



Theses and Dissertations

2021-12-08

Trademarks and Genericide: A Corpus and Experimental Approach to Understanding the Semantic Status of Trademarks

Richard B. Bevan

Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Arts and Humanities Commons](#)

BYU ScholarsArchive Citation

Bevan, Richard B., "Trademarks and Genericide: A Corpus and Experimental Approach to Understanding the Semantic Status of Trademarks" (2021). *Theses and Dissertations*. 9332.

<https://scholarsarchive.byu.edu/etd/9332>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Trademarks and Genericide: A Corpus and Experimental Approach
to Understanding the Semantic Status of Trademarks

Richard B. Bevan

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Arts

Earl K. Brown, Chair
William Egginton
Wendy Baker Smemoe
Jim Law

Department of Linguistics
Brigham Young University

Copyright © 2021 Richard B. Bevan
All Rights Reserved

ABSTRACT

Trademarks and Genericide: A Corpus and Experimental Approach to Understanding the Semantic Status of Trademarks

Richard B. Bevan
Department of Linguistics, BYU
Master of Arts

Genericide is the process by which a trademarked term is used generically by the public and ultimately loses its legal trademark protections. The linguistic methods that courts have used to determine whether a given term is in the process of or has undergone genericide have historically relied on dictionaries. However, there has been a recent push to use corpus linguistics as a tool to aid in that determination for not only trademarks but word meaning in general (Hoopes, 2019; Lee & Mouritsen, 2018). In addition to corpus data, I argue that the use of experimental data via a linguistic questionnaire can support, validate, and clarify corpus findings and can be an additional means to aid in the determination of the semantic status of trademarked terms. Corpora comprised of texts from the social media website Reddit were created and concordance lines exhibiting uses of 24 terms (10 generic and 14 trademarked) were judged based on their semantic senses as interpreted by two raters. These concordance lines were compared to the responses of a linguistic questionnaire asking participants how they used those 24 terms. Results show that the questionnaire responses are comparable to and validate many of the results of the judging of the Reddit corpora. The questionnaire data provided clarity on use of terms deemed ambiguous by previous research. I assert that the use of questionnaire data is a useful option in researching the genericide phenomenon either in conjunction with corpus data or independently. Both methods are considered helpful for courtrooms and businesses in investigating genericide, but based on the findings of this thesis I advocate that neither method can determine genericide alone but should be only considered as aids to work in conjunction with other evidence and data.

Keywords: genericide, trademark, corpus, questionnaire

TABLE OF CONTENTS

TITLE PAGE	i
ABSTRACT	ii
TABLE OF CONTENTS.....	iii
1 INTRODUCTION	1
2 REVIEW OF LITERATURE	5
2.1 Genericide and Semantic Change	5
2.2 Notable Trademark Court Cases	8
2.3 Linguistic Tools Used for Contested Meaning	11
2.4 Corpus Studies for Legal Interpretation.....	15
2.5 Experimental Data for Increased Reliability.....	17
2.6 Surveys in a Semantic Framework	18
3 METHODOLOGY	19
3.1 Word Selection.....	19
3.2 Reddit Corpus	20
3.3 Survey of Terms.....	25
4 RESULTS AND DISCUSSION	30
4.1 Rater Agreement	30
4.2 Crock-Pot, Kleenex, Band-Aid, and Xerox	32
4.2.1 Survey Results	38
4.3 Ten Generic Terms	40
4.3.1 Survey Results	43
4.4 Ten Trademarked Terms.....	46

5 CONCLUSION.....	54
APPENDICES	61
REFERENCES	83

1 Introduction

Trademarks are among a company's most valuable assets (Hoopes, 2019) and it is critical for companies to protect their trademarks if they want to retain their value. A trademark's value is comprised of intangible goodwill between consumers and companies. For example, the trademarks for Apple, Google, and Microsoft were estimated at \$170 billion, \$101.8 billion, and \$87 billion, respectively, in 2017 (Hoopes, 2019). This amount of value trademarks bring to companies makes unsurprising how much money and time is spent by companies to protect their trademarked terms.

However, many brand names that were once protected as trademarks have undergone semantic broadening through a process known in the legal community as "genericide". Genericide is "the evolution of a trademark's meaning from a single source of products to a word for the product itself" (Lalonde et al., 2007). In other words, genericide occurs when a word that is trademark protected has broadened semantically to the point where the word is no longer identifying the source of the product but is a word for the product itself regardless of the company or brand that produces it. Previously trademarked terms that have undergone genericide include *Murphy Bed, aspirin, heroin, escalator, thermos, yo-yo*, and many more (Hoopes, 2019; Hughes, 2018; In, 2002). Companies, such as Google, have been involved in legal battles and have retained their trademarked status and protections (Hughes, 2018). Google's legal battles have focused on the term *google* being used as a verb meaning 'to search on the Internet'.

There have been several court cases that have attempted to determine the semantic status of trademarks and whether they have truly become generic or whether they are specific enough to retain trademark protections. Some of these cases include *Bayer Co. v. United Drug Co.*¹

¹ 272 F. 505 (S.D.N.Y. 1921)

(regarding the product *aspirin*), *Kellogg Co. v. National Biscuit Co.*² (regarding the product *shredded wheat*), and, more recently, *Elliott v. Google Inc.*³ (regarding the word *google*). These cases will be discussed more fully in the following chapter. Lawyers and judges often refer to the Lanham Act of 1946⁴ as a guide in these cases. The Lanham Act is the primary federal trademark statute of law in the United States. It prohibits trademark infringement, trademark dilution, false advertising, and much more. However, this thesis focuses on its prescription of the primary significance test as a measure to determine the primary semantic status of a trademarked product name in the minds of the consuming public. Regarding primary significance, the Lanham Act states:

The primary significance of the registered mark to the relevant public rather than the purchaser motivation shall be the test for determining whether the registered mark has become the generic name of goods or service on or in connection with which it has been used.⁵

This has become known as the Primary Significance Test. This “test” requires challenging parties to “show that, in the minds of the consuming public, the ‘primary significance’ of the mark describes a class of products rather than a particular product made by the trademark holder” (Hoopes, 2019, p. 409).

Courts have used various linguistic tools in determining the primary significance in the minds of the consuming public. Some have used dictionaries and media usage such as in *Murphy Door Bed Co. v. Interior Sleep Sys., Inc.*⁶ when determining the trademark status of *Murphy Bed*

² 305 U.S. 111 (1938)

³ 860 F.3d 1151 (9th Cir. 2017)

⁴ 15 U.S.C. § 1051 et seq.

⁵ *Id.* at 1064(3)

⁶ 874 F.2d 95, 101 (2d Cir. 1989)

or in *Elliott v. Google Inc.* for the term *google*. Hoopes (2019) states that corpora can be used as an additional aid in determining meaning, but that they cannot prove genericide. This is because there are many uses of trademarked terms that must be considered ambiguous due to the inability to know for certain how the user of a particular word intends that word to be perceived.

It is important to note that the classification of terms as being trademarked or generic, and even the process of genericide, are legal terms rather than linguistic. However, this thesis uses such legal terms as though they were linguistic in nature for ease of understanding. This thesis in no way purports to be legal in nature or hold legal bearing. The goal of this thesis is to expand the steps taken by Hoopes (2019) in using corpus data to explore the genericide phenomenon through a linguistic lens. As dictionaries and corpora alone have not proven flawless tools in determining genericide, the importance of linguistic evidence and data is undeniable. I show in this thesis how experimental data, specifically those from a linguistic questionnaire, can validate and substantiate the effectiveness of corpus data.

This thesis combines texts gathered and compiled into multiple corpora from the online social media platform Reddit, as well as questionnaire responses to answer and explore three questions:

1. How do the combination of corpora and survey data compare to those of Hoopes (2019)?
2. How do corpora data compare and contrast with the experimental survey data for terms that have become legally generic?
3. How do corpora data compare and contrast with the experimental survey data for terms that have retained their trademarked statuses?

To answer and explore these questions, I collected concordance lines of various trademarked terms, both current and former, within the Reddit corpora and categorize them based on whether they are used generically. The categorization of concordance lines is compared to responses to a linguistic questionnaire. I hypothesize that the more common responses from the questionnaire are positively correlated to trademarked terms and are categorized as being generic usages.

2 Review of Literature

This review of literature begins with a look at the factors that fuel the semantic change of trademarks and the genericization process. I will then provide a synopsis of notable court cases of trademark disputes that have shaped the idea of primary significance and the evolution of the Lanham Act. I will also detail recent court cases involving trademark disputes and the implementation of the Primary Significance Test. I then detail the linguistic tools typically used in determining meaning and previous studies that have implemented corpus data in legal interpretation. Finally, I review the necessity of experimental and participant data as an extra measure of reliability and validity to corpus data and their contribution to semantics.

2.1 Genericide and Semantic Change

There has been little research performed in the actual linguistic factors that determine the genericization of a trademark. In fact, the idea of genericization formed outside of linguistics as purely a marketing problem or the “misuse” of the trademarked terms by the consuming public (Clankie, 2000). However, the definition of genericization itself is indicative of semantic change. That is, that genericization is “the evolution of a trademark’s meaning from a single source of products to a word for the product itself” (Lalonde et al., 2007).

Clankie (2000) explores semantic reasons why trademarks become generic. He postulates that there are four primary processes by which genericization of trademarks occurs: 1) Novelty – a brand name for a product that did not exist before will become synonymous with the product itself, such as *rollerblades* 2) Length and Predominance – if the predominant brand name in a semantic class is shorter (that is, shorter in word length) than the corresponding class noun, the predominant brand name will become generic, such as *Jell-O* for *gelatin*; 3) Genericization as a Regular Process – trademarks follow a systematic pattern of semantic change. They begin as

proper adjectives (specific) followed by a common noun. Examples of this use of trademarks are *Apple computers* or *Ford truck*. The trademarked terms *Apple* and *Ford* are “proper adjectives” and precede the product (common noun) associated with the trademark. Then by the process of ellipsis⁷ they become proper nouns (specific) without any succeeding common noun. For example, one might say in a conversation about trucks: *I don’t like Fords. I like Chevys more.* The common noun, *truck(s)*, can be omitted based on the context of the conversation but the use of *Fords* or *Chevys* can still be deemed as trademarked uses. The trademarked terms can then become common nouns (generic). The common nouns can then, in some cases, become verbs or attributive adjectives; 4) The Single Association Hypothesis – There must be a distinct association between a brand name and a single product. For example, a brand name, such as Chanel, that produces a variety of products cannot become generic because there is no single item association to be made.

While Clankie’s (2001) reasons for genericide highlight the processes by which genericide can occur, there are two major mechanisms by which semantic change can occur: metaphor and metonymy (Traugott & Dasher, 2001). Nerlich and Clarke (1992) argue that “the trick of being innovative and at the same time understandable is to use words in a novel way” and that “there are only two main ways of going about that: using words for the near neighbours of the things you mean (metonymy) or using words for the look-alikes (resemblars) of what you mean (metaphor)” (p. 137). It is clear, then, that the mechanism by which genericide, and semantic change in general, occurs is metonymy.

Lakoff and Johnson (1980) explain that metonymy is the use of one entity to stand for another. Trademarks that are used generically exemplify this by using the trademark – an entity

⁷ Omission of a word or words that are superfluous or understood from context

that is a subset of a category of product, to represent the category as a whole (better known as PART FOR THE WHOLE metonymy). Example (1) below shows one general instance of PART FOR THE WHOLE metonymy as provided by Lakoff and Johnson (1980, p. 36).

(1) I've got a new *set of wheels*.

In this example, the *set of wheels*, a part of a car, is representative of the entire car as a whole. The metonymic usage of trademarks is not surprising seeing how metonymy in general is prevalent throughout everyday language use (Denroche, 2014; Kövecses & Radden, 1998; Lakoff & Johnson, 1980; Radden, 2005).

The connection between a *set of wheels* being used as PART FOR THE WHOLE metonymy and trademarks doing the same is a broad interpretation of this kind of metonymy. Seto (1999) argues that an example such as (1) would be considered a true part-whole relationship. That is to say that *wheels* are actual components that comprise the car as a whole. Conversely, *aspirin* within the broader category of *pain killers* does not have the same relationship. *Aspirin* is not a foundational component that is found in every pain killer like wheels to cars, and so it is not explicitly a part of a whole. The relationship between *aspirin* and *pain killers* may be more taxonomic in nature and a better example of hypernymy and hyponymy. One key difference between the two is that *aspirin*, despite its subordinate position within the category of *pain killers*, is a pain killer itself, whereas *wheels* are not cars in and of themselves. This distinction between example (1) and generic trademark usage is valid, but this thesis considers the broader interpretation that each represents a hierarchical relationship between a superordinate item (*car* or *pain killer*) and a subordinate item (*wheels* or *aspirin*).

One of the key elements to metonymy is the selection of a metonymic vehicle (i.e. which term is to be used metonymically). This selection is based on the relative salience of a feature

associated with the word. For example, in the early 20th century when films went from silent to including sound and one wanted to express the idea of this new kind of film, the term *talkie* places salience to the speech in a film (Denroche, 2014). Kövesces and Radden (1998) propose that salience can be placed on more common members of a category and that they will be used metonymically over less common members, or simply COMMON OVER LESS COMMON. They provide *aspirin* as an example of this salient feature which suggests that because *aspirin* became the most common brand of its type of product at the time it ultimately became generic.

