one of the most compelling and debated issues of our time. Yet, the complex nature of the manifestations of this phenomenon makes its definition difficult, as Faloppa (2020a) and Sellars (2016) claim, and the many diverse descriptions available today do not seem to be complete and rigorous enough to be universally accepted. Nevertheless, hate speech has established itself as an umbrella term, or, in Hughes' words (2006, p. 220), as "a significant new categorizing term" encompassing a range of linguistic forms of hate: single words, phrases, and longer expressions that communicate derision, contempt, and hostility towards individuals or social groups targeted for identity traits such as ethnicity, religion, gender, sexual orientation and (dis)ability (Bianchi, 2022; Faloppa, 2020b). As defined by the United Nations on their website and in the Strategy and Plan of Action on Hate Speech (UN, 2019, p. 2), the latter 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". Similarly, in Google's English dictionary, HATE SPEECH is described as "abusive or threatening speech or writing that expresses prejudice on the basis of ethnicity, religion, sexual orientation, or similar grounds" and it is meaningfully exemplified in "we don't tolerate any form of hate speech".

Hate speech reflects "the power of language as the bearer of prejudice" (Hughes, 2006, p. 220), but the degree of tolerance towards hate words differs across space and time, since it depends on the values and belief systems of societies and these change over the years (Allan & Burrige, 2006, p. 105). As concerns racial prejudice and abuse, ethnic slurs have acquired an increasingly offensive status since the late 20<sup>th</sup> century, when the civil rights movements led to the development of political correctness (see Green, 2005; Pinnavaia, 2020; Wachal, 2002;

Zgusta, 1998/2000). In this sense, indeed, ethnic slurs “qualify unambiguously” as politically incorrect (Hughes, 2010, p. 12).

Since the late 20<sup>th</sup> century, the evolution of ethnic slurs into the most derogatory area of language has affected lexicography too. First, the increasing awareness and sensitivity in society have been reflected by monolingual dictionaries and “revealed in changes to lexicographic conventions” (Allan & Burrige, 2006, p. 105), meaning that lexicographers have become “much more regulative in their policy” and they have started to “clearly explain, label and exemplify offensive senses and uses in the dictionary’s metalanguage” (Allan & Burrige, 2006, p. 108). Secondly, also scholarly work from this critical perspective has intensified since the late 20<sup>th</sup> century and has established a long and productive line of studies on the treatment of ethnophaulisms in English lexicography. However, more relevantly here, very little attention has been paid to online general-purpose dictionaries. To the best of my knowledge, they are indeed scrutinised in the following four publications, which are briefly reviewed here in chronological order.

Henderson (2003) examines ethnic slurs used for black and white Americans in five monolingual dictionaries of English, two of which are online sources, namely *Merriam-Webster’s Online Collegiate Dictionary* (MWOCD 2001) and the historical *Oxford English Dictionary Online* (OEDO 2002). In her research, Henderson shows that the OEDO records the highest number of slurs, but does not consistently use labels to treat them, as opposed to the MWOCD, which tends to describe words applied to black people and words applied to white people as *offensive* and *disparaging* respectively. Nissinen (2015) describes the treatment of 37 potentially offensive nationality words in a total of twenty dictionaries of different types and sizes, British and American, among which the OEDO (2015) is the only non-learner’s online work. Like in Henderson’s study (2003), this reference tool records the highest number of terms, 69% of which are described as *offensive* through the use of labels, definitions, usage notes, or a combination of these sections. Žugić and Vuković-Stamatović (2021) present a study based on the qualitative multilingual analysis of the definition of one single lemma, namely the word for Albanian, in nineteen online and freely accessible monolingual dictionaries. As to English, the authors investigate only three general-purpose works including the Merriam-Webster.com dictionary site (2020), *The Random House Unabridged Dictionary* hosted on the Dictionary.com site (2020) and *The American Heritage Dictionary of the English Language* (2020) hosted on Thefreedictionary.com site. Lastly, Pettini (2023a), which is the research this paper aims to extend, explores the treatment of 285 ethnophaulisms in the free online edition of the *Oxford Dictionary of English* (ODEO) hosted on the “powered by Oxford” Lexico.com site (Note 2). Her analysis shows that the dictionary “quite clearly reflects the taboo nature of ethnophaulisms and quite consistently tends to warn the Internet user against the potentially racist and xenophobic power of these words” (Pettini, 2023a, p. 313). In more detail, out of a total of 227 terms included (80%), 64% are indicated as ethnophaulisms with clear usage data. These include the labels *derogatory* and *offensive*, which characterise 91% of slurs, mostly alone (74%) or in combination with other entry sections (17%), namely definitions, usage notes, and word origin. What seems to emerge, according to Pettini (2023a, p. 314), is “a quite prescriptive approach of the dictionary to ethnophaulisms and, thus, to racial abuse, which might be interpreted as symptomatic of greater public awareness and sensitivity to possibly offensive racial references”. In this light, the study presented here aims to replicate Pettini’s work (2023a) to provide a more complete picture of the interface between hate words like ethnic slurs and the general user of the Internet.

