

Language and Ethics: A Better Way to Treat "Derogatory Terms" and Corresponding Punishments

Yanru Jiang*

Department of Philosophy, University of California, Santa Barbara, CA, USA

*Corresponding author: yanrujiang@ucsb.edu

Abstract. Language and words are mostly considered to have consistent semantics, so many scholars classify phrases that generally appear to be associated with harm or undesirable consequences as "derogatory terms" and strictly prohibit their usage in order to establish a more ethical social environment. However, the question of whether the origin of the harm comes exclusively from a word and whether the word itself has a negative nature still needs to be discussed further. This paper will analyze and criticize traditional theories and conclude that there is no constant negative nature to a single word based on the research of historical changes and examples of semantic plurality in different geographical areas. It is only necessary to make people aware that language and semantics are dependent on the intentions of the speaker and the rational analysis of the listener, and that too much ethical criticism will only make the communication more unethical or counterproductive.

Keywords: Semantics; derogatory terms; intentions.

1. Introduction

The philosophy of language is a branch of philosophy that deals with the nature of language and meaning as well as their relation to reality. Clearly, there are many theoretical debates in the process of development in this field, such as whether the significance of words is merely a reference to the real world or whether it can be taken further to make judgments and inferences about reality through linguistic logic or whether the meaning of words just lies in communication through social practices. None of these ideas, however, deny that language has a powerful function that may either demonstrate the truth or affect social behaviors. It is commonly accepted that language has the ability to promote understanding and respect among people in a positive, pleasant manner while also having the ability to harm or offend others through violent speech, slurs, or other types of misleading language used in negative ways.

In many cases, conflicts over language can be highly severe and traumatic. For example, in a donut store, a 77-year-old man was accidentally killed after making racist comments to the clerk. It is a story that begins with a single word and results in the death of one individual, which actually leads to the tragedy and sorrow of two families [1]. More catastrophic is the annihilation of an entire race, such as the genocide of Tutsi in Rwanda in 1994, which was stimulated by the increased open usage of anti-Tutsi pejorative words like "inyenzi" or "cockroach." Consequently, the attack or murder of Tutsi became even more legitimate in their community due to the widespread speech act [2]. In this sense, the use of derogatory terms under these circumstances engenders significant detrimental actions for both speakers and audiences, which further raises ethical issues that need to be explored.

It is widely acknowledged that there is a moral obligation to avoid causing harm to others through using disparaging language, and it also appears reasonable to penalize people when they mention these words no matter what reason they hold, even without the intention to hurt them. Generally, people believe that language itself can be used as a weapon to harm others, so the nature of aggression is assigned and defined by those "derogatory terms". Because of this, evil people are able to grasp the opportunity to unite and adopt such a weapon to violate others, and virtuous people are convinced that all those who use this term are malicious and hurtful in their hearts. They criticize the immorality of these speech acts, especially for the deeply hurt listeners or referred party, and may provoke their subsequent anger and resistance.

However, is it justifiable that language can be endowed with such a potential for harm inherently? Moreover, should the harsh punishment that follows for the speaker be entirely ethically sanctioned? It is true that deliberate harm should not be excused, and those, especially for a group, who use words to insult others brutally need to be identified and publicly shamed. However, in many cases, the speaker does not intend to harass others or does not realize that he or she is causing harm. As a result, the same word should not always be assigned the same derogatory meaning. Only by understanding that any language is merely a morally natural tool for expressing intentions, rather than being given an intrinsically unchanging definition, can harm be truly reduced, and ethical conflicts be ethically minimized. Furthermore, this proper way of treating and understanding language will also lead to a more friendly social environment with less anger or meaningless punishment for interlocutors.