2.2 Notable Trademark Court Cases

One of the earliest and influential cases of genericide is the term *aspirin* (Bellifemine, 1984). Aspirin, the technical name being acetyl salicylic acid, was the trademark of Bayer Co. starting in 1899. In 1921, Bayer Co. sued United Drug Co. for trademark infringement. The drug was initially distributed strictly by prescription and was labelled as *aspirin*. Bayer Co. eventually began selling the drug directly to consumers but failed to include the company's name anywhere on the bottle. Bayer Co. eventually added its company name two years before the patent on the product expired and sued United Drug Co. when they released their own product of acetyl salicylic acid under the name *Aspirin*. Ultimately, the court decided that there was no trademark infringement occurring. It concluded that *aspirin* had become generic, that it was used to describe any pain-relieving medication, and that it had passed into public domain.

This case did not rely on linguistic tools or methods to determine the semantic status of the term *aspirin*. However, it did set the groundwork for the idea of primary significance. This is evident in the deciding statement by Judge Learned Hand⁸:

⁸ Judge Billings Learned Hand was a judge of the United States District Court of the Southern District of New York at the time of the Bayer Co. case. He eventually retired as the Senior Judge of the United States Court of Appeals for the Second Circuit.

“The single question, as I view it, in all these cases, is merely one of fact: What do buyers understand by the word whose use the parties are contending? If they understand by it only the kind of goods sold, then, I take it, it makes no difference whatever what efforts the plaintiff has made to get them to understand more. He has failed...”⁹

Linguistic or semantic evidence is not the only factor to play a role in determining genericide. Business marketing and decisions, as well as the extent to which a company takes measures to protect its trademark, are additional factors that can play a role. However, Judge Learned Hand makes clear that the perception and understanding of consumers of trademarked terms is the primary question to be considered.

The statement by Judge Learned Hand may have set the foundation for the idea of primary significance, but the term itself came to prominence in 1938 in the case of *Kellogg Co. v. National Biscuit Co.* (Coverdale, 1984). At the heart of this case was the term *shredded wheat*. National Biscuit Co., the plaintiff, sued Kellogg Co. for using the term as a descriptive term for a newly produced cereal. The Supreme Court ruled that the term was descriptive and that it did not qualify to be trademarked. Most importantly, the Court stated that in order “to establish a trade name in the term *shredded wheat* the plaintiff must show more than a subordinate meaning which applies to it. It must show that the primary significance of the term in the minds of the consuming public is not the product but the producer.”¹⁰ This is the first occurrence of the phrase “primary significance” and the Supreme Court emphasized that the population in question was the consuming public.

⁹ *Supra* note 1 at 509

¹⁰ *Supra* note 2 at 118

As a result of the two cases discussed above and as other litigation involving trademarks became increasingly prevalent, The Lanham Act was passed in 1946 as the primary federal trademark statute of law in the United States. The original language of the act stated that “the primary significance of the registered mark to the relevant public rather than purchaser motivation shall be the test...”. The Lanham Act was later amended in 1984 with the Trademark Clarification Act. This amendment removed the phrase “rather than purchaser motivation” and enforced the idea of primary significance as the central test in determining whether a trademarked term has truly become generic or not (In, 2002).

The primary significance test does not always provide a clear answer to the semantic status of trademarks. *Elliott v. Google, Inc.* was a dispute in which the trademarked term, *google*, was argued to be a generic term primarily due to the fact that the term is often used as a verb meaning ‘to search something online’ (Hughes, 2018). Despite this association the U.S. Court of Appeals for the Ninth Circuit ruled in the case of Google that genericide had not occurred with the term. They stated that “(1) the claim of genericide did not relate to a good or service and (2) Google’s verb usage does not automatically constitute a generic use” (Hughes, 2018, p. 271). Both rulings have had and may yet have a profound effect on current and future trademarks. Hughes (2018) states that a company such as Xerox benefits greatly from this ruling as their trademark, *xerox*, has been in danger of genericide due to the secondary meaning of the term, ‘to photocopy’. Companies other than Xerox and Google have their trademarks used as verbs, as well. *Photoshop*, *Uber*, and *Zoom* are modern examples that are used in such a way. Examples (2), (3), and (4) below exhibit instances of each of these terms used as verbs as seen from the Reddit data collected in this thesis.

- (2) It looks like you are on really bad terms if you **photoshop** her out of pictures.

(3) ...like we will drive and they are like no we will Uber.

(4) ...about using our sick days!!!11!! We could just zoom instead.

This thesis focuses on both arguments found within the decision of the Ninth Circuit when analyzing the terms from the corpus and survey data. That is that 1) the claims of genericide must relate to a good or service and 2) verb usage does not automatically constitute a generic use (Hughes, 2018).

2.3 Linguistic Tools Used for Contested Meaning

Courts have frequently relied on linguistic tools to aid in cases that require the determination of the meaning of a term or phrase. This is the case not only in trademark disputes, but in legal questions of statutory interpretation and ordinary meaning, as well. One of the most common linguistic tools used to determine meaning are dictionaries (Hoopes, 2019; Mouritsen, 2010). Hoopes (2019) is critical of the use of dictionaries and argues that while lexicographers may use objective data in their compilation of meanings, the creation of dictionaries is a subjective and human endeavor (Hoopes, 2019, p. 417). This sentiment is also felt by other scholars who criticize the idea that the dictionary is a linguistic bible (Goldfarb, 2017; Lee & Mouritsen, 2018; Mouritsen, 2010). This criticism from scholars is relatively recent and has made a push in the last decade or so. However, in 1945 Judge Learned Hand, who was involved in *Bayer Co. v. United Drug Co.*, said of dictionaries: “[I]t is one of the surest indexes of a mature and developed jurisprudence not to make a fortress of the dictionary.”¹¹

Trademark cases are not the only legal contexts in which the use of dictionaries have been criticized. Court cases in which dictionaries have been used in determining statutory interpretation and ordinary meaning of the law have been criticized, as well (Goldfarb, 2017; Lee

¹¹ Cabell v. Markham, 148 F.2d 737, 739 (2d Cir. 1945)

& Mouritsen, 2018; Mouritsen, 2010; Solan, 2001). One Supreme Court case that reflects this criticism is *Muscarello v. United States*¹². The United States Supreme Court had to determine whether the meaning of *carry a firearm* included having a gun in a car and, more specifically, in a glovebox. Justice Breyer's majority opinion acknowledges that the term *carry* can mean either transport in a vehicle or to have on one's person (Lee & Mouritsen, 2018). Justice Breyer explains that the former is the primary definition because it is the first definition in the Oxford English Dictionary, and the latter is a special definition (Lee & Mouritsen, 2018). The dissenting opinion by Justice Ginsburg disagreed with this notion and asserted that just because a definition was considered "special" does not reduce its value. These uncertainties on what exactly was meant by ordinary meaning and how to best interpret the meanings led to a 5-4 split among Justices. Due to the split opinions in the *Muscarello* case, the case has been said to show "a need for improvement in judicial reasoning about statutory concepts" (Solan, 2001).

One issue with dictionaries is simply that they treat words with multiple senses differently. Some dictionaries might include a sense that another does not. Some might lump similar senses into a single, broader sense whereas others will keep them separate (Goldfarb 2017). Fillmore and Atkins (1994) compared senses of ten dictionaries for the verb *risk*. They first compared three dictionaries of similar size and purpose (single-volume dictionaries for native speakers of English): 1) Collins English Dictionary (1986), 2) Longman Dictionary of the English Language (1988), and 3) Chambers 20th Century Dictionary (1983). Each dictionary showed two senses for the verb. However, after dissecting the wording of the two senses from each dictionary, Fillmore and Atkins showed that at least three, not two, senses could reasonably be deduced. This was not unique to just the verb *risk*, but to other common English words, such

¹² 524 U.S. 125, 1998

as *jeopardy*, *danger*, *peril*, *threaten*, and *warn* (Fillmore & Atkins, 1994). The incongruency between dictionaries has the potential for dictionaries to be selected simply because their available definitions match better with a certain argument.

Mouritsen (2010) argues that the belief that the first listed definition of a word is the most frequent and most common definition is something that needs to be avoided due to its problematic use in the *Muscarello* case. This is what is known as the Sense-Ranking Fallacy and it most likely comes from the human presumption that the most important things ought to be first in a sequence (Mouritsen, 2010). Mouritsen provides detailed examples of how the different dictionaries, namely the Oxford English Dictionary and the Random House Dictionary of the English Language, used by the Supreme Court in the *Muscarello* case prove that sense ranking is fallacious.

Each of these issues about the role of dictionaries in determining ordinary meaning have led to the push by many researchers to use corpus linguistics as a quantifiable and objective means to determine ordinary meaning to supplement, not necessarily replace, the use of dictionaries (Hoopes, 2019). Previous literature provides excellent research methods and models in using corpora in cases of contested meaning (Cunningham & Egbert, 2019; Lee & Mouritsen, 2018; Phillips et al., 2016; Phillips & White, 2018). Central to these models is the use of multiple corpora such as the Corpus of Contemporary American English¹³, or COCA, (Davies, 2008-), the Corpus of Historical American English¹⁴, or COHA, (Davies, 2010-) and the BYU Law Corpora¹⁵.

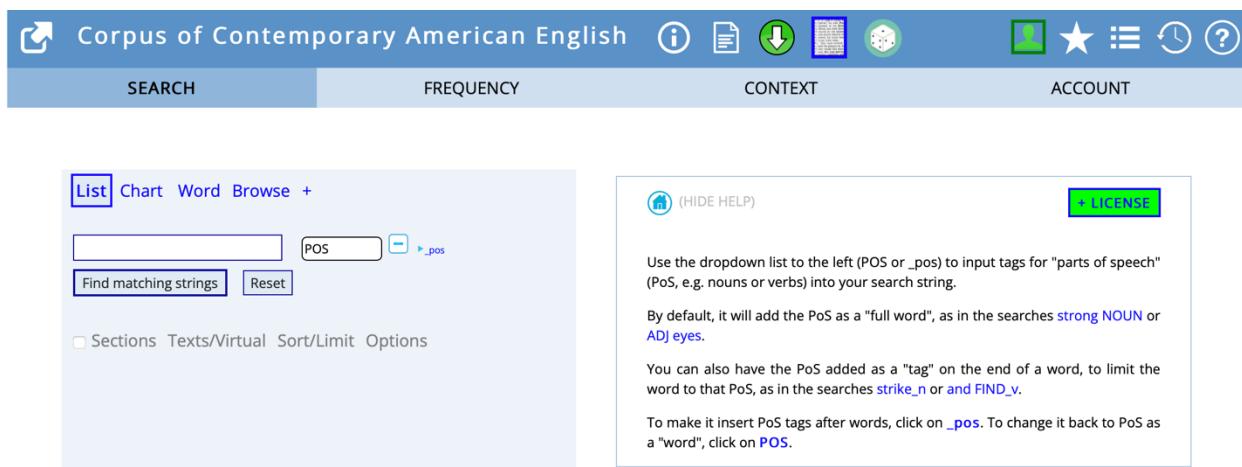
¹³ Available at <https://www.english-corpora.org/coca/>

¹⁴ Available at <https://www.english-corpora.org/coha/>

¹⁵ Available at <https://lawcorpus.byu.edu/>

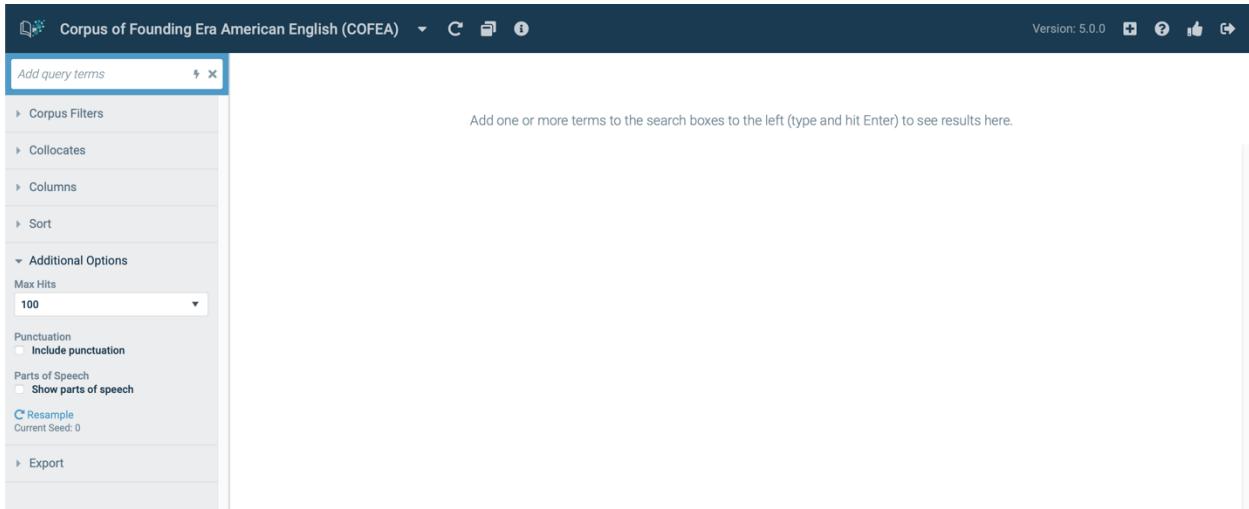
COCA has been used most extensively in legal linguistic research (Hoopes, 2019; Lee & Mouritsen, 2018; Mouritsen, 2010) but COHA (Phillips et al., 2016), and the Corpus of Founding Era American English, or COFEA, (Cunningham & Egbert, 2019; Phillips & White, 2018) have also been used. Each of these corpora provide valuable and convenient interfaces for linguistic research. Each interface can query their respective texts for individual words and phrases, lemmas, parts of speech, collocates, mass retrieval of concordance lines, and trends over time. Figures 2.1 and 2.2 below show the interfaces for COCA and COFEA. COCA and COHA are near identical and so COHA is not represented.