## 2. Materials and Method

This research investigates the treatment of ethnophaulisms in the free and pervasive “powered by Oxford Languages” Google’s English dictionary (GED hereafter). The aim is to explore first whether these words are recorded and, secondly, if so, whether and how dictionary entries signal their hateful nature. The methodology used for this study draws on Pettini (2023a) and combines quantitative and qualitative approaches. The following paragraphs describe the most important methodological aspects as regards 1) the dictionary examined, and 2) data selection and analysis. Both choices, as mentioned above, have been guided by this study’s focus on the experience of the Internet general user.

### 2.1 Google’s English Dictionary

Launched in 2009, Google’s Dictionary is the online dictionary service of Google which, as regards English, can be accessed via operators like ‘define ...’, ‘... definition’, ‘meaning of ...’ ‘... meaning’, ‘what does ... mean’, and similar phrasing in Google search. More relevantly, its content is licensed by Oxford Languages (n.d., online), OUP’s provider of digital language data. In this respect, it is worth adding that this partnership is explicitly mentioned in the “Definitions from Oxford Languages” wording placed in the top left-hand corner of each entry, next to a clickable “Learn more” which directs users to a support webpage explaining that Google’s “dictionary boxes” show data “from third-party expert sources” (Google Search Help, n.d., online).

Regarding the use of this dictionary, when users type in proper operators with the word or phrase of interest in Google's search bar, if the entry is included, they are offered a dictionary box containing the headword and its definition(s). If available, users can also find data like phonetic transcription, pronunciation audio, word class, grammatical information, the list of senses, one or more usage examples, synonyms, phrases, origin, also in a chart form, and use over time via the Ngram Viewer tool. As lexicographers at Oxford Languages clarify (n.d., online), "this dictionary is regularly updated with evidence from one of the world's largest lexical research programmes, and features over 350,000 words and phrases" belonging to many English varieties. Like other Oxford dictionaries, Google's content is created with the evidence-based approach of descriptive lexicography, aimed at monitoring development in real-life examples of spoken and written language gathered through a series of corpora curated by lexicographers at Oxford Languages (n.d., online).

More relevantly, as regards content, editors at Oxford Languages (n.d., online) explain that, given the descriptive nature of this tool, "vulgar and offensive words" are included "because such terms are part of a language's lexicon", but their status is clearly identified with labels, their changes over time, if any, are continuously monitored and integrated into the dictionary to reflect current usage thanks to lexicographers' language research programme. However, as they state to conclude this section about offensive language, feedback from users is always welcome, especially to signal cases which, in users' view, do not meet Oxford's "rigorous quality standards, whether due to *changing cultural sensitivities* or for other reasons" (Oxford Languages, n.d., online, emphasis added). Similarly, on Google's support webpage users learn that, since this global platform licenses data, "dictionary results don't reflect the opinions of Google" and, as to "how Google handles offensive definitions", these are included "for more complete results", but Google's "partners label these terms as vulgar, derogatory, or otherwise offensive to provide proper context" (Google Search Help, n.d., online). Moreover, like Oxford Languages, at the end of this webpage, Google explains how to "report a problem with a dictionary box", but they simultaneously recommend users to "also send feedback directly to the third-party source that provided the definition" (Google Search Help, n.d., online). In this respect, it is worth repeating that Google's partners are explicitly cited in each dictionary box.

The sections about offensive language provided by Oxford Languages and Google seem to suggest that they are fully aware of potential issues arising from cultural sensitivities, and they try to forestall criticism with specific explanatory notes. As to Oxford Languages in particular, this might come as no surprise given the gender-related controversy targeting OUP a few years ago (see Pettini, 2021, 2023b). After all, given the influence of the Internet on dictionary consulting, since most users tend to google their language issues in this digital age (Béjoint, 2016; Jackson, 2017), OUP's partnership with the most visited website in the world inevitably makes them more prone to public criticism, especially as far as sensitive issues are concerned.

## 2.2 Data Selection and Analysis

As in Pettini (2023a), the lexemes to be analysed in this study have been extracted from Wikipedia, because it is the world's largest online encyclopaedia and one of the top ten most visited websites globally (Semrush, 2024). More specifically, data were derived from the "List of ethnic slurs" (Wikipedia, 2024, online), which is the first site that appears in Google search results when a general user of the Internet googles 'ethnic slur', and in which the latter is defined as "a term designed to insult others on the basis of race, ethnicity, or nationality" (Wikipedia, 2024, online). As of March 2024, out of a total of 430 items listed in this Wikipedia entry, comprising almost 150 terms of several different languages, 285 English terms have been collected and the most relevant lexicographic data in their entries, if any, have been observed to describe their treatment in GED. Concerning this point, the analysis has concentrated on usage labels and definitions (Note 3). Even though their description is detailed in Section 3, some of their characteristics are worth mentioning here.