2. Language and Harm in Traditional Views

When scholars tend to analyze derogatory terms, they typically adopt a semantic approach to define the term, which targets specific classes of people, such as the specific category of racial or religious groups, a gender, or sexual orientation, regardless of the context of the conversation. The reason people choose semantic theories of those slurs is that they can easily explain why the same words or language can have so much power to remain consistent in doing harm in different contexts; therefore, they conclude that derogatory terms only entail the literal meaning and themselves have the function and features to demean the others. As a result, for example, when the n-word is uttered in the context, the black people are extremely sensitive to this way of speaking and believe it is being used to attack them.

Historically, the most prevalent approach to defining derogatory terms was either a purely descriptive one or a purely expressive one, and the essential differences between the two made them incompatible with each other [3]. To be specific, a purely descriptive way implies a completely objective representation or description of the meaning of a word, including the object and the characteristics it is used to describe. For example, the stigmatized n-word is used to describe African Americans, but this approach cannot explain why only the n-word is a derogatory term since it has the same denotation as the term "African American," i.e., the same descriptive meaning. Therefore, another group of scholars proposed an expressive way to analyze derogatory terms. They argue that only this approach can emphasize the reasons why language has the power to hurt others and undermine the peaceful social environment, all of which are caused by the person's heightened emotional state of mind. However, this view ignores the denotative aspect of words. For example, it leads to the question of why a person in a terrible mood uses the N-word against whites and blacks at the same time yet only makes black people angry.

In response to this phenomenon, other researchers have proposed a third interpretation of literal meaning, the combination of descriptive and expressive perspectives. They claim that slurs or derogatory terms differ in nature from other languages, and it is possible to make the two approaches compatible and interrelated with each other. Adam Croom, for example, argues that when people choose to use a derogatory term, they at first endorse the negative attitudes to the term and then use words and attitudes to refer to the undesirable characteristics shared by a particular group of people [4]. Adam's innovation is to suggest that derogatory terms are not necessarily limited to a specific and fixed group but can also be used to describe other people with the same undesirable attributes. This approach does explain derogatory terms more comprehensively than the purely descriptive and pure expressive approaches, but it still confines the term to its own negativity and cannot serve all cases. For example, when I just mention the N-word in my article to give an example of my own opinion, but only neutrally used, without attributing any negative attitude or describing any bad characteristics. Then it is clear that the N-word in my article, which does not contain any pejorative references, cannot be applied to Adam's meaning theory.

In addition to the study of literal semantics, there is another category of scholars who focus more on the impact of the usage of derogatory terms, which is an extension of the study when the negative

nature is predetermined. According to Tirrell, there are five major roles deeply derogatory terms playing in speech acts. They first mark the in and out groups, implying that the specific term can always divide people into groups, and the utterers take it for granted to target the word with a derogatory meaning to other groups of people who are different from him. Second, the term is negatively essentializing in terms of identifying people and delivering information, and this feature is the one that addresses the negative nature of the words. Thirdly, the use of the word is a socially ingrained activity, and its power of it stems from connecting with social networks of oppression or discrimination, which can also be explained by the predetermined negativity of words. Fourth, the term can have diverse functions under different language games, and the last feature is that these profoundly derogatory terms can be action-engendered, and ultimately it leads to the inevitable tragedy [2].

Clearly, Tirrell is a typical scholar who summarizes and analyzes the derogatory terms in one single conclusion, listing the negative effects or functions that these terms can have. Although she mentions that deep derogatory terms can have a variety of functions and must be used in the context of social activities, these characteristics are always limited to the word's harmful nature. Once a thing is given an essence, even for a single word, it signifies that such a property is constant and unchangeable, like the negativity of derogatory words. For example, if a knife is regarded in advance as hazardous, whether it is used to harm others or to save oneself, whether or not the knife is utilized and causes actual injury in social activities, it can only be a harmful object that should be discarded and eventually strictly forbidden to be used according to ethical requirements. Similarly, if the derogatory term is initially perceived negatively, then the only way to cope with it is to repulse the usage and utterly eliminate it.