Figure 2.1 COCA Interface



The screenshot shows the COCA interface with a blue header bar. The header includes the logo and name 'Corpus of Contemporary American English', a search icon, a frequency icon, a context icon, an account icon, and a help icon. Below the header is a navigation bar with tabs: 'SEARCH' (selected), 'FREQUENCY', 'CONTEXT', and 'ACCOUNT'. The main search area on the left has a dropdown menu ('List', 'Chart', 'Word', 'Browse', '+') and a search input field with a 'POS' dropdown and a 'pos' link. Buttons for 'Find matching strings' and 'Reset' are present. Below the input field are links for 'Sections', 'Texts/Virtual', 'Sort/Limit', and 'Options'. To the right of the search area is a help panel with a 'LICENSE' button. The help panel contains text explaining the use of PoS tags in search strings, the default behavior of adding PoS as a full word, and how to add PoS as a tag. It also explains how to change PoS back to a word.

Figure 2.2 COFEA Interface



2.4 Corpus Studies for Legal Interpretation

Linguistic research regarding genericide is scarce. Most research regarding genericide and primary significance is legal in nature and does not attempt to show what linguistic methods are best used to aid in determining genericide. Hoopes (2019) appears to be one of the first to begin to question and look at the methods employed to garner meaning for trademarks.

As an alternative aid in determining or recognizing genericide, Hoopes suggests the use of corpora to show “how the public actually uses language” (Hoopes, 2019, p. 421). Hoopes queried four trademarks within COCA to examine instances of generic use. The trademarks he queried were (i) Xerox®, (ii) Crock-Pot®, (iii) Band-Aid®, and (iv) Kleenex®. He collected a random sample of 100 concordance lines for each trademark and coded them as either a generic, trademark, or ambiguous sense. He also looked at collocates of each trademark to show patterns of surrounding words. The results show that unmistakably generic usages for the trademarks were scarce and incontrovertible evidence of genericide was not possible.

Hoopes found many instances of ambiguous uses of each of the four trademarks. These are the main cause for the conclusion that corpora cannot determine or prove genericide. Gilquin and

Gries (2009) suggest that corpus researchers (and linguistic research in general) should consider combining corpus analysis with other experimental methods in order to make more accurate and concrete conclusions. Gilquin and Gries (2009) provide four reasons to do so:

1. “...even the smallest results <in corpora> will often be significant and additional experimental evidence will help separate the wheat from the chaff.”
2. “Different corpora will yield different results and additional experimental evidence will help obtain a more precise understanding of phenomena.”
3. “Corpus-based results can, and should, be validated against corpus-external findings.”
4. “Combining corpus and experimental data would also help gain insight into the relation between the two types of data.”

Primary significance is difficult to prove (Bellifemine, 1984; In, 2002) and has been criticized as being purely lexically based and not focused on the effect of the term on the market (Coverdale, 1984), that is to say that primary significance is more than just what a given trademark means, but that business-related factors such as marketing and other economical factors can play a role in how the consuming public perceives a trademark. Primary significance is difficult to prove because of the difficulty of completely understanding the minds of the consuming public especially in how they interpret meaning.

Conversely, Hoopes (2019) argues that despite the difficulty of using linguistic data to prove primary significance in the minds of consumers, the primary significance test ought to remain the standard to determine genericide as opposed to simply relying on the majority usage of a term. Hoopes states that “[c]ourts should not rely on majority usage because usage does not necessarily track knowledge” (2019, p. 438). That is, that raw frequency data does not necessarily prove or disprove the intent, knowledge, or subjectivity behind the use of the term.

2.5 Experimental Data for Increased Reliability

Perhaps the two most important points given by Gilquin and Gries (2009) as reason to combine corpus and experimental data are numbers 2 and 3 as outlined above: additional experimental evidence will help obtain a more precise understanding of the phenomena and corpus-based results ought to be validated by external results. As has already been discussed, the application of the Lanham Act and Primary Significance Test is not a perfect science. Determining whether a trademark has become generic can be based on several factors that make the phenomenon of genericide very difficult to adjudicate precisely.

This combination has been suggested by other researchers, namely that of Lee and Mouritsen (2018). In their discussion of further research of determining ordinary meaning in the law, especially as it relates to their interpretations of the phrases *carry a weapon* and *vehicle in the park*, they suggest that a survey be constructed that is “aimed at assessing not just the first sort of vehicle that comes to mind but also the range of meanings encompassed within a prohibition on vehicles in the park. Survey data could give us quantitative information about these notions of ordinary meaning.”

Such a combination of corpus and survey data has recently been performed by Garceau (2020) in an attempt to clarify the prevailing definitions of *sex* and *gender*. Garceau used concordance lines from the COCA, iWeb, and COHA corpora that provide usages of the two terms. The concordance lines were placed in a survey and distributed to linguistics college students and to individuals not necessarily associated with linguistics via Mechanical Turk. The survey requested that participants read a concordance line and provide their interpretation of what the term (*sex* or *gender*) was referring to. One point of this study is that it was an attempt to increase objectivity in the interpretation of corpus data. Garceau notes that the motivation behind

this was partly because much of the previous literature on legal interpretation of semantic meaning has been performed by one or two individuals (Cunningham & Egbert, 2019; Lee & Mouritsen, 2018; Mouritsen, 2010; Phillips & White, 2018) whether that be interpretation of concordance lines and how they are rated by the researcher(s), or the selection of corpus examples that only support their arguments. This is a valid point of concern and the results of the study show that agreement between participants from Mechanical Turk was above 80% as compared to about 70% in Phillips and White (2018). This thesis acknowledges this issue and the possible subjectivity of interpretation as a result.

2.6 Surveys in a Semantic Framework

The use of surveys to perform semantic fieldwork is a rather new approach. Theoretical syntax and semantics are typically the approaches by which researchers investigate meaning and semantic statuses of words. These approaches are usually performed solely by the author of a paper with possible feedback from colleagues (Gibson et. al, 2011). Non-quantitative studies, such as qualitative analyses and interpretations of corpus data, that rely on the opinions of one or two people are prone to biases and subjectivity. Vander Klok (2014) advocates for semantic survey because the semantics of many units of language are less understood and opaque. Rather than relying on one or two individual judgements, semantic surveys or questionnaires allow for the collection of not only quantitative data, but the judgements and interpretations of a more diverse set of speakers of a language. Vander Klok (2014) provides possible methods when creating semantic surveys: elicitation in an acceptability or judgement task; a semi-forced task; and a Likert scale rating task. This thesis draws from the semi-forced task method as the basis of its semantic survey.

3 Methodology

This thesis analyzes the usage of 24 different words that have either had their trademark protections lost due to genericide, questioned because of generic uses, or protected and upheld in court. Fourteen currently trademarked brands/products and 10 generic brands/products that lost their patents or trademarks due to genericide will be investigated. The data for this thesis come from two sources: a corpus of Reddit texts compiled personally using a Python web scraping script and a linguistic questionnaire distributed on social media and email. Section 3.1 will review the selection of the 24 words to be investigated. Section 3.2 details the creation of the Reddit corpora, the size of the corpora, and its overall reflection of the target population. Section 3.3 explains the survey, participants (demographics and total number), and its overall reflection of the target population.

3.1 Word Selection

The Wikipedia article '*List of generic and genericized trademarks*' provides a list of 20 terms that were "originally legally protected trademarks, but which have subsequently lost legal protection" (2021). These terms will be referred to as 'generic terms' going forward. The Wikipedia article also includes a list of 145 products and brands that are currently trademark protected but have been used as generic terms. These will be referred to as 'trademarked terms' going forward.

Each term in the Wikipedia lists was queried in COCA and the raw frequency counts for each term were collected. The 10 most frequent terms in each list were selected to be included in this thesis, with one exception. This exception is regarding the term *heroin*. It was within the 10 most frequent generic terms but was chosen to be replaced by *flip phone* due to the mature nature of the product and had a higher risk to potential mental or emotional harm to the participants of

the survey. *Flip phone* was chosen as the replacement because it was the next most frequent generic term from the list gathered from Wikipedia. In addition to the 20 terms selected (10 generic products and 10 trademarked products), four additional trademarked products were included in this thesis: *Crock-Pot*, *Kleenex*, *Band-Aid*, and *Xerox*. These four terms fell outside of the top 10 most frequent trademarked terms but were investigated by Hoopes (2019) and so they will be included in this thesis for comparison. This results in a total of 24 terms – 10 generic and 14 trademarked (see Table 3.1).

Table 3.1 – *Lists of Generic and Trademarked Terms*

Generic Terms	Trademarked Terms
Videotape	Google
Aspirin	Coke
Kerosene	Nintendo
Linoleum	Zoom
Escalator	Uber
Teleprompter	Photoshop
Laundromat	PowerPoint
Trampoline	LEGO
Cellophane	Realtor
Flip phone	Styrofoam
	Xerox
	Band-Aid
	Kleenex
	Crock-Pot

3.2 Reddit Corpus

The purpose of creating a corpus of Reddit text is to find actual language that reflects the language use of the consuming public for a given product better than a generalized corpus, such as COCA. In 1921, Judge Learned Hand in *Bayer Co. v. United Drug Co.*, when determining the trademarked status of *aspirin*, emphasized the need to determine how buyers, or consumers,

understand the trademark in question. This was the beginning of the idea of primary significance in genericide cases that were eventually included in the Lanham Act.

Reddit provides a platform for users to have discussions about anything in forums (known as subreddits) that are focused on specific and narrow topics. For instance, there are subreddits focused on the stock market, individual sports teams, TV shows, cooking advice, and much more. These specialized topics provide the ability for targeted corpora to be created based on relevant topics for each term in question. For example, a corpus can be compiled using texts from subreddits focused on food, slow cooking, and cooking when looking at a term such as *crock-pot*. These subreddits include naturally occurring language by individuals who are likely to use, buy, and consume the products or service referred to by the terms in question.

Each term in question (above) was queried on Reddit.com to find the top subreddits associated with the term. Five subreddits were chosen for each term (with some subreddits being repeated across terms) based on relevance to the product (e.g. the *cooking* subreddit for the trademarked term *crock-pot*). If fewer than five or no relevant subreddits were available, the largest and most popular subreddits were used (e.g. AskReddit, todayilearned, etc.). This resulted in 115 unique subreddits to be used to create a corpus for each term in question.

A Python script (see Appendix A) was used to scrape language data from up to 1000 posts from users within each of the 115 subreddits. The language data included the text from the posts themselves and every comment other users submitted to each post. Table 3.2 below details the subreddits used for each term, the number of posts and words comprising each corpus created for each term, and the average word per post. On average there were 26 million words collected from an average of 4,785 posts for each term.

Table 3.2 – *Reddit Corpus Summary per Term*

Term	Subreddits Used	Total Posts	Total Words	Average Word per Post
Aspirin	Health, conspiracy, explainlikeimfive, medicine, science	4,943	55,121,154	11,151
BandAid	WTF, Warts, firstaid, mildlyinfuriating, teenagers	4,710	23,953,114	5,086
Cellophane	PlayingCardsMarket, Random_Acts_Of_Amazon, bettafish, crafts, playingcards	4,998	3,569,026	714
Coke	AskCulinary, SodaStream, ToFizzOrNotToFizz, food, soda	4,989	11,839,434	2,373
CrockPot	Cooking, PlantBasedDiet, PressureCooking, food, slowcooking	4,983	20,264,322	4,067
Escalator	Whatcouldgowrong, deadmalls, escalator, holdmybeer, therewasananattempt	3,933	19,184,275	4,878
FlipPhone	PhonesAreBad, apple, dumbphones, nostalgia, nosurf	4,988	20,776,698	4,165
Google	AskReddit, artificial, internet, programming, technology	4,958	49,399,846	9,964
Kerosene	DaysGone, Outdoors, camping, energy, preppers	4,977	13,048,567	2,622
Kleenex	DoesAnybodyElse, ForeverAlone, depression, firstworldproblems, sad	4,965	14,835,148	2,988
Laundromat	Entrepreneur, Frugal, Laundromats, childfree, shameless	4,030	26,916,499	6,679
LEGO	Showerthoughts, Toys, legodnd, pokemon, todayilearned	4,925	31,073,341	6,309
Linoleum	CleaningTips, DIY, HomeImprovement, InteriorDesign, printmaking	4,980	21,065,058	4,230
Nintendo	GameDeals, Games, consoledeals, gamecollecting, gaming	4,986	37,492,412	7,520
Photoshop	Art, DigitalPainting, graphic_design, picrequests, postprocessing	4,991	9,085,180	1,820

Powerpoint	FellowKids, LifeProTips, consulting, softwaregore, teenagers	4,898	30,619,432	6,251
Realtor	FirstTimeHomeBuyer, RealEstate, homeowners, legaladvice, realestateinvesting	4,981	22,066,035	4,430
Styrofoam	MadeOfStyrofoam, environment, mildlyinfuriating, mildlyinteresting, oddlysatisfying	3,937	24,653,499	6,262
Teleprompter	Filmmakers, PoliticalHumor, news, politics, videography	4,976	49,312,432	9,910
Trampoline	Trampoline, TrampolineTricks, Tricking, Wellthatsucks, woahdude	3,871	15,089,916	3,898
Uber	AskReddit, CreditCards, SelfDrivingCars, business, sanfrancisco	4,966	29,649,603	5,971
Videotape	ObscureMedia, TrueFilm, boxoffice, movies, unpopularopinion	4,957	42,185,840	8,510
Xerox	OfficeDepot, mechmarket, sysadmin, todayilearned, vintagecomputing	4,978	30,323,923	6,092
Zoom	Professors, Teachers, interestingasfuck, teenagers, todayilearned	4,918	39,549,486	8,042

Each term was queried with a Python script within each term's respective subreddits using regular expressions (see Appendix B) to retrieve a concordance line that included up to 50 characters before and after the queried term. All concordance lines that did not have any preceding context or were simply just the terms themselves were filtered out. A random sampling of 100 concordance lines (or all concordance lines if there were fewer than 100 overall lines collected) for each term was compiled for a total of 2271 concordance lines – only *kleenex* (n=25) and *videotape* (n=46) had fewer than 100 lines.