In GED usage labels are highlighted in bold small capitals and placed above or at the beginning of the sense they describe. More precisely, this study scrutinises "effect labels", those relating to "the effect that a word or sense is intended by the speaker or writer to produce in the hearer or reader", namely "derogatory" and "offensive" (Jackson, 2013, p. 113). Their difference, as the author explains (Jackson, 2013, p. 113), lies in people's intention and/or perception, since *derogatory* means "intending to be disrespectful", while *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). Likewise, GED itself defines DEROGATORY as "showing a critical and disrespectful attitude" and OFFENSIVE as "causing someone to feel resentful, upset, or annoyed". Regarding the second entry section analysed, definitions include a list of senses and subsenses. If the lemma is polysemous, each of them is displayed on a new line, senses are numbered and subsenses are bulleted. In cases of polysemy, this study has examined only the ethnicity-related sense of each term, a delimitation used to interpret data as semantically relevant.

As to the criteria adopted in the analysis, whose findings will be discussed in Section 3, the 285 ethnophaulisms have been examined first in terms of inclusion, to verify whether they are recorded or not, and, if included and semantically pertinent, meaning ethnicity-related lemmas or senses of lemmas, their offensiveness has been assessed according to the data contained in (effect) labels and definitions.

### 3. Results and Discussion

As illustrated in Table 1, based on the criteria mentioned above, namely inclusion, semantic relevance and offensiveness, the analysis has shown three sets of expressions: 1) 66% of Wikipedia's ethnic slurs are recorded in GED, of which 2) 80% are semantically relevant because they present an ethnicity-related sense, of which 3) 80% are ethnophaulisms according to the dictionary.

Table 1. Quantitative data

Inclusion		Semantic relevance		Offensiveness	
Included	188/285	Relevant	151/188	Ethnophaulisms	121/151
				Not offensive	30/151
		Irrelevant	37/188		
Not included	97/285				

Regarding the first set, for example, the dictionary does not record *armo*, “a racial epithet” for “a white person of Armenian descent” (Dalton, 2007, p. 139), *Aunt Jemima*, a noun referring to “a black woman who seeks approval from white people by obsequious behaviour” (Dalzell & Victor, 2013, p. 62), *Buddhahead*, an offensive noun for “a Japanese person” (Dalzell, 2018, p. 100), *cabbage-eater*, an offensive epithet for “a German or Russian immigrant” (Dalzell & Victor, 2013, p. 372), *cheese-eating surrender monkeys*, a phrase meaning “the French” (Dalzell, 2018, p. 139), *dot head* for “an Indian or Pakistani” (Dalzell, 2018, p. 238), *Eyetalian* for “an Italian” (Dalzell, 2018, p. 267), *jungle bunny*, a highly offensive noun for “a black person” (Dalzell, 2018, p. 457), *Leb*, *Lebo*, or *Lebbo*, used derogatorily in Australian English to refer to “a Lebanese person, or any person from an Arabic background” (Dalzell & Victor, 2013, p. 1375), *Mister Charlie* for “a white man” (Dalzell, 2018, p. 524), *nig nog*, used in British English to denote “any non-white person” (Dalzell & Victor, 2013, p. 1580), *Portagee* for “a person from Portugal” (Dalzell, 2018, p. 614), *round eye* for “an American or European” (Dalzell, 2018, p. 662), *sand nigger*, a highly offensive phrase for “an Arab; an Indian or Pakistani person” (Dalzell, 2018, p. 671), *taco*, an offensive name for “a Mexican or Mexican-American” (Dalzell, 2018, p. 777), *timber nigger* and *wagon burner*, both representing offensive epithets for “a Native American Indian” (Dalzell, 2018, pp. 795, 830).

As concerns the second set, among the 37 terms which do not present any ethnicity-related senses, there are several common nouns, mostly of polysemous nature, with no relevant semantic extensions recorded. Examples include the following lemmas, which target the group between brackets in Wikipedia's List of ethnic slurs (Wikipedia, 2024, online), some of which only in specific varieties of English, as highlighted in italics: *ape* (US black people), *apple* (NAm native Americans), *banana* (NAm Asian people), *coconut* (US, UK, NZ Hispanics, or Latinos), *pancake* (Asian people), *snowflake* (US white people), and *teapot* (black people). Interesting expressions in this group are *eight ball*, *goombah*, *kebab*, and *kimchi*, whose meaning might motivate the use of these lemmas as stereotypical allusions to ethnic groups.

In the dictionary, EIGHT BALL (black people) refers to “the black ball, numbered eight, in the game of eight ball”. GOOMBAH (Italian people) is not a US derogatory name for an Italian American, as defined by Dalzell (2018, p. 350), but an informal North American noun denoting “an associate or accomplice, especially a senior member of a criminal gang”, whose origin dates back to the 1960s and probably comes from “a dialect alteration of Italian *compare* ‘godfather, friend, accomplice’” (original emphasis). KEBAB (Muslims, usually of Arab or Turkish descent) is “a dish of pieces of meat, fish, or vegetables roasted or grilled on a skewer or spit”. Similarly, KIMCHI (Korean people) denotes “a Korean dish of spicy pickled cabbage”.

Lastly, in the third set, the 30 semantically relevant lemmas which cannot be interpreted as ethnophaulisms according to the dictionary include, for instance, ABC, AUSSIE, FRITZ, GYPSY, INDON, KIWI, PAK and ROSBIF. Table 2 illustrates the pertinent data contained in the entries for these lemmas in GED.