3. Response to Traditional Views

Based on the preceding study of other academics' research findings, the first doubt that arises is whether the so-called derogatory term can truly have the essence of negative attributes. Adam Croom introduces the mixture of descriptive and expressive meaning as destructive descriptions and emotions from the words themselves, while Tirrell Lynne outlines the ensuing effects and features of their uses, given the deeply derogatory meaning as an inherent precondition. If one follows their argument that a category of words or language is inherently harmful, then the main challenge is that once this nature is stated, then all references to the same words must have offensive meanings consistently. Obviously, the idea that language itself can reveal the nature of things is problematic as long as there are examples that do not meet this condition.

Consider the most prevalent and widely examined N-word as an example. The evolution of this word demonstrates that the so-called essence, both in the descriptive and expressive ways of analysis, is changing throughout time; that is, there is no eternally fixed negative essence. Initially, the N-word in Latin is a neutral term that does not contain any derogatory or discriminatory meaning but simply stands for the black color. Gradually, however, the word became more than a mere symbol or reference when an enslaved race in history was associated with a number of negative characteristics. Since then, the N-word has referred to black people as possessing a horrible nature, which implied the degradation of the entire race with injury, blood, and pain. It took a long time after the abolition of slavery for the people to accept the concept of racial equality. Although the term was still considered sensitive since it was used to express racism, non-blacks using it politely in their conversation without endorsing any negative connotation was acceptable. However, that was not sufficiently satisfactory for many people to achieve their goal of racial equality. They begin to argue that in order to achieve greater progress in society today, it is important to entirely renounce all usage or discussion of the N-word. More specifically, even citing others or referring to the word without any aggressive or malicious purpose, such as for academic linguistic research, should be blamed or even strictly punished.

A longitudinal look at the N-word from a historical perspective reveals that its meaning and usage have constantly been shifting and have largely evolved in response to people's changing attitudes and social circumstances. From the initial reference to blackness, it evolved into a deeply derogatory term that included a significant racist insult directed at the referred African Americans by the speaker. Although after the achievement of equality, the degree of derogatory meaning of the same word was reduced, debates still existed, as the previously mentioned scholars have suggested that since derogatory terms have been used offensively, there should be no tolerance toward the term in the future as well. In addition to studying a word longitudinally across history, cross-sectional research is also feasible. For example, it is worthwhile to investigate whether the N-word means the same thing in different parts of the country. According to Michael Patrick MacDonald's record, the same word denotes class hierarchy but has nothing to do with race; to be specific, it is a term used to designate white people rather than black people as the conventional meaning [5]. Although it has a relatively negative connotation, it is far from the genuinely pejorative phrase Tirrell advocated for, and there is no need to abolish such a description entirely.

It can be concluded that if none of the previous usages possess a constant and unchanging meaning or essence, then there is no reason to determine a permanent meaning in the present and reject the potential possibility of multiple semantic changes in the future. By studying history, "the eroded aspects of meaning are retrieved" [6], and people can find that a derogatory term is just a negative label artificially put on a morally neutral word, and as long as the label is withdrawn, the word itself should not have any risk of harm. Specifically, it is just a tool for people to communicate, and all the essence and meaning are artificially constructed, so there is no ultimate truth to language itself, including its inherent harm. However, does this nullify all of the unfortunate realities brought about by racial slurs and appease those who are racist with malicious intent?

In fact, the generation of harm is not denied; the only difference is that the source is not the same as the analysis from traditional views. Terms are assigned derogatory meanings not because the words themselves have such meanings; instead, it's people's subjective consciousness and thinking that manipulate the words. As Ludwig Wittgenstein argues, language is not about "propositions" that express or reveal abstract entities, but rather these words and sentences are like game pieces to use in social practices [7]. Different people have different strategies for the game, and the expression of language is the choice of the speaker's own strategies or intentions. To complete such a game, speakers have their own option to use the N-word either as a medium to express discrimination against African Americans or to express other meanings without endorsing any negative attitudes or racial insults, such as conducting academic research.