I examined each of the 2,271 concordance lines and marked each one in one of four ways: (i) Specific – if the use of the trademark was source identifying, (ii) Generic – if the use of

the trademark was product, or category, identifying, (iii) Ambiguous – if, for any reason, the referent associated with the trademark is unknown or unclear, and (iv) Other – if the use of the term is not associated in any way with the source of product or category of product. A second coder marked in the same way as the first coder 10 random concordance lines for each term selected from the 2271 concordance lines from the first coder (240 lines; 10.5%).

Below are examples of concordance lines that were coded based on the four categories listed above. Each of these examples were coded the same by both coders.

(5) ...real estate as well as the National Association of Realtors. Giving out the keys to buyers without the...

This example was coded as being a specific usage (as defined above) of the term *realtor*. The term is used within the context of the entity that holds the trademark (the National Association of Realtors) and the trademarked term's connection to the entity is obviously. Conversely, (6) below is an example of *realtor* that was coded as a generic usage.

(6) ...and keep values propped up as best they can. My realtor says he expects a big "boom" once the lockdown...

This example does not show any obvious connection in its surrounding context to the National Association of Realtors. I do not know the reasoning behind the second coder's choice to mark this as generic, but as I read this example my first thought when reading the word *realtor* was not of the trademark or the National Association of Realtors, but of any real estate agent that aides others in buying homes whether or not they belong to the association.

(7) ly" powerpoint ?

This concordance line was marked as ambiguous. The context surrounding the term is not sufficient to determine in any way how *powerpoint* is being used.

(8) ...you can take a picture from really far away and zoom it in. The moon will appear bigger next to the...

This use of *zoom* was marked as other. The context shows that the usage is related to taking pictures with a camera and is unrelated to the video conferencing software Zoom.

A basic agreement percentage, as well as Cohen's kappa (Cohen, 1960), between the two coders was computed to show to what extent the coders agreed. This is by suggestion of McHugh (2012) as Cohen's kappa adjusts the agreement percentage based the possibility of coders agreeing and disagreeing by chance. Providing both metrics shows to what extent coders agreed and to what extent that agreement has been adjusted by Cohen's kappa due to chance.

3.3 Survey of Terms

The second source of data to be investigated in this thesis is from a linguistic questionnaire. While the purpose of the Reddit corpus is to collect naturally occurring language to reflect the consuming public, the questionnaire is intended to elicit language directly from the public regarding which word, or words, they use to refer to the items and products associated with the terms explored in this thesis. By using both corpus and survey language data, this thesis seeks to substantiate the reasons Gilquin and Gries (2009) give to consider the combination of data: (i) to help separate the wheat from the chaff, (ii) to help obtain a more precise understanding of the genericide phenomenon, (iii) to help validate the corpus-based results, (iv) to help gain insight into the relation between the two types of data.

One goal of distributing the linguistic questionnaire is to collect data directly from individuals and thus gain insight into the minds of the consuming public as required in the Primary Significance Test. As members of society, it is assumed that all participants of the survey are consumers and are aware of terms and products in question. Certainly, elicited

language is subjective and can have external, undue influence on the results. However, this is to be counteracted by the more objective and naturally occurring language of the Reddit corpus.

The survey was created on Qualtrics.com and was comprised of 30 questions – 24 questions (one per term in question) eliciting the word, or words, used by the participant to refer to each product and 6 demographic questions: age, gender, country and state (if from the U.S.) of birth, and country and state of residence. Participants were provided an image of the product in question along with a brief text description of the product. Participants were asked to either provide the word, or words, they use to refer to the product shown and describe or select the option indicating that they do not have a specific word for the product. Figure 1 below shows an example of one of the questions posed to participants. The complete survey can be found in Appendix C.

Figure 1 – *Example Survey Question*



Please provide the word, or words, you would use to refer to the cooking appliance used to cook food over a long period of time.

Answers:

- A blank text box for participant input
- I don't have a word for this

Only one question at a time was given to participants and the order of questions was randomized prior to the distribution of the survey to reduce any inadvertent priming or bias. Each

participant was given the opportunity to enter a drawing to win one of four \$25 Amazon gift cards as compensation for their participation. A separate survey was created for the gift card drawing to make survey responses as confidential and anonymous as possible.

Distribution of the link to the survey was done personally and by the committee members of this thesis via social media (Facebook and Reddit) and email via the snowball method. All participants were encouraged to share the survey themselves to find the most amount of participants. There was no restriction on which demographics were requested to complete the survey. However, native to near-native English speakers over 18 were requested to participate.

There was a total of 288 participants who answered at least one question. Only one was under the age of 18 and their responses were not included. There were also some participants whose responses were removed from the data pool completely due to suspected spamming of the questionnaire. There were about 20 participants whose answers were the same including many of the demographic questions. Based on the information provided from Qualtrics.com, the geographical coordinates of these participants were near identical. Due to the online distribution of the questionnaire, it is suspected the URL to the questionnaire was received by a number of individuals spamming the survey with the chance of winning a \$25 Amazon gift card.

Of the 288 participants, there were 150 females, 114 males, and 3 non-binary or 3rd gender individuals. There were 21 participants who chose not to answer. Table 3.3 below shows the number of participants by age and gender. There were 246 participants (85.4%) from 41 states across the United States. Of these 246 participants, 78 were from Utah (31.7%) and 67 were from California (23.3%). Other countries of origin for participants include Canada (6), the United Kingdom (3), and Sweden (2).

Table 3.3 *Age and Gender Demographics of Participants*

Age and Gender	Count
18 - 24	36
Female	21
Male	13
Non-binary or third gender	2
25 - 34	110
Female	55
Male	54
Non-binary or third gender	1
35 - 44	60
Female	34
Male	26
45 - 54	26
Female	19
Male	7
55 - 64	13
Female	7
Male	6
NA	21
NA	21
Older than 65	21
Female	13
Male	8
Under 18	1
Female	1
Grand Total	288

All participant responses were exported from Qualtrics.com as a CSV spreadsheet file to be easily imported into RStudio for data analysis and visualization. All responses that were from individuals under the age of 18 were removed from the data pool. Although those under the age of 18 might still be considered the consuming public, it was decided that they be excluded from the data for legal reasons due to the difficulty of receiving consent from a legal parent or guardian for an anonymous, online survey. All questions that participants completed were considered valid. For example, if a participant only answered five questions, those five responses are considered in the data, but the remaining blank and incomplete responses are not. All

responses for each term were counted and percentages calculated. Responses that were either singular or plural (e.g. *LEGO* and *LEGOS*) were considered the same.

4 Data and Discussion

The following review of the Reddit corpus and survey data, and the resulting discussion and comparison between the two, will be discussed in four parts. In Section 4.1, I first review the rater agreement and inter-rater reliability scores between the two raters of the Reddit corpus data. Section 4.2 is a comparison between the four words that were previously studied by Hoopes (2019), namely *crock-pot*, *Kleenex*, *band-aid*, and *xerox*, and the corpus and survey data from this thesis will be discussed. Section 4.3 reviews the results of the ten generic words from the corpus and survey data. Finally, Section 4.4 review the corpus and survey data of the ten trademarked terms. These ten words are grouped together and discussed based on similar results. For example, trademarked terms that show strong generic usage will be discussed with other words that show the same.

4.1 Rater Agreement

As was outlined in the previous chapter, I randomly selected concordance lines (n=2271) from the Reddit corpus and coded each one, one of four ways: i) Specific (s) – if the use of the trademark was source identifying, (ii) Generic (g) – if the use of the trademark was product, or category, identifying, (iii) Ambiguous (a) – if, for any reason, the referent associated with the trademark is unknown or unclear, and (iv) Other (o) – if the use of the term is not associated in any way with the source of product or category of product. Another random 10 concordance lines for each of the 24 trademarked terms (n=240; 10.5%) were given to a second rater and coded with the same criteria of the first. Table 4.1 below shows the individual counts of the 240 lines both raters coded.

Table 4.1 – *Frequency count of each coding criterion*

	Specific	Generic	Ambiguous	Other
Rater 1	71	153	16	0
Rater 2	82	148	10	0

Of the 240 lines that were coded by both human judges, none of them were coded as being ‘Other’ – or having a meaning or use unrelated to the brand or product concerned. This is indicative that the corpus data consist of examples of trademark language that is, at the very least, relevant to the present thesis.

As a measurement of reliability of the raters’ coding, both a raw agreement percentage and Cohen’s kappa (Cohen, 1960) were computed. Table 4.2 below shows the number of codings that matched between raters, the total possible matches (n=240) and the agreement percentage between raters.

Table 4.2 – *Agreement percentage between raters*

Matches	Total Possible	Agreement Percentage
217	240	90.42%

The 90.42% agreement, however, may include random chance agreement. In order to account for this, the Cohen’s kappa coefficient was calculated. It is a “robust statistic useful for either inter-rater or intra-rater reliability testing” (McHugh, 2012, p. 279). Figure 4.3 below is the Cohen’s kappa statistic as it was calculated using the *R* programming language (R Core Team, 2021).

Table 4.3 – *Cohen’s Kappa for Inter-rater Reliability*

	lower	estimate	upper
Unweighted kappa	0.74	0.81	0.88

The kappa statistic of 0.81 estimates that 81% agreement between the two raters in this study when accounting for random chance with a confidence interval ranging from 0.74 to 0.88, or plus-or-minus 7% from 81%. Cohen suggests that kappa statistics between 0.61 and 0.80 be interpreted as “substantial” agreement and statistics between 0.81 and 1.00 be interpreted as “near perfect” agreement (McHugh, 2012). The agreement of the two raters of this thesis can thus be interpreted as being “substantial” to “near perfect” agreement. However, McHugh (2012) suggests that this may be too generous an interpretation as statistics of 0.61 suggest that near 40% of the data may be erroneous and suggests a narrower window of statistics between 0.80 and 0.90 be considered strong agreement. Whether Cohen’s original interpretation of the statistic or McHugh’s modified interpretation is considered, the inter-rater reliability score (0.81) of the raters of this thesis is dependable and can be considered very strong.

4.2 Crock-Pot, Kleenex, Band-Aid, and Xerox

It is important to note that Hoopes (2019) focused his coding strictly on whether the context of the concordance lines indicated a specific or generic usage. The raters in the present thesis were instructed to use their individual interpretations of how the terms were being used. They were encouraged to use the context as a guide, but the focus was on how they, as consumers, interpreted the terms. This instruction was given because the focus of the Primary Significance Test is how the terms are interpreted in the minds of the consuming public. Because of this, ambiguous codings are far less common in the present thesis when compared to Hoopes and the comparison.

Hoopes personally coded 100 concordance lines for each of the four terms *Crock-Pot*, *Kleenex*, *Band-Aid*, and *Xerox*. As seen in Table 4.4, the majority of uses for each term were

labelled as ambiguous with the exception of *xerox* which had the majority of instances being flagged as specific uses.

Table 4.4 – *Hoopes' (2019) Coding Frequencies*

	Specific	Generic	Ambiguous	Other
<i>Crock-Pot</i>	20	4	76	0
<i>Kleenex</i>	6	2	90	2
<i>Band-Aid</i>	3	15	82	0
<i>Xerox</i>	75	7	18	0

The frequencies of the coding of the Reddit corpus concordance lines in the current thesis showed similar results as Hoopes. The difference being that the majority of uses of *Crock-Pot*, *Kleenex*, and *Band-Aid* were coded as being generic rather than ambiguous (see Table 4.5).

Table 4.5 – *Reddit Corpus Coding Frequencies of Hoopes Terms*

	Specific	Generic	Ambiguous	Other
<i>Crock-Pot</i>	2	92	6	0
<i>Kleenex</i>	0	17	0	0
<i>Band-Aid</i>	2	96	2	0
<i>Xerox</i>	85	6	8	0

The substantially lower total frequency count of *Kleenex* is because there were only 17 concordance lines in total that were found in the entire Reddit corpus. Even though Hoopes' coding does not show substantial clear use of these terms generically, the number of ambiguous uses may be indicative that the terms are in a fluctuating state and are in the process of changing from a proper noun (specific) to a common noun (generic) (Clankie, 2000). In fact, many of the generically coded instances of the Reddit data may have been considered ambiguous by Hoopes. Should another individual read the concordance lines, there are many instances that may be specific uses to them. The subjectivity of meaning is one of the more difficult things about

primary significance. For example, (9) below was coded as generic by both raters of the Reddit corpus.

(9) ...Everyone else loves it but it's made in a *crock pot* and cooked for way too long and it somehow ends...

An example such as (9) may very well be considered ambiguous by some individuals. It is possible that this use is specific as a proper adjective with the omission of the common noun *slow cooker* (e.g. *a crock pot slow cooker*) which could make the use of the term ambiguous. However, as the raters of the Reddit corpora were instructed to rate concordance lines with their own interpretation in mind this was coded as generic by raters. If it is an omission of the term *slow cooker*, then this is more indicative of the term *crock pot* in (9) as being used metonymically to represent any slow cooker and thus used generically. This is the reasoning behind my coding of the concordance line as generic. The second rater was instructed to code their concordance lines according to how they perceived the use of the term. This difference in coding philosophy between the two raters in this thesis and the data from Hoopes (2019) can explain why the two sets of data are so different.