Table 2. Examples of ethnicity-related but not offensive lemmas

	Labels	Definitions
ABC <sup>3</sup>	informal, US	an American-born person of Chinese descent.
AUSSIE	informal	a person from Australia.
FRTIZ	dated, informal	a German, especially a soldier in the First World War (often used as a nickname). <ul style="list-style-type: none"> <li>the Germans collectively.</li> </ul>
GYPSY	-	1. a member of a people originating in South Asia and traditionally having an itinerant way of life, living widely dispersed across Europe and North and South America and speaking a language (Romani) that is related to Hindi; a Romani person.
INDON	informal, Australian	a person from Indonesia.
KIWI	informal	2. a New Zealander.
PAK	-	(in South Asian use) Pakistan.
ROSBIF	informal, humorous	(originally among French-speakers) an English person.

As data in Table 2 show, the senses of these lemmas are associated with ethnicity or nationality, but there is no sign concerning their derogatory or offensive use. Worth noting, however, are some usage-related pieces of information contained in the definitions, like “often used as a nickname” for FRITZ, and “originally among French-speakers” for ROSBIF. As concerns PAK, it must be highlighted that this noun does not refer to a person from Pakistan, but it denotes the country in the dictionary.

### 3.1 Ethnophaulisms in Google’s English Dictionary

After describing the general findings of the analysis above, this section focuses on the treatment of the 121 ethnophaulisms found in the dictionary. As Figure 1 illustrates, the main markers of offensiveness are effect labels (L), mostly alone or in combination with definitions (LD), thus appearing in a total of 118 entries (97%). In addition, to a very limited extent, definitions alone (D) can also contain pertinent information.

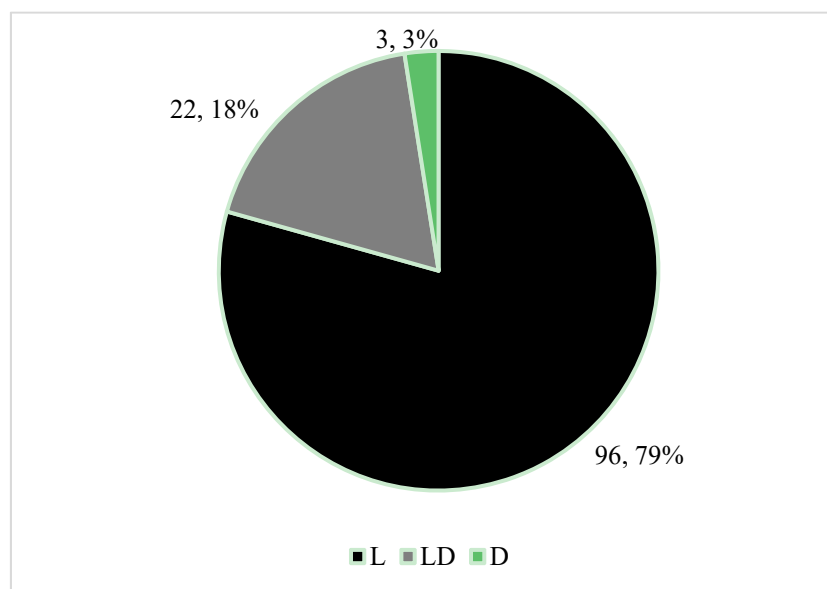


Figure 1. Markers of offensiveness in Google’s English dictionary

Highlighted in bold small capitals and placed above or at the beginning of the sense they describe, depending on whether the lemma is monosemic or polysemic, effect labels characterise the use of the large majority of ethnophaulisms in the dictionary, comprising 64% of lemmas labelled as *offensive* (76/118) and 36% of lemmas labelled as *derogatory* (42/118). Interestingly enough, as to the metalanguage used in their treatment, it is important to stress that in GED these labels always follow a clear warning symbol, which is a triangle with an exclamation point, and that in 25 instances (21% of the total) the relevant description is preceded by a clickable “View definition” wording, meaning that these offensive or derogatory senses are visible only if users click on them, an option which

turns into “Hide definition” after the latter has been displayed. An entry presenting this feature is, for example, the offensive US SLOPE in sense 2, that is “a person from East Asia, especially Vietnam”. Despite the small number of lemmas involved, this is a very distinctive and meaningful metalexigraphic characteristic which motivates the title of this paper.

Going back to the two effect labels, they can both occur with other usage labels (temporal, stylistic and geographical labels), but, while *offensive* is not further specified, the *derogatory* effect of a term is also described in terms of frequency: this label also appears as *often derogatory* (8 occurrences) and as *sometimes derogatory* (2 occurrences), which, however, represent limited instances out of the total. For example, POMMY and SEPO are informal, often derogatory names for “a British person” and “an American person” respectively, while GORA is a “sometimes derogatory” noun referring to “a white person”.