Furthermore, the specific generation of harm can be better understood in the context of H.P. Grice's meaning theory. According to Grice, meaning is based on the speaker's intention to make the listener aware of his articulated intention, and only after that can the listener successfully understand or accept the speaker's meaning [8]. Similarly, derogatory terms like the N-word have offensive meanings only because the speaker deliberately tends to make the listener or the target aware of his discriminatory intention, and afterward, the victims are hurt, and their anger is triggered. Therefore, the cunning use of language is to present one's evil purpose, and make the audience finally recognize such a purpose is the complete mechanism of harm. Clearly, simply defining a word as a derogatory term is insufficient to explain its negative impact.

Only by not viewing language to be intrinsic in nature can the speaker's intentions be clarified, allowing the listener to differentiate between intentional and unintentional harm. To be precise, listeners will no longer be sensitive to a single word, and they will be more rational in discerning the speaker's intention and whether there is any harm in it or not. Such rationality will also play a vital role in better controlling irritable emotions. Even if the audience recognizes the speaker's malice, they will be more rational about the insult and not develop further rage to satisfy the villain. After all, verbal attacks are the most cost-effective way for evil people to adopt and attack, and if the listener can recognize this, they will select a more appropriate and harmless way to respond.

4. Punishment and Problems

Aside from questioning whether derogatory terms are intrinsically negative, it is critical to consider how to handle and punish the corresponding serious harms as indicated in traditional views. If the harm caused by derogatory terms is inevitable, then only the complete elimination of these words seems to be morally right. However, that leads to a second concern. Is it possible to make a term disappear forever, and are the punitive measures put in place to restrict its use successfully in practice?

To begin with, even if the prohibition is fully implemented, this will not reduce the frequency with which the N-word is used or mentioned from a psychological perspective. The eradication of such a phrase exacerbates the potential danger or conflict in society. According to the White Bear Effect in psychology, also known as the rebound effect, suppressing thoughts often increases their occurrence. Specifically, when people are urged to restrain their thoughts from thinking of a white bear, they may succeed in suppressing the emergence of this object at the moment. However, after a while, they are more likely to have the thought of the white bear than those who were not asked to suppress it before. Likewise, when people are requested to refrain from saying a word, they are consciously suppressing their thoughts, which in turn increases the probability that they will say the word in the future. And when a controversial term like the N-word is misused and widely spread, the cost to society is enormous.

Secondly, from an ethical standpoint, the elimination of words and severe punishment leads to an unexpected and terrible effect that harms innocent people. For example, by providing various examples of famous people who were blamed or lost their jobs for mentioning the N-word in their show or on social media, John McWhorter states how the unreasonable judgment in these cases hurt the innocent speakers and, at the same time, presupposed the fragility of black people. As he believed, sometimes the N-word is mentioned unconsciously and non-judgmentally as referring to the word itself without any further implication at all, and these situations may be inevitable in the quote, in the classroom, or in relative historical study. If the audience consciously insists on identifying such mentions as slurs as well, they are engaging in a critical act based on the assumption that blacks are particularly vulnerable and cognitively impaired in distinguishing between the use and mention of a word. This presupposition and overprotection can be considered another devaluation of black people. [9] However, even if people carefully design the rules to differentiate the treatments among different situations, such as using the word as a slur and simply mentioning the word itself, the problems persist. Actually, the distinction can never set a perfect standard because the use of language in the language game is complicated and varied, and there is no fixed rule that can be set up to deal with all circumstances fairly.

Thirdly, retributive punishment based on the employment of these ritual slurs fails to educate and make the guilty aware of his wrongdoing to others while also failing to compensate the victims. The most common and widespread means of punishment is mass shaming online, which is an accusation and criticism of the speaker's speech act for mentioning derogatory terms in the past. According to Kane's analysis, this critique often originates from the idea of individuals imposing their own values onto others and attempting to interfere with their behaviors [10]. Moreover, such motivation is not helpful in solving moral issues or treating the presence of harm in a rational and meaningful way. The actual effect of hasty mass shaming cannot make the guilty realize and reflect on the wrongness of their actions because it does not provide them the time and space to empathize with the victims, who are the people they have insulted with words, especially for celebrities who rapidly apologize after saying something wrong. Instead of considering others, they only strive to redeem or rebuild their own self-image. Therefore, no matter what standpoint we choose to examine the outcome of punishment, it is almost impossible to eradicate words completely and ethically.