An instance of the common noun not being omitted is in example (10) below. The presence of *rep* in (10) is clearly indicative of a specific use and there is nothing in the context of (9) that makes a specific use clear.

(10) ...there is no risk and it is “literally” what the *crock pot* rep said is taking some liberties with interpretation...

The context of (10) indicating a specific use follows the life-cycle of a trademark beginning with its status as a proper adjective paired with a common noun (Clankie, 2000).

Perhaps the most surprising frequency counts from the Hoopes study and the Reddit corpus are those of *xerox*. As a term that has, over time, been in danger of genericide, it is somewhat unexpected to see such frequent occurrences of specific uses. At the very least, it may seem more reasonable to have a larger balance between specific and generic use. Indeed, the difference is substantial and statistically significant. Using the tool `table.plotter`, a robust Chi-square statistics function developed using the R coding language by Gries (2009), the overall effect of the difference can be seen. Figure 4.1 ($\chi^2(6) = 197.04$; $p = 8.12\text{e-}40$) and Figure 4.2 ($\chi^2(6) = 254.86$; $p = 3.75\text{e-}52$) below are the Chi-square graphics generated by this tool for both the Hoopes and Reddit data. The numbers in blue represent counts that occur more frequently than expected and red numbers represent those that occur less frequently than expected. The size of each number represents how far away the actual frequency count deviates from the expected count (i.e., the residual) – the larger the number, the greater the deviation. For example, in Figure 4.2, the ambiguous count of *crock-pot* is 6. The expected value is 5.06 and so the font size of the number is small because of the difference between 6 and 5.06 is small. In contrast, the expected count for specific uses of *xerox* is 27.88. The actual count of 85 is significantly larger and results in the size of the number being large.

Figure 4.1 – Hoopes’ Coding Effect Plot

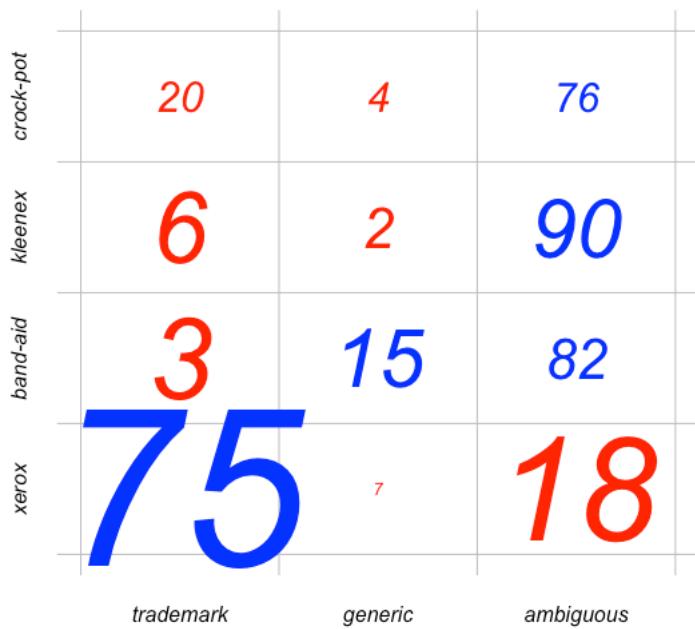
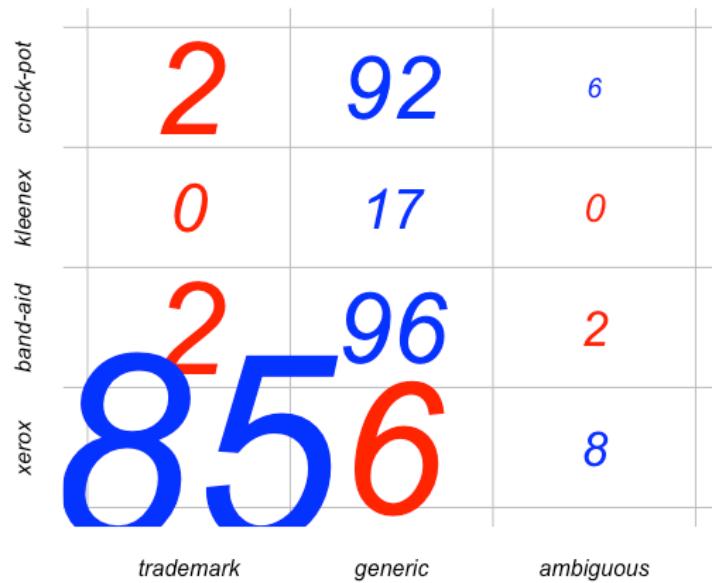


Figure 4.2 – Reddit Corpus Coding Effect Plot



As the data suggest, *xerox* is a much stronger trademark than expected given the other three. As seen in examples (11) and (12) below, *xerox* is used, as expected, as a proper adjective followed by a common noun, as well as a proper noun. Instances such as (4) were the most common of the specific uses.

(11) ...not much more secure I've been seeing some of our Xerox printers only take the serial number as the password...

(12) ...only real rebuttal I've ever heard to that is that Xerox later sued them for IP infringement...

Both examples are source-identifying to the Xerox brand, but they are different. Example (11) indicates that *xerox* is a subset of printers and (12) refers to the greater Xerox Corporation. This duality of referent is likely a factor in the term's ability to maintain its trademarked status.

Hoopes (2019) suggests this duality of referent, as well, and the data from the Reddit corpus support it. Indeed, the duality is also hypothesized as a reason for the term's generic stagnation by Clankie (2000). Clankie hypothesizes that there must be a psychological association between a brand name and a single product. A brand name, such as Chanel, that produces a variety of products cannot become generic because there is no single item association to be made.

However, this hypothesis does not always hold true. Crock-Pot, for instance, is a company and brand that is associated with a large variety of products. They are most well-known for their slow cooker products, but that is not the sole item they produce. As evident by the corpus data, *crock-pot* is being used generically despite the term's connection with multiple products. This suggests that other factors, such as marketing, market presence, competition, and possibly other linguistic factors may take precedence.

4.2.1 Crock-Pot, Kleenex, Band-Aid, and Xerox Survey Results

The survey results for the four terms both strengthen and somewhat contradict the coding of the Reddit corpus and the findings of Hoopes. Table 4.6 shows the complete results from the survey for the four terms. The column labeled as “Generic” are the number of responses that participants provided the trademarked term as their choice of word for the product in general. This is considered as using the trademarked term generically. The “Other” column is the number of responses that included a response that was something other than the trademarked term.

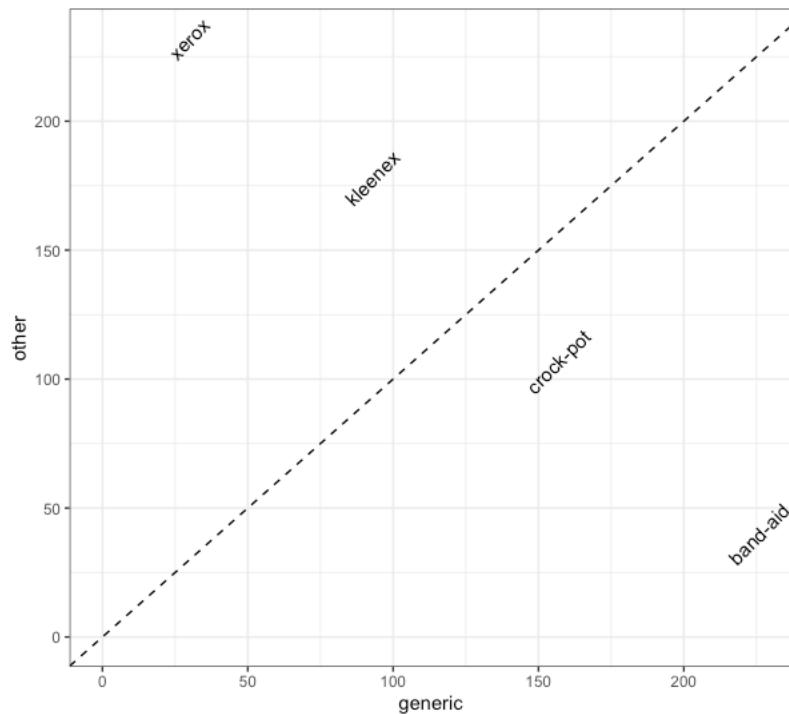
Table 4.6 – *Survey Results of Four Terms*

	Generic	Other	No Word	N/A
<i>Crock-Pot</i>	157	107	9	15
<i>Kleenex</i>	93	178	3	14
<i>Band-Aid</i>	226	40	2	20
<i>Xerox</i>	30	232	5	21

The survey data suggest that *crock-pot* and *band-aid* are the more generic of the four words. This agrees with the coding results of both the Reddit corpus and the findings of Hoopes. There were 226 of a possible 268 responses (84%) where participants gave *band-aid* as a preferred term for an adhesive bandage. This was the strongest term with generic tendencies in both the survey and corpus data. *Crock-pot* was not as frequently used generically, but over half of the responses (57%) were. The strongest trademark according to the survey data was *xerox* with only 11% of responses being a generic use of the term. The difference between the status of these words can be better visualized in Figure 4.3 below. The diagonal dashed line represents a central point for the data where words below the line are used more generically and those above the line are used more strongly as trademarks. The farther away from the dashed line a term is,

the stronger its preference toward either generic usage (below the line) or trademarked usage (above the line).

Figure 4.3 – *Chart of Four Words – Survey Data*



The corpus data show that *crock-pot*, and *kleenex* are used more generically than the survey data suggest. Hoopes marked many of the uses of these terms as ambiguous whereas the coders of the Reddit corpus labelled them as generic. The substantially smaller sample size of *kleenex* (n=17) may not be sufficient for the Reddit corpus data to accurately determine the use of that term. If the sample size was larger, it is possible more specific and trademarked usage would be seen to be more like the survey responses.

Based on Hoopes' findings, nearly all instances of *crock-pot*, *kleenex*, and *band-aid* concordance lines were considered ambiguous – 76%, 90%, and 82%, respectively. The survey data provide insight into these terms as to how they are perceived by the public and help disambiguate the terms. *Band-aid* is strongly considered to be generic, whereas *crock-pot* and

kleenex seem to be more in flux and in the middle ground of being either more generic or specific in use.

The survey data strongly support the findings by both Hoopes and the raters of the Reddit corpora that *xerox* is not being used generically and its perception as a trademark is strong. One explanation for this is, as stated previously, that there is a dual sense of the term. One is the brand itself as a type of photocopying machine and the other as a corporate entity. *Crock-pot*, *Kleenex*, and *Band-Aid* are all owned by corporations with different names – Holmes Product Corp., Kimberly-Clark, and Johnson & Johnson Consumer Inc., respectively. It is also possible that using *Xerox* generically is generational. Technological improvements happen rapidly and photocopiers produced by Xerox may be less ubiquitous than in the past as the technology has become obsolete or, at the very least, less prevalent. This is not readily apparent from the demographics of the participants of the survey in this thesis as responses that indicated *xerox* being used generically were equally spread out across age groups.

4.3 Ten Generic Terms

Hoopes (2019) only used COCA, a generalized corpus, in his research and so the inclusion of the four terms in this thesis is to show what insight a non-generalized corpus such as the Reddit corpora and survey data can provide in disambiguating the terms and how the survey data support or contradict corpus findings. These goals for the comparison of the two sets of data are the same hereafter, but the previous four terms were considered separately from the rest of the data as a strict comparison with previous research conducted on those four terms.

The Reddit corpus coding results for the 10 generic terms were overwhelmingly one-sided. Every concordance line for 8 of the 10 terms was coded as being generic (see Table 4.7). It is important to note that *videotape* only had a total of 34 concordance lines from its corpora

and so its results may not be as representative as the others. This overwhelming representation of generic use is expected as the 10 terms have all lost any trademarked status and are legally considered generic terms with the exception of *aspirin* which still hold its trademark in many countries outside of the United States. The two terms that were not unanimous in their coding were *teleprompter* and *aspirin*.

Table 4.7 – *Reddit Corpus Coding Frequencies of Generic Terms*

	Specific	Generic	Ambiguous	Other
<i>Cellophane</i>	0	100	0	0
<i>Escalator</i>	0	100	0	0
<i>Flip Phone</i>	0	100	0	0
<i>Kerosene</i>	0	100	0	0
<i>Laundromat</i>	0	100	0	0
<i>Linoleum</i>	0	100	0	0
<i>Trampoline</i>	0	100	0	0
<i>Videotape</i>	0	34	0	0
<i>Aspirin</i>	0	88	12	0
<i>Teleprompter</i>	0	98	2	0

The two ambiguous uses of *teleprompter* (see examples (13) and (14) below) were ambiguous because there was little to no context in the concordance lines. There is no clear way to know without context which meaning is attributed to the word. However, because the remaining 98 lines are coded as generic, it is likely that they would be generic, but there is no linguistic evidence to say one way or another.

(13) eL **tElEpRoMpTeR** !

(14) * **Teleprompter**

Example (13) could be considered generic with the use of the Spanish *el* as it could be interpreted as a definite article meaning “the”. This may be a Reddit comment or post that was written in Spanish and the use of *el* may be indicative a generic use. Without looking deeper into

the individual posts for these types of concordance lines as they are seen on the Reddit website or app, a better understanding of these uses of *teleprompter* may be impossible to apprehend. For instance, there may be a picture or video that accompanied the posts as viewed on the Reddit website with a web browser, whereas the humans who gave their perception only had the text available.