Going into details, among the lemmas labelled as *offensive* there are: ABO and BOONG, both used in Australian English to mean “an Aboriginal person”, BEANER, a North American expression for “a Mexican or person of Mexican descent”, BOGTROTTER for “an Irish person”, COOLIE for “a person from South or East Asia”, DAGO for “a Spanish, Portuguese, or Italian-speaking person”, GINZO and GREASER, both US English terms for “an Italian person” and “a Hispanic American, especially a Mexican” respectively, GYPPO used in British English for “a Romani person; a gypsy”, GIN and LUBRA, both Australian English epithets for “an Aboriginal woman”, ORIENTAL, a name for “a person of East Asian descent”, the US English SPIC for “a Spanish-speaking person from Central or South America or the Caribbean, especially a Mexican”, WOP for “an Italian or other southern European”, and ZIPPERHEAD for “a person from SE Asia” in US English. Labelled as *derogatory*, and mainly *informal*, are, among others, ARGIE for “a person from Argentina; an Argentinian”, EYETIE for “an Italian person”, FROG for “a French person”, GRINGO for “a person, especially an American, who is not Hispanic or Latino”, JOCK for “a Scotsman”, KRAUT for “a German”, LIMEY for “a British person”, PADDY and PAT for “an Irishman”, the North American POLACK for “a person from Poland or of Polish descent”, RUSSKI for “a Russian”, TAFFY for “a Welshman”, WETBACK for “a Mexican living in the US, especially without official authorisation”, and YANK for “an American”.

As illustrated in Figure 1, in GED the second entry section containing relevant data is the definition, mostly combined with effect labels (18%) or, more rarely, alone (2,5%). In the latter case, the description more or less explicitly suggests the offensive effect of lemmas. Examples include GALLA, defined as “another term for Oromo” which is “not favoured by the Oromo themselves and is regarded as pejorative or offensive”, and PALEFACE, which refers to “a white person (formerly used in stereotyped representations of the speech of North American Indians)”. When combined with effect labels, more importantly, two strategies in the phrasing of definitions can be observed. The first and most frequently adopted one contains the following elements in 19 entries out of 22 in total (86%): A/an + effect adjective + term for + a/an + ethnic/nationality adjective + person. Effect adjectives in this pattern include *contemptuous* (13 instances), *derogatory* (5), and *insulting* (1). For example, defined as *contemptuous* terms are CHINK “for a Chinese person”, JAP and NIP “for a Japanese person”. Defined as *derogatory* terms are OFAY “for a white person, used by some black people” and SAMBO “for a black person”. The second strategy observed involves the description of usage data in brackets, like in the entry for RAGHEAD and TOWELHEAD, both meaning “a person who wears a turban or keffiyeh (often used as a term of abuse for an Arab or Muslim)”.

### 3.2 Comparative Analysis

This section discusses the findings of this study in comparison with those of Pettini’s paper (2023a), which, as already mentioned, investigates the treatment of ethnophaulisms in the free online edition of the Oxford Dictionary of English (ODEO) hosted on “powered by Oxford” Lexico.com (see Note 2).

Based on the criteria used in the analyses, as to inclusion, out of a total of 285 ethnic slurs collected from Wikipedia, the number of terms recorded in the two reference tools differs and GED includes a smaller number in absolute terms, namely 188 (66%) as opposed to the 227 lemmas (80%) recorded in the ODEO. As to semantic relevance, 179 and 151 are the ethnicity-related lemmas found in the two dictionaries, and, relative to the size of those included, this result is very similar (79% and 80% respectively). As to the third analytic criterion, which is offensiveness, findings are also comparable, in relation to the numbers of semantically relevant lemmas recorded in the two works: ethnophaulisms are 146 in the ODEO and 121 in GED, accounting for 82% and 80% of their respective ethnicity-related entries.

Overall, out of the total number of Wikipedia’s ethnic slurs explored, in the ODEO “the majority are not only included (80%) and with the relevant sense (63%), but more than half of them (51%) are treated as ethnophaulisms” (Pettini, 2023a, p. 309). Similarly, in GED, still out of the total, despite lower percentages, the majority are included (66%), they mostly present ethnicity-related senses (53%), but less than half of them are ethnophaulisms (42,5%).

In more detail, treated as ethnophaulisms in the ODEO and not in GED are 28 lemmas, encompassing, among

others, as defined in the former, AUNT JEMIMA, an informal, derogatory, US name for “A black woman, (in later use) specifically one considered submissive or servile towards white people”, CHEESE-EATING SURRENDER MONKEYS, a derogatory US phrase to mean “The French”, DOTHEAD, a slang, derogatory and offensive noun used in US English to refer to “A person of South Asian origin or descent”, the informal, offensive US EYETALIAN for “An Italian”, the derogatory FRENCHER and FROG-EATER for “A French person”, the derogatory MISS ANN which “in African-American usage” represents “(a name for) a white woman, especially one who is considered hostile to or patronizing of black people”, the derogatory Canadian NEECHEE for “A North American Indian”, the slang, derogatory and offensive US RASTUS, which is “a nickname for a hypothetically average or typical African-American man. Originally applied jestingly in a number of songs and films, and later with more deliberately offensive intent”, and UNCLE TOMAHAWK, a derogatory and offensive North American expression meaning “A North American Indian who is considered to be excessively obedient to or cooperative with the white establishment”.