5. Conclusion

Past linguistic studies have concentrated on exploring the nature of language, and certain words have been labeled as "derogatory terms" in their traditional views. These studies have generally

agreed that language has power and can be hurtful in nature which is why it frequently leads to extremely horrible disasters. In order to avoid harm, these derogatory terms need to be strictly regulated and banned from use. However, the first and most serious problem with the traditional viewpoint is that language cannot be considered to have the essence of negativity as they believed. The source of tragedies, including discrimination, death, and even war, is not from the presentation of language or a single word, but the speaker's subjective intention to hurt or the listener's subjective fragility. Examining the historical use of words is the greatest technique to shatter the misunderstanding about language in this scenario. When it is recognized that a word has different meanings in numerous dimensions throughout history, people will learn that there is no consistent meaning or usage of words, which will better protect both listeners and speakers.

Furthermore, from both the psychological and ethical points of view, these derogatory terms will not disappear, nor will punitive measures be effective in reducing their frequency of appearance or the extent of the harm they cause. The harm will not occur unless the term's inherent definition is broken. For the malicious speaker, they will realize that a term is not harmful in nature, and there will be no need to use it as a weapon to attack. For the unintentional speaker, when they make a friendly and polite reference to these terms, it will not harm the listener as well because they understand that no discriminatory intent is present in such communication.

Overall, while this study cites multiple relevant theories as support, there are limitations to the specific cases investigated. The conclusions seem to be drawn from the analysis of only one derogatory term, the "n-word," and lack more examples of whether there are other contemporary uses of this term that are completely unrelated to negativity. In addition, the study does not provide actual strategies to respond to deliberate insults or modify their damaging repercussions, except for researching historical usage of the word. Therefore, future studies should address the preceding shortcomings and provide more viable and ethical methods of reclaiming terms with derogatory meanings and minimizing harm.

References

- [1] Cokley, K. The psychological impact of racist slurs, the painful history and consequences of the weaponization of the n-word. *Psychology Today*, 2021. Retrieved January 9, 2023, from <https://www.psychologytoday.com/us/blog/black-psychology-matters/202105/the-psychological-impact-racist-slurs>
- [2] Lynne, T. Genocidal Language Games. In Ishani Maitra & Mary Kate McGowan (eds.), *Speech and Harm: Controversies Over Free Speech*. Oxford University Press, 2012: 174--221.
- [3] Potts, C. *The expressive dimension*, 2007: 165-198.
- [4] Croom, Adam M. "Slurs." *Language sciences (Oxford)* 33.3, 2011: 343–358.
- [5] Coles, G. The Exorcism of Language: Reclaimed Derogatory Terms and Their Limits. *College English*, 2016,78(5): 424–446.
- [6] Zwart, Hub. "Care for Language: Etymology as a Continental Argument in Bioethics." *Journal of bioethical inquiry* 18.4, 2021: 645–654.
- [7] Lycan, W.G. *Philosophy of Language: A Contemporary Introduction* (3rd ed.), 2018. Routledge. <https://doi.org/10.4324/9781315146119>
- [8] Grice, H. P. Meaning. *The Philosophical Review*, 1957, 66(3): 377–388. <https://doi.org/10.2307/2182440>
- [9] Mcwhorter, J. The new n-word standard isn't progress. *The New York Times*, 2022. Retrieved January 10, 2023, from <https://www.nytimes.com/2022/02/11/opinion/use-mention.html>
- [10] Kane, Laura Wildemann. Accountability and Community on the Internet: A Plea for Restorative Justice. *Journal of Applied Philosophy*, 2020, 37 (4): 594-611.