Aspirin is different than *teleprompter* in that there were some ambiguous uses that had sufficient context surrounding the term. One line of note (example (15) below) was coded as ambiguous by rater one and generic by rater two.

(15) ...testifies it IS ABSOLUTE TRASH. Do I write for 81mg aspirin BID? Of course I do, but I also include...

This is one of the few instances both raters disagreed in their coding. When Clankie's (2000) hypotheses are considered, it is more likely that this is a generic usage. Typically, if the use of a trademarked term is specific then it is either a proper adjective that modifies a common noun or as a proper noun, as has been discussed previously. The example seen in (15) shows *aspirin* as the head noun that is being modified by *81mg*. Proper nouns other than geographical locations or names are rarely modified by adjectives or other nouns and proper adjectives are the modifiers. Figure 4.4 shows the 25 most frequent (occurrences per million words) proper nouns preceded by adjectives from COCA¹⁶.

¹⁶ This query was entered into the COCA interface as: ADJ NAME (any adjective preceding a proper noun)

Figure 4.4 – Adjective-*Proper Noun* Combinations from COCA

1	□	LATIN AMERICA	9527	9.59
2	□	SOUTHERN CALIFORNIA	8985	9.05
3	□	EASTERN EUROPE	6082	6.12
4	□	NEW YORKER	6029	6.07
5	□	NEW YORKERS	3495	3.52
6	□	UNINTELLIGIBLE)	3471	3.50
7	□	CENTRAL AMERICA	3398	3.42
8	□	INAUDIBLE)	3277	3.30
9	□	NORTHERN CALIFORNIA	3231	3.25
10	□	CORPORATE AMERICA	2182	2.20
11	□	CENTRAL ASIA	1989	2.00
12	□	BLACK FRIDAY	1960	1.97
13	□	CENSORED)	1776	1.79
14	□	UPSTATE NEW	1770	1.78
15	□	NORTHERN VIRGINIA	1750	1.76
16	□	SUB-SAHARAN AFRICA	1597	1.61
17	□	GREAT BRITAIN	1533	1.54
18	□	NAZI GERMANY	1496	1.51
19	□	LOWER MANHATTAN	1465	1.48
20	□	GRAND CANYON	1418	1.43
21	□	CENTRAL EUROPE	1223	1.23
22	□	NORTHERN IRAQ	1162	1.17
23	□	LATE OCTOBER	1077	1.08
24	□	LATE AUGUST	1069	1.08
25	□	MAGIC JOHNSON	1041	1.05

Figure 4.4 shows that locations such as America, Africa, Germany, Europe, etc. are the most frequent occurrences of adjective-*proper noun* combinations with *Magic Johnson* as the sole person that occurs in this environment. It is also apparent that many of the terms in Figure 4.4 seem to be multi-word proper nouns that happen to be comprised of words that are classified as adjectives on their own. *Southern California* or *late October* are two examples that seem to be genuine adjective-*proper noun* combinations, but a term such as *Black Friday* may be seen more as a proper noun in and of itself. Because of this, it is likely that (15) is generic rather than ambiguous or specific.

4.3.1 Ten Generic Terms Survey Results

The survey data do not show that these terms are as strongly used generically as the Reddit corpora do. Some terms such as *trampoline* or *escalator* showed strong generic use while terms such as *kerosene* and *videotape* showed completely the opposite. In this regard the survey

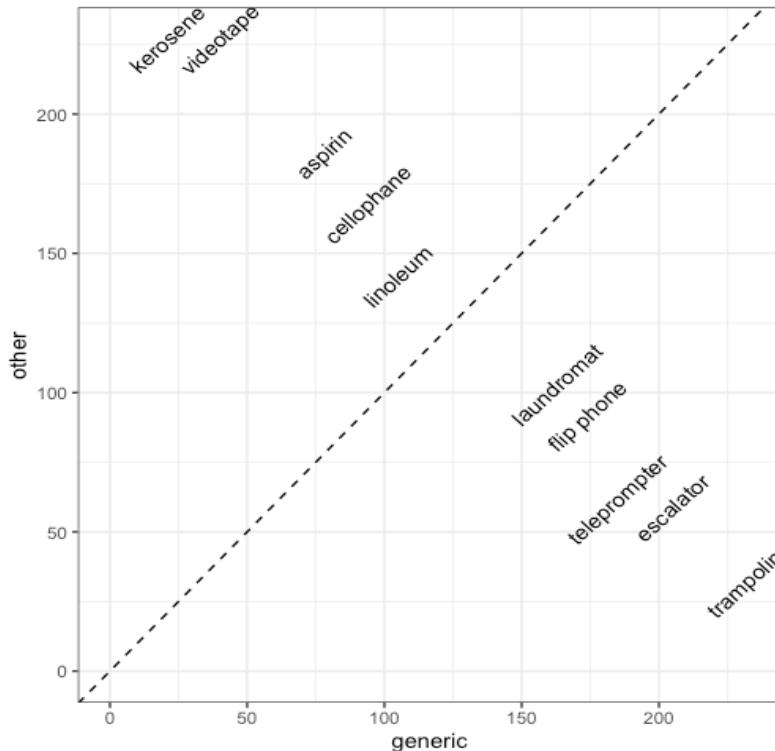
data and the Reddit corpora data do not agree. Table 4.6 below shows the complete results of the survey for the ten generic terms.

Table 4.8 – *Survey Results of Ten Generic Terms*

	Generic	Other	No Word	N/A
<i>Cellophane</i>	94	168	22	4
<i>Escalator</i>	205	59	4	20
<i>Flip Phone</i>	174	92	12	20
<i>Kerosene</i>	21	227	20	20
<i>Laundromat</i>	163	103	3	19
<i>Linoleum</i>	105	142	21	20
<i>Trampoline</i>	233	33	2	20
<i>Videotape</i>	40	227	6	15
<i>Aspirin</i>	78	186	7	17
<i>Teleprompter</i>	185	62	22	19

Because these generic terms have already lost their trademark, this disagreement does not suggest that these terms are associated with any specific brands or companies. Half of the generic terms were strongly used generically, while the other half were not as demonstrated in Figure 4.5 below. Again, this is not necessarily because these terms are being used as trademarks, but based on the other responses, the public is using other terms than these to describe the products. For example, only 40 responses from participants showed *videotape* as the preferred term. However, 104 of responses (37.7%) showed that *VHS* was the preferred term.

Figure 4.5 – Survey Responses of Generic Terms



The presence of other competing products also plays a part in how these terms are being used by the survey participants. For example, there are several over-the-counter medicines to reduce pain other than aspirin. Generic drugs such as ibuprofen or acetaminophen and their trademarked counterparts (Advil® and Tylenol®, respectively) were as equally represented as aspirin. Conversely, terms such as *escalator* or *trampoline* do not seem to have as many, if any at all, competing products. The most common responses other than *escalator* was *elevator* which occurred 44 times.

Overall, the survey data do not show as strong of generic use of these terms, but the Reddit corpora data are not necessarily invalidated because of this. The lack of generic use of terms such as *videotape*, *kerosene*, or *aspirin* can also indicate the obsolescence of the terms or products, or the stronger presence of other products.

4.4 Ten Trademarked Terms

The coding results of the ten trademarked terms yielded more varied results. There were two terms, *realtor* and *styrofoam*, that were coded primarily as generic, whereas others were coded primarily as specific such as *coke* or *nintendo*, the latter having 100% of its instances coded as such. The complete results of the trademarked terms (Table 4.9) vary and are discussed here in two different groupings: 1) trademarked terms that are used generically primarily as verbs; 2) trademarked terms that only exhibited generic use as nouns or proper adjectives. This second grouping is discussed in even smaller groupings based on their similarity of coding in the corpora data. For example, *coke* and *nintendo* were both coded primarily as specific usages and are thus analyzed together.

Table 4.9 – *Reddit Corpus Coding Frequencies of Trademarked Terms*

	Specific	Generic	Ambiguous	Other
<i>Coke</i>	91	0	10	0
<i>LEGO</i>	54	16	30	0
<i>Nintendo</i>	100	0	0	0
<i>PowerPoint</i>	42	45	13	0
<i>Realtor</i>	5	93	2	0
<i>Styrofoam</i>	0	98	2	0
<i>Google</i>	94	5	1	0
<i>Photoshop</i>	83	6	11	0
<i>Uber</i>	91	3	6	0
<i>Zoom</i>	62	10	7	21

The four trademarked terms that were used as verbs were coded mostly as specific uses. *Zoom* showed significant usage (21%) in ways other than as a trademark. This is due to its sense relating to photography and cameras as in example (16). Phrases such as *zoom in* or *zoom out* are integral ideas to those fields. Of the remaining 79 concordance lines, 62 of them (78%) were in direct reference to the Zoom product or company, as in example (17).

(16) ...you can take a picture from really far away and zoom it in. The moon will appear bigger...

(17) ...can't you have kids read some sentences aloud on zoom like in class?

In example (17), this use of *zoom* is indicative of the specific use of the platform Zoom. It could be in reference to any video conferencing software. However, that type of use is shown better in example (10) where the term is preceded by the indefinite article *a*.

(18) ...students can't come to school seems tacky. Do a zoom instead.

This example could be in direct reference to the Zoom platform, but the use of the indefinite article further removes the meaning from its source. For example, with other proper nouns and common nouns, one might say “There is a cat on Gary”. This example is similar to (17) where there is no article. One might also say “There is a cat on a table” – as in example (18) as it relates to *zoom*. *Gary* is a proper noun that cannot be preceded by neither *a* nor *the*. This behavior is seen with the term *uber*, as well. (19) and (20) are similar uses to (16) and (17), respectively.

(19) ...I see people going out of their way to use uber even when a regular taxi would be cheaper...

(20) ...forward so thought I imagined it. I woke up in an uber going to a random address in a city near mine...

The survey data also show *uber* and *zoom* are used more generically than not. However, not as strongly as what the Reddit corpora data suggest. Just over half of responses show that participants use the term *zoom* for a video conference (see Table 4.10). The other half of responses show that participants used either other trademarks such as Skype, WebEx, or Google Meet as their preferred term or just the generic term *video conference*.

Table 4.10 – Survey Responses for Uber and Zoom

	Generic	Other	No Word	N/A
<i>Uber</i>	167	95	15	11
<i>Zoom</i>	179	174	21	22

The fact that *zoom* is not used generically according to the questionnaire data as strongly as the Reddit corpora data suggest may be due to the recent wide-scale need for video conferencing software with the COVID-19 pandemic. Some age groups of participants of the survey may not have been in situations that required the use of video conference software and so they have not been placed in a situation for the term to become the prevailing term in their minds.

As for *google* and *photoshop*, any generic use was strictly as a verb as in (21) and (22). There were no instances of the terms being used as nouns like in (19) and (20) for *zoom* and *uber*.

(21) ...recommend that you take mathematical formula and Google it or look it up in a text...

(22) It looks like you are on really bad terms if you photoshop her out of pictures

Clankie's (2000) hypotheses suggest that trademarks that become generic may be used as verbs after becoming common nouns. As there are no instances from the Reddit corpus of *google* or *photoshop* being used this way, it is not evident either of these terms are currently subject to genericide.

The survey data show that the verb usage of these terms is very strong. While the Reddit corpora data only show a few instances of verb usage, the survey data show that the verb usage of these terms are very strong (see Table 4.11 below).

Table 4.11 – Survey Responses for Google and Photoshop

	Generic	Other	No Word	N/A
<i>Google</i>	150	128	6	4
<i>Photoshop</i>	172	84	15	17

Google is similar to *zoom* above in that it is only used as a verb by just over half (52%) of all responses. There were 98 responses (76% of responses labeled “Other”) that show participants use *search the internet* or *internet search* as their preferred term. *Photoshop* is used more frequently as a verb than *google* with 63.4% of participants stating that as their preferred term. The survey questions for these two terms may not be as useful when comparing the two data sets. The survey questions were specific to the verb usage and so it is not surprising that the two terms showed more generic usage that way than the Reddit corpora data. The data from the corpora were not specific to the noun or verb usage and so, in this case, the corpora data are probably more useful in understanding the status of the two terms than the survey data.

The remaining six terms showed varying results in how they were coded. Two terms, *realtor* and *styrofoam*, were coded almost completely as generic with *realtor* having 93% of its lines coded as generic and *Styrofoam* coded as 98% generic with 100% agreement between raters. The ambiguous concordance lines of *styrofoam* (n=2) were similar to those of *teleprompter*. There was so little context surrounding the term that there was no way to determine the intended meaning from its context alone as can be seen in examples (23) and (24) below.

(23) ...is **Styrofoam**.

(24) ...or **styrofoam**)

Realtor is a unique trademark because it is a title for real estate agents who have joined the National Association of Realtors. The ambiguous instances of *realtor* were primarily instances where the trademarked title and its use as a generic title were indistinguishable from the context as in (25). Nevertheless, it, too, was seen by the raters as primarily being generic in use.

(25) ...so many buyers looking for the same thing. I'm a realtor and it has been an awkward year to say the least.

It is likely that this individual is part of the National Association of Realtors or is simply a real estate agent referring to themselves as a *realtor* and thus (25) was considered ambiguous.

The survey data for these two terms differ quite substantially with the data from the corpora. Each of these terms trended generic in their use, but not as strongly as the corpora data suggest. Table 4.12 shows that less than half of responses (49.5%) showed that participants use *realtor* as their preferred term and 61% used *styrofoam* as their preferred term.