On closer inspection, regarding the treatment of ethnophaulisms in the two dictionaries, important similarities and differences can be observed. First, concerning similarities, effect labels are the first and most frequent type of information related to offensiveness users find in both tools, the same labels are used in both (*offensive* and *derogatory*) and the frequency of use of the label *offensive* is higher in both. However, in the ODEO the two labels are also used together as illustrated above for DOTHEAD, RASTUS, and UNCLE TOMAHAWK. Moreover, the other relevant entry sections are more numerous in the ODEO. Indeed, if labels alone characterise the use of 74% and 79% of ethnophaulisms in the ODEO and GED respectively, they combine with definitions in 10% of entries in the ODEO (as opposed to 18% in GED), and also with usage notes (5%), with definitions and usage notes (1%) and with word origin (1%). As to sections other than labels, definitions alone contain relevant data in 5% and 3% of entries of the ODEO and GED respectively, while usage notes are used to signal offensiveness in the ODEO only (4%). Given the focus on usage, the absence of this section in GED is noteworthy. In this sense, for example, an expression which is treated as an ethnophaulism in the ODEO and not in GED is GYPSY. As Pettini (2023a, p. 312) explains, the ODEO’s usage note in this entry states that “the word Gypsy is now sometimes considered derogatory or offensive and has been replaced in many official contexts by Romani or Roma”, although it is still “the most widely used term for members of this community among English speakers”.

Regarding definitions, compared to the findings shown in Pettini (2023a, p. 312), the following similarities and differences emerge. When this entry section is the only marker of offensiveness, relevant usage data are provided in brackets in GED too. When definitions combine with effect labels, the two dictionaries adopt exactly the same pattern in the phrasing of descriptions, the effect adjectives found are the same, among which the most frequently used one is *contemptuous* in both, but in GED it co-occurs with the *derogatory* label too, and not only with *offensive* as in the ODEO.

Lastly, although examples of use were not included in the analysis, for comparative purposes it is worth mentioning that out of a total of 121 ethnophaulisms in GED, almost all the entries for these lemmas (109, 90%) do not contain any illustrative examples. This result is significant because it confirms an aspect that emerged in Pettini’s work too (2023a, p. 314), where 80% of ethnophaulisms were not exemplified in the ODEO and, thus, it might signal an omission strategy affecting sensitive content which is worthy of further investigation.

#### 4. Conclusions

Google is the most visited website in the world and given the tendency of online users to look for lexical information via general search engines, special attention should be paid to the analysis of the data they are presented with by the dictionary service of this technology giant. The quantity and quality of lexicographic information online users find when dealing with their language issues matter, and this is particularly important when it comes to semantic areas linked to sociocultural aspects. This is all the more so where language expresses intolerance and discrimination towards diversity, when words are used as weapons to attack individuals or social groups on the basis of who they are, of identity factors like ethnicity, nationality, colour, and descent. Although linguistic racism and xenophobia represent an ancient and cross-linguistic phenomenon, the alarming increase in cases of hate speech targeting people on these grounds in the present cultural moment is a red flag to be alerted to and all-out efforts should be made to counter this global issue. In this sense, dictionaries can play a role too and can contribute to reflecting the hateful power and politically incorrect status of ethnophaulisms in order to offer users linguistic guidance, as Cloete claims (2014).

The explanatory notes about offensive language discussed in Section 2.1 seem to suggest that Oxford Languages and Google are conscious of the issues which may arise from changing sociocultural sensitivities and that they try to forestall criticism with informative commentaries on the descriptive nature of GED. On Google’s support



webpage (Google Search Help, n.d., online), users learn that offensive words are included because they are part of the language, they are recorded for more complete results, and that Google's partner clearly labels their status to provide proper context. Indeed, since this global platform licenses data, it is specified that "dictionary results don't reflect the opinions of Google", a statement which might be interpreted as a way to prevent public debate on their content, to avoid the risk of facing backlash from users, as happened in the past when OUP was the target of a campaign against the representation of women in the free and pervasive "powered by Oxford" dictionary content (see Pettini, 2021, 2023b).

The results of the analysis suggest a twofold interpretation process. In absolute terms, out of the total of 285 lexemes scrutinised, findings show that the majority are included (66%), they mostly present ethnicity-related senses (53%), but less than half of them are treated as ethnophaulisms (42.5%), as discussed in Section 3. This means that if an online user googles those ethnic slurs, they have a 40% chance of finding them. In relative terms, on the contrary, out of the total of the semantically relevant lemmas included in GED, ethnophaulisms represent the large majority (80%). More specifically, relative to their quantity in GED, what online users always find when they google these hate words is, together with unambiguous data, a visible warning symbol, which is a triangle with an exclamation point, a sign urging caution in the use of these expressions. In this respect, the major dictionary markers indicating offensiveness are effect labels, which characterise the use of almost all ethnophaulisms (97%). More rarely, pertinent usage data are found in definitions (3%). Relative to their size, thus, ethnophaulisms in Google's English dictionary are clearly labelled and/or described as offensive or derogatory words, thus making online users aware of their hurtful nature.

With respect to the research questions this study aimed to answer, thus, results show that on the most popular website in the world general users of the Internet can learn that ethnophaulisms are instances of hate speech, data openly indicate their abusive nature and outrageous power, and the dictionary's distinctive "view and hide definitions" option found in some entries (21% of the total) might be understood as an attempt to treat ethnic slurs in a very special way. By using this mechanism, lexicographers seem to suggest the idea that even though racist language exists, and thus must be recorded in a dictionary, these extreme forms of linguistic intolerance are so dangerous that they must be concealed, in what seems to be a descriptive but slightly censoring approach.