Table 4.12 – *Survey Responses for Realtor and Styrofoam*

	Generic	Other	No Word	N/A
<i>Realtor</i>	135	130	8	15
<i>Styrofoam</i>	177	106	4	1

Responses such as *foam* or phrases that included the term *foam* account for 50 of the 106 other responses (47%). It is possible that these phrases were caused by backformation from the term *styrofoam* or the generic product name *polystyrene foam*. If they were created via backformation from the trademarked term, then that would suggest that *styrofoam* is more generic than shown from the current survey results. 96 of the 130 other responses for *realtor* (73.8%) comprised of either the term *real estate agent* or *housing intermediary*. The 96 responses from participants that

used these terms still make up less than half of the responses for *realtor*, but their presence suggest that *realtor* is not as generic as the data from the corpora suggest.

Opposite to the results of *realtor* and *styrofoam* were the terms *nintendo* and *coke* which were considered to have strong statuses as trademarks. *Nintendo* concordance lines were coded 100% as specific and 91% of *coke* as specific with 100% agreement between raters. Neither showed any instances of generic use, as well. Some instances of *coke* were ambiguous due to *coke* also being a slang term for *cocaine*. Instances such as (26) do not indicate whether the use of the term is the brand, the generic use of the term for any soda, or short for cocaine. Other ambiguous instances were similar to those discussed previously where there was little to no context in the concordance lines to provide context for the use.

(26) Obviously OP is a *coke* fiend

The survey results for *coke* and *nintendo* are very similar to the data from the corpora as seen in Table 4.13 and strongly support those findings. Only 5% of responses indicate *coke* as the preferred term, or in other words, 95% of responses were something other than the term *coke*. This strongly supports the coding of the concordance lines with found that 91% of those lines were clearly specific and trademarked uses. The strong trademarked usage of *coke* may likely be due to regional variability in American English. Many American English speakers in the southern United States use *coke* instead of *soda* (2021). However, of the 14 responses of generic use, only 4 were from participants hailing from southern States – 1 from Washington D.C. and 3 from Texas.

Table 4.13 – *Survey Responses for Coke and Nintendo*

	Generic	Other	No Word	N/A
<i>Coke</i>	14	263	3	8
<i>Nintendo</i>	25	227	19	17

The survey responses for *nintendo* show that 9.2% of responses indicate *nintendo* as the preferred term of participants. The remaining 90.8% were either competing brands such as PlayStation or Xbox, or a generic term such as *video game console* or *gaming console*. The survey data for both *coke* and *nintendo* confirm the coding of the corpora data.

The final two terms to discuss, *LEGO* and *powerpoint*, were the most varied in terms of coding of the trademarked terms. *Powerpoint* showed generic usages that were similar to *uber* and *zoom* where the term was used after an indefinite article as a common noun rather than as a specific product as in (27). There are instances of *powerpoint* being pluralized, as seen in (28), which are only found among generic usages, as well.

(27) ...time you made that comment you were not making a ***PowerPoint*** you were
actually making a common on this Reddit

(28) ...like such an absurd use of my time to make pretty ***powerpoints*** with absolute
garbage findings/recommendations...

In comparison with the corpora data, the survey data show that both *lego* and *powerpoint* are perceived more generically as seen in Table 4.14.

Table 4.14 – *Survey Responses for LEGO and PowerPoint*

	Generic	Other	No Word	N/A
<i>LEGO</i>	216	62	2	8
<i>PowerPoint</i>	200	55	18	15

The data from the corpora show that *lego* had 30 concordance lines (30%) that were considered ambiguous. The survey data show that 216 responses (77.1%) were instances of the term being used generically. This suggests that many of the lines coded as ambiguous may be more likely generic uses than specific. Because of LEGO's strong market presence and lack of competitors,

it is more likely that the trademark is the preferred term to the public than something else. One of Clankie's (2000) hypotheses suggests that trademarks that are shorter and simpler than their counterparts are more likely to become generic. This may be the case with *lego* as the other most frequent responses among participants was *plastic block toy* (11 total responses) or *building block* (15 responses).

The data from the corpora show that 45% of *powerpoint* concordance lines were generic and so it seems as though the term is somewhere in the middle of being generic or a trademark. However, the 73.3% of responses from participants that show the term being used generically suggest that the term is considered more generic than the corpora suggest.

For all 24 terms analyzed in this thesis, the survey data and the corpora data show varying degrees of agreement. While there is some variation and divergence between the survey and corpora data, the discussions above show that, overall, the survey data support the findings from the corpora data. With some terms such as *coke*, *nintendo*, or *xerox*, the survey data are in clear support of the corpora data. They also show some ability to disambiguate terms such as *band-aid*, *crock-pot*, and *lego*. The survey results were not as helpful in determining the status of *google* and *photoshop* primarily due to the phrasing of the survey questions. The fact that the survey data were not as reliable with these two terms do not invalidate the effectiveness of the data to support the corpora data or provide insight into the ambiguous usage overall.

5 Conclusion

The results from this thesis in no way seek to be a definitive answer to the legal status of the trademarks studied. Researchers have recently called for the use of corpus data in determining the semantic status of any word or phrase, not just trademarks, that come under the proverbial legal microscope (Cunningham & Egbert, 2019; Hoopes, 2019; Lee & Mouritsen, 2018; Mouritsen, 2010; Phillips et al., 2016; Phillips & White, 2018). The goal of this thesis is to show how the data from experimental methods, via a linguistic questionnaire, compare and contrast with, as well as support, observational data from corpora. The use of both experimental data and observational from a corpus was suggested by Gilquin and Gries (2009) and this thesis uses both to determine the semantic status of trademarks that have undergone genericide or have “flirted” with genericide (Hoopes, 2019, p. 411). These comparisons were done in an effort to answer three questions:

1. How do the combination of corpora and survey data compare to those of Hoopes (2019)?
2. How do corpora data compare and contrast with the experimental survey data for terms that have become legally generic?
3. How do corpora data compare and contrast with the experimental survey data for terms that have retained their trademarked statuses?

As is evident by the analyses and discussions above, the degree to which these two methods agree with each other can vary. Overall, however, the survey data support the findings from the corpora data despite some variation between the survey and corpora data.

For the terms that showed strong trademarked usage in the corpora data, namely *coke*, *nintendo*, and *xerox*, the survey data showed overwhelmingly similar numbers in agreement with

the codings of the raters. Other terms, such as *styrofoam*, *realtor*, and *zoom*, showed stronger generic usage in the corpora data than in the questionnaire data. These differences, however, can possibly be explained by other social and linguistic factors that are not the focus of this thesis.

Zoom, for example, exhibited a more mixed response of use from participants of the questionnaire compared to the strong generic usage found in the corpora. I hypothesize that this could be a difference in life experience among participants of the questionnaire. The questionnaire included individuals from different ages and countries of residency who may not have required the use Zoom video conferencing software during the COVID-19 pandemic. Thus, some participants may not have acquired the term in a generic sense. These types of circumstances that may facilitate the acquisition of a certain term in a generic sense rather than a trademarked sense may be a topic for future research.

This thesis places great emphasis on a comparison with the findings of Hoopes (2019). He found that many instances of trademarked terms found within COCA are ambiguous despite context. In comparison with Hoopes' findings, the questionnaire data and the coding of the concordance lines from the Reddit corpora were in strong agreement with one another. *Xerox* was one term in particular that was consistent between Hoopes and the data and analysis in the present thesis. There was some divergence between how the two raters of the Reddit corpora compared to the coding of Hoopes, but the questionnaire data did succeed in disambiguating and providing further insight into the statuses of *crock-pot*, *kleenex*, and *band-aid*. Hoopes coded a majority of his concordance lines from COCA for these three terms as ambiguous while the raters in this thesis coded the concordance lines from the Reddit corpora primarily as generic.

Hoopes (2019) focused his coding on determining the intent of the utterer whereas the raters of the Reddit corpora in this thesis focused on their interpretation of the terms based on

context and their own perception of the term. The difference between the results and the motivations behind each rater give rise to a question that may be critical for future research and in the interpretation of genericide and the Primary Significance Test when using corpora. That is, what is more important when determining primary significance in the minds of the consuming public – the intent of the utterer (speaker/writer) or the interpretation of the perceiver (listener/reader)? The fact that the survey data agreed more with the raters of the Reddit corpora suggests that the latter – the interpretations of the perceiver – may be more indicative of the status of a term in the minds of the consumers.

The approaches of the two methodologies are fundamentally different and this could explain why some terms such as *videotape*, *aspirin*, and *Kleenex* show variation between data sets. The analysis of corpus data is inherently semasiological and looks to identify meaning of a particular term, whereas the survey data is inherently onomasiological and asks participants to identify the term given a particular meaning. This is evident in the analysis of the ten generic terms in which the Reddit corpora data showed almost 100% generic usage for all ten generic terms from this study as would be anticipated. However, the survey data showed more varied results. This is not necessarily because these terms are being used as trademarks, but based on the other responses, the public is using other terms than these to describe the products. For example, the generic term *videotape* was only represented by 14.6% of responses whereas 38.1% of responses showed *VHS* as the preferred term showing that participants use a trademarked term twice as often as the generic equivalent. The presence of other competing products also plays a part in how these terms are being used by the survey participants. For example, there are several over-the-counter medicines to reduce pain other than aspirin. Other generic drugs such as ibuprofen or acetaminophen and their trademarked counterparts (Advil® and Tylenol®,

respectively) were responses that were as equally given as aspirin. Conversely, terms such as *escalator* or *trampoline* do not seem to have as many, if any at all, competing products.

To address the difference in approach between the two data sets, future research may want to focus on provided participants a more semasiological-oriented survey. In such a survey, participants may be presented with a generic or trademarked term and asked to use it in a sentence or define it. This would match the approach of the corpus data in that participants are asked to identify meaning given a term, rather than identify a word given a meaning.

There were also discrepancies between survey and corpora data for the terms *google* and *photoshop*. This was primarily due to a difference in scope between the wording survey prompts and corpus queries. The survey prompts for these terms were focused on asking which terms participants use to refer to the actions relating to these terms. For example, the prompt given to participants for the term *google* was:

Please provide the word, or phrase, you would use to refer to performing a search on an internet browser.

The corpus query for this term, however, looked for all uses of the term as a noun, verb, or otherwise. This resulted in the data from the Reddit corpora having only a few instances of verb usage (5%) and the survey data showing over half of responses from participants (52%) being the term used as a verb.

The varying results between the two sets of data for terms such as *videotape* or *google* show that corpus queries and survey prompts can be both too narrow or broad in scope and a greater effort must be taken in order to ensure the two are more unified in the word and its parts of speech to be included in the data. This could be accomplished by either tagging the corpus data for parts of speech so that queries can focus on verb or noun usage. Also, survey prompts

can be worded more carefully in order to elicit results that match corpus queries. This issue of scope is one of the more significant limiting factors of the research in this thesis and must be considered for further research. The questionnaire prompt for *videotape*, for instance, was too broad or gave too much freedom to participants because the target generic term was not properly elicited as was evident by *VHS*, a trademarked term, being the majority response provided. A fixed response questionnaire in which participants select terms rather than provide their own terms may be the solution to this type of issue. However, that in turn loses the authenticity of a participant spontaneously producing the term on their own. The actual generation of language is more desirable, I believe, in a study such as this in which the evaluation of how a term is used and perceived is the focus.

The free response format of the survey itself provided other limiting factors that did not impact the ability of the data to answer the research questions but did result in data that was irrelevant to the present research. Participants were not required to answer any questions other than the consent form. This resulted in some participants not answering every question and thus an unequal amount of data points associated with each term. Fortunately, the number of participants (n=317) that completed every question was enough to provide an adequate number of results for each term as a sample of 30 or more is generally accepted as being adequate in size. Additionally, participants were permitted to provide any answer to any question. Some participants gave jovial answers or answers completely disassociated with the term in question. For example, for the prompt associated with *powerpoint*, one participant responded with the phrase *the winds of change*. This was obviously not a serious response and was therefore not considered. Answers such as these were only considered on an individual basis. That is, if a participant gave a non-serious answer for one prompt but serious answers for the rest of the

prompts, the serious answers were still considered in the data. This causes sample sizes to be uneven, but as no sample size were below 30 even after discarding non-serious answers this is not a major concern.

The survey was also distributed online via social network websites. This mass distribution method may have resulted in the link to the survey being given to individuals who spammed the survey with multiple entries under different email address. I do not believe the intent was malicious but simply an opportunity for someone, or a group of individuals, a greater chance at winning a \$25 gift card to Amazon. This possibility was evident by there being about 20 entries that were identical in every response and showed strong similarity in how the email addresses associated with the responses were comprised. These responses were excluded from the research for this reason.

Because both the survey data and the corpora data support each other overall, they show that each method may be a viable way to determine the semantic status of a trademark. It may be that corpus analyses are more conducive to the needs of a courtroom. Lawyers, judges, and expert witnesses alike can learn to use corpus data to show evidence of how trademarks are being used concurrently. It may be more efficient and useful for a court to find and use two raters, such as was done in the present thesis, than to conduct a survey. However, the intent of the raters, as well as how they code concordance lines, must be considered. The research done in this thesis shows that corpus data can provide quick and beneficial insight to the status of trademarks to courtrooms and ought to be considered.

Linguistic questionnaires and other experimental methods can be expensive and time-consuming and thus not an economical way to research the status of a trademark within a courtroom. However, businesses and corporations seeking to protect their trademarked terms

may want to consider consumer feedback via surveys or other means to determine how their trademarked terms are being perceived and used by the public. The research from this thesis shows how questionnaires support and validate corpus findings and suggest that they are a viable method to determining how trademarks are being perceived by the public.

Further linguistic research in what constitutes a generic term can provide further substantiation of corpus and experimental data. Clankie's (2000) hypotheses on how trademarks become generic, especially his hypothesis on the regular process of a trademark's life (i.e. the trademarked term's movement as a proper adjective or noun (specific) to a common noun (generic)) is a strong foundation to be built upon. Introducing more syntactic, morphological, and other semantic methods in researching the linguistic differences between generic (formerly trademarked) and currently trademarked terms may provide insight on patterns and processes that have not yet been realized.