Concerning future research, this study will be extended to other "powered by Oxford Languages" platforms like Apple and Microsoft operating systems, whose preinstalled dictionaries represent another free and extremely widespread reference tool. Moreover, further research will be carried out to explore the features of Google's English dictionary more in depth, and special attention will be paid to the treatment of other manifestations of hate speech, in semantic areas related, for example, to gender, sexual orientation and (dis)ability.

#### **Acknowledgements**

Not applicable.

#### **Authors' contributions**

Not applicable.

#### **Funding**

Not applicable.

#### **Competing interests**

The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### **Informed consent**

Obtained.

#### **Ethics approval**

The Publication Ethics Committee of the Canadian Center of Science and Education.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

#### **Provenance and peer review**

Not commissioned; externally double-blind peer reviewed.

#### **Data availability statement**

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## Data sharing statement

No additional data are available.

## References

- Allan, K., & Burridge, K. (2006). *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511617881>
- Béjoint, H. (2016). Dictionaries for General Users: History and Development; Current Issues. In P. Durkin (Ed.), *The Oxford Handbook of Lexicography* (pp. 7–24). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199691630.013.2>
- Bianchi, C. (2022). *Hate speech: Il lato oscuro del linguaggio*. Bari: Laterza.
- Cloete, A. E. (2014). The Treatment of Sensitive Items in Dictionaries. In R. Gouws, U. Heid, W. Schweickard & H. E. Wiegand (Eds.), *Dictionaries. An International Encyclopedia of Lexicography Supplementary* (Volume: Recent Developments with Focus on Electronic and Computational Lexicography, pp. 482–486). Berlin: De Gruyter. <https://doi.org/10.1515/9783110238136.482>
- Dalton, C. H. (2007). *A Practical Guide to Racism*. Sheridan, WI: Gotham Books.
- Dalzell, T. (Ed.). (2018). *The Routledge Dictionary of Modern American Slang and Unconventional English*. London: Routledge. <https://doi.org/10.4324/9781315195827>
- Dalzell, T., & Victor T. (Eds.). (2013). *The New Partridge Dictionary of Slang and Unconventional English*. London: Routledge. <https://doi.org/10.4324/9781136283055>
- Faloppa, F. (2020a). *L'hate speech, questo sconosciuto*. Lingua Italiana Treccani. Retrieved from [https://www.treccani.it/magazine/lingua\\_italiana/speciali/Hate\\_speech/01\\_Faloppa.html](https://www.treccani.it/magazine/lingua_italiana/speciali/Hate_speech/01_Faloppa.html)
- Faloppa, F. (2020b). *Manuale di resistenza alla violenza delle parole*. Torino: UTET.
- Ferrett, E., & Dollinger, S. (2021). Is Digital Always Better? Comparing Two English Print Dictionaries with Their Current Digital Counterparts. *International Journal of Lexicography*, 34(1), 66–91. <https://doi.org/10.1093/ijl/eca016>
- Filmer, D. (2011). *Translating Racial Slurs: A Comparative Analysis of Gran Torino Assessing Transfer of Offensive Language between English and Italian*. MA thesis. Durham, UK: Durham University.
- Filmer, D. (2012). *The 'gook' goes 'gay': Cultural Interference in Translating Offensive Language*. InTRAlinea 14. Retrieved from <http://www.intralinea.org/archive/article/1829>
- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering Online Hate Speech*. Paris: UNESCO Publications.
- Google Search Help. (n.d.). *Learn about dictionary boxes on Google*. Retrieved from <https://support.google.com/websearch/answer/10106608?hl=en>
- Green, J. (2004). Language: Wash Your Mouth Out. *Critical Quarterly*, 46(1), 107–111. <https://doi.org/10.1111/j.0011-1562.2004.00557.x>
- Henderson, A. (2003). What's in a Slur? *American Speech*, 78(1), 52–74. <https://doi.org/10.1215/00031283-78-1-52>
- Hughes, G. (2006). *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-Speaking World*. Armonk: M.E. Sharpe.
- Hughes, G. (2010). *Political Correctness: A History of Semantics and Culture*. Oxford: Wiley Blackwell. <https://doi.org/10.1002/9781444314960>
- Jackson, H. (2013). *Lexicography: An Introduction*. London: Routledge. <https://doi.org/10.4324/9780203467282>
- Jackson, H. (2017). English Lexicography in the Internet Era. In P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography* (pp. 540–553). London: Routledge. <https://doi.org/10.4324/9781315104942-34>
- Lew, R. (2011). Online Dictionaries of English. In P. A. Fuertes-Olivera & H. Bergenholtz (Eds.), *E-Lxicography: The Internet, Digital Initiatives and Lexicography* (pp. 230–250). London: Continuum.
- Lew, R., & de Schryver, G. M. (2014). Dictionary Users in the Digital Revolution. *International Journal of Lexicography*, 27(4), 341–359. <https://doi.org/10.1093/ijl/ecu011>
- Lorentzen, H., & Theilgaard, L. (2012). Online Dictionaries—How Do Users Find Them and What Do They Do

- Once They Have? In R. Vatvedt Fjeld & J. M. Torjusen (Eds.), *Proceedings of the XV EURALEX International Congress* (pp. 654–660). Oslo, Norway: University of Oslo. Retrieved from <https://euralex.org/publications/online-dictionaries-how-do-users-find-them-and-what-do-they-do-once-they-have/>
- Müller-Spitzer, C., & Kopleinig, A. (2014). Online Dictionaries: Expectations and Demands. In C. Müller-Spitzer (Ed.), *Using Online Dictionaries*, Lexicographica Series Maior 145 (pp. 143–188). Berlin: De Gruyter. <https://doi.org/10.1515/9783110341287>
- Nissinen, S. (2015). *Insulting Nationality Words in Some British and American Dictionaries and in the BNC*. MA Thesis. Tampere, Finland: University of Tampere.
- Nuessel, F. (2008). A Note on Ethnophaulisms and Hate Speech. *NAMES*, 56(1), 29–31. <http://doi.org/10.1179/175622708X282929>
- Oxford English Dictionary online. (2024). *Ethnophaulism*. Retrieved from <https://www.oed.com/search/dictionary/?scope=Entries&q=ethnophaulism>
- Oxford Languages. (n.d.). *Oxford Languages and Google*. Retrieved from <https://languages.oup.com/google-dictionary-en>
- Palmore, E. B. (1962). Ethnophaulisms and Ethnocentrism. *American Journal of Sociology*, 67(4), 442–445. <https://doi.org/10.1086/223168>
- Pettini, S. (2021). “One is a woman, so that’s encouraging too”: The representation of social gender in “powered by Oxford” online lexicography. *Lingue e Linguaggi*, 44, 275–295. <http://doi.org/10.1285/i22390359v44p275>
- Pettini, S. (2023a). The Language of Ethnic Conflict in English Online Lexicography: Ethnophaulisms in “powered by Oxford” Lexico.com. *Lexikos*, 33, 299–317. <https://doi.org/10.5788/33-1-1816>
- Pettini, S. (2023b). The ‘Social’ Dimension of Online Lexicography: Gender, Dictionaries and Users. In D. Faraci, G. Iamartino, L. Lopriore, M. Nied-Curcio & S. Zanotti (Eds.), *When I use a word, it means just what I choose it to mean—neither more nor less: Studies in honour of Stefania Nuccorini* (pp. 189–203). Roma: Roma Tre Press.
- Pinnavaia, L. (2020). Tracing Political Correctness in Bilingual English-Italian Dictionaries. *Textus*, 33(1), 87–106. <https://doi.org/10.7370/97353>
- Roback, A. A. (1944). *A Dictionary of International Slurs* (Ethnophaulisms). Cambridge, MA: Sci-Art Publishers.
- Sellers, A. (2016). Defining Hate Speech. *Berkman Klein Center Research Publication*, 20, 16–48. <https://doi.org/10.2139/ssrn.2882244>
- Semrush. (2024). *Top Websites*. Retrieved from <https://www.semrush.com/website/top>
- Statcounter. (2024). *Search Engine Market Share Worldwide: March 2024*. Retrieved from <https://gs.statcounter.com/search-engine-market-share>
- UN (United Nations). (2019). *UN Strategy and Plan of Action on Hate Speech*. Retrieved from [https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action\\_plan\\_on\\_hate\\_speech\\_EN.pdf](https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf)
- Wachal, R. (2002). Taboo or Not Taboo: That is the Question. *American Speech*, 77(2), 195–206. <https://doi.org/10.1215/00031283-77-2-195>
- Wikipedia. (2024). *List of ethnic slurs*. Retrieved from [http://en.wikipedia.org/wiki/List\\_of\\_ethnic\\_slurs](http://en.wikipedia.org/wiki/List_of_ethnic_slurs)
- Zgusta, L. (1998/2000). Some Developments in Lexicography, Past and Present. In J. E. Mogensen, V. Hjørnager Pedersen & A. Zettersten (Eds.), *Symposium on Lexicography IX* (pp. 11–26). Halle: Max Niemer Verlag. <https://doi.org/10.1515/9783110915044.11>
- Žugić, D., & Vuković-Stamatović, M. (2021). Problems in Defining Ethnicity Terms in Dictionaries. *Lexikos*, 31, 177–194. <https://doi.org/10.5788/31-1-1635>

## Notes

Note 1. As of March 2024, Google accounts for more than 90% of the search engine market share worldwide (Statcounter, 2024) and it is the most visited website in the world (Semrush, 2024).

Note 2. As Pettini explains (2023a, p. 300), Lexico.com was OUP’s domain of the free online version of the Oxford

Dictionary of English (ODEO) from June 2019 to August 26, 2022, when the website was closed and redirected to Dictionary.com, the original website operator. However, the ODEO was not moved to this website, whose main, proprietary source is the *Random House Unabridged Dictionary*.

Note 3. Other entry sections, if included at all, have not been examined because they do not present pertinent data (audio pronunciation, phonetic transcription, grammatical information, illustrative examples, usage notes, word origin, and phrases).

### **Copyrights**

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).