Appendices

Appendix A – Python script used to scrape Reddit for text

```
import os, praw

# SUPPLY YOUR APP'S INFO
reddit = praw.Reddit(client_id='ID NUMBER HERE', # 14-character code
                      client_secret='SECRET HERE', # 27-character code
                      user_agent='AGENT NAME HERE', # your app's name
                      username='REDDIT USERNAME HERE', # your reddit username
                      password='REDDIT PASSWORD HERE') # your reddit password

# SPECIFY THE SUBREDDIT COMMUNITIES YOU WANT TO SCRAPE
subreddits_to_search = ['technology', # google
                        'programming', # google
                        'AskReddit', # google, uber
                        'internet', # google
                        'artificial', # google
                        'soda', # coke
                        'ToFizzOrNotToFizz', # coke
                        'food', # coke
                        'AskCulinary', # coke
                        'SodaStream', # coke
                        'GameDeals', # nintendo
                        'Games', # nintendo
                        'gamecollecting', # nintendo
                        'gaming', # nintendo
                        'consoledeals', # nintendo
                        'RealEstate', # realtor
                        'legaladvice', # realtor
                        'realestateinvesting', # realtor
                        'homeowners', # realtor
                        'FirstTimeHomeBuyer', # realtor
                        'teenagers', # zoom, powerpoint
                        'Professors', # zoom
                        'Teachers', # zoom
                        'interestingasfuck', # zoom
                        'SelfDrivingCars', # uber
                        'sanfrancisco', # uber
                        'business', # uber
                        'CreditCards', # uber
                        'picrequests', # photoshop
                        'DigitalPainting', # photoshop
                        'postprocessing', # photoshop
```

```
'Art', # photoshop
'graphic_design', # photoshop
'consulting', # powerpoint
'LifeProTips', # powerpoint
'FellowKids', # powerpoint
'softwaregore', # powerpoint
'pokemon', # lego
'Showerthoughts', # lego
'todayilearned', # lego
'legodnd', # lego
'Toys', # lego
'MadeOfStyrofoam', # styrofoam
'mildlyinteresting', # styrofoam
'oddlysatisfying', # styrofoam
'environment', # styrofoam
'mildlyinfuriating', # styrofoam
'OfficeDepot', # xerox
'mechmarket', # xerox
'sysadmin', # xerox
'ventagecomputing', # xerox
'todayilearned', # xerox
'teenagers', # band-aid
'firstaid', # band-aid
'Warts', # band-aid
'WTF', # band-aid
'mildlyinfuriating', # band-aid
'firstworldproblems', # kleenex
'DoesAnybodyElse', # kleenex
'depression', # kleenex
'sad', # kleenex
'ForeverAlone', # kleenex
'slowcooking', # crock pot
'PressureCooking', # crock pot
'PlantBasedDiet', # crock pot
'food', # crock pot
'Cooking', # crock pot
'ObscureMedia', # videotape
'movies', # videotape
'boxoffice', # videotape
'unpopularopinion', # videotape
'TrueFilm', # videotape
'PhonesAreBad', # flip phone
'apple', # flip phone
'nostalgia', # flip phone
'nosurf', # flip phone
'dumbphones', # flip phone
```

```

'science', # aspirin
'Health', # aspirin
'explainlikeimfive', # aspirin
'conspiracy', # aspirin
'medicine', # aspirin
'energy', # kerosene
'preppers', # kerosene
'DaysGone', # kerosene
'camping', # kerosene
'Outdoors', # kerosene
'printmaking', # linoleum
'HomeImprovement', # linoleum
'DIY', # linoleum
'CleaningTips', # linoleum
'InteriorDesign', # linoleum
'Whatcouldgowrong', # escalator
'therewasanattempt', # escalator
'holdmybeer', # escalator
'deadmalls', # escalator
'escalator', # escalator
'politics', # teleprompter
'news', # teleprompter
'Filmmakers', # teleprompter
'videography', # teleprompter
'PoliticalHumor', # teleprompter
'shameless', # laundromat
'Laundromats', # laundromat
'childfree', # laundromat
'Frugal', # laundromat
'Entrepreneur', # laundromat
'TrampolineTricks', # trampoline
'Trampoline', # trampoline
'Wellthatsucks', # trampoline
'woahdude', # trampoline
'Tricking', # trampoline
'PlayingCardsMarket', # cellophane
'Random_Acts_Of_Amazon', # cellophane
'playingcards', # cellophane
'crafts', # cellophane
'bettafish' # cellophane
]

# SPECIFY THE MAXIMUM NUMBER OF COMMENTS TO SCRAPE
max_comments = 1000
os.chdir("FILE PATH TO SAVE .TXT FILES TO")

```

```

for s in subreddits_to_search:
    with open(s + ".txt", "w") as fout:
        print("Working on subreddit:", s)
        subreddit = reddit.subreddit(s)

    top_subs = subreddit.top(limit=max_comments)

    topics_dict = {}
    topics_dict["title"] = []
    topics_dict["id"] = []

    for i in top_subs:
        topics_dict["title"].append(i.title)
        topics_dict["id"].append(i.id)

    for v in topics_dict["title"]:
        fout.write(v + "\n")

    ids = topics_dict["id"]

    print(f"There are {len(ids)} posts gathered from the '{s}' subreddit")

    for index,id in enumerate(ids):
        print("Working on id #" + str(index + 1) + ":", id)
        submission = reddit.submission(id)
        submission.comments.replace_more(limit=0)
        for comment in submission.comments.list():
            fout.write(comment.body + "\n")

```

Appendix B – Python Script Used to search regexes and create concordance lines

```
import os, re

dir_name = 'Aspirin'
# dir_name = 'BandAid'
# dir_name = 'Cellophane'
# dir_name = 'Coke'
# dir_name = 'CrockPot'
# dir_name = 'Escalator'
# dir_name = 'FlipPhone'
# dir_name = 'Google'
# dir_name = 'Kerosene'
# dir_name = 'Kleenex'
# dir_name = 'Laundromat'
# dir_name = 'LEGO'
# dir_name = 'Linoleum'
# dir_name = 'Nintendo'
# dir_name = 'Photoshop'
# dir_name = 'Powerpoint'
# dir_name = 'Realtor'
# dir_name = 'Styrofoam'
# dir_name = 'Teleprompter'
# dir_name = 'Trampoline'
# dir_name = 'Uber'
# dir_name = 'Videotape'
# dir_name = 'Xerox'
# dir_name = 'Zoom'

os.chdir(f'FILEPATH FOR INDIVIDUAL FOLDERS BASED ON TERM')
with open(f'FILEPATH WHERE TO SAVE CONCORDANCE LINES/{dir_name}_lines.csv', mode = 'w') as fout:
    fout.write('pre\tmatch\tpost\n')
    filenames = [f for f in os.listdir() if re.search(r"\.txt$", f, flags=re.I)]
    for f in filenames:
        with open(f, encoding="utf8") as infile:
            for line in infile:
                line = line.strip()
                line = re.sub(r'\t', "", line)
                matches = re.finditer(r'REGULAR EXPRESSION HERE', line, flags=re.I)
                for m in matches:
                    try:
                        pre = line[m.start()-50: m.start()]
                    except IndexError:
                        pre = line[:m.start()]
                    fout.write(f'{pre}\t{m}\t{line[m.end():]}\n')
```

```
try:  
    post = line[m.end(): m.end() + 50]  
except IndexError:  
    post = line[m.end():]  
fout.write(pre + '\t' + m.group() + '\t' + post + '\n')
```

Appendix C – Complete survey

Q1. My name is Richard Bevan, I am a graduate student at Brigham Young University, and I am conducting this research under the supervision of Professor Brown, from the Department of Linguistics. You are being invited to participate in this research study about trademarked brand names and items. I am interested to learn more about how people view, use, and understand trademarked brand names. Being in this study is optional.

If you choose to be in the study, you will be asked to complete a survey that should take approximately 10-15 minutes of your time.

You can skip questions that you do not want to answer or stop the survey at any time. The survey is anonymous and no one will be able to link your answers back to you. Please do not include your name or other information that could be used to identify you in the survey responses. At the end of the survey, you may choose to enter a drawing to win one of four \$25 Amazon gift cards. Your chances of winning will vary depending on the number of people who enter the drawing, but you are expected to have about a 1% chance to win.

If you have questions about the survey, please contact Richard Bevan at richard.b.bevan@gmail.com. If you have questions or concerns about your rights as a research participant, you can call the BYU Institutional Review Board at 801-422-1461 or irb@byu.edu. Please select one of the answers below to choose whether or not you wish to continue with the survey.

Thank you!

I wish to continue with the survey

I do not wish to continue with the survey



Q2. Please provide the word, or words, you would use to refer to the typically white material that is used for packing or as a material used to make cups, bowls, and plates (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q3. Please provide the word, or phrase, you would use to refer to performing a search on an internet browser. (see picture)

Please enter the word, or words, below:

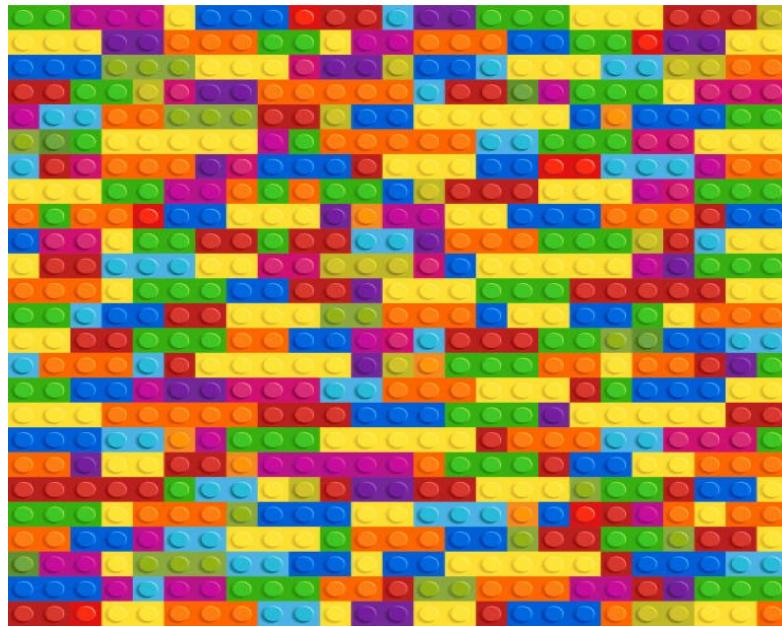
I do not have a word for this



Q4. Please provide the word, or words, you would use to refer to the plastic-like material used for food packaging (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q5. Please provide the word, or words, you would use to refer to the small, plastic building block toys that are typically made for children (see picture).

Please enter the word, or words, below:

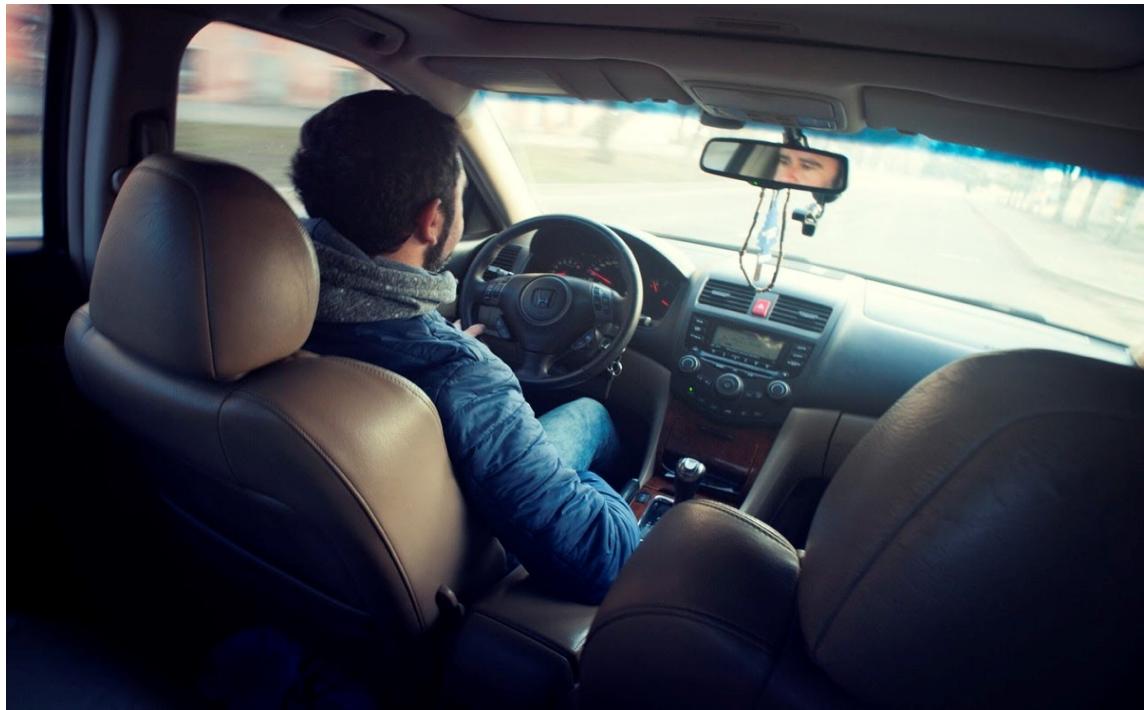
I do not have a word for this



Q6. Please provide the word, or words, you would use to refer to a flavored, carbonated beverage (see picture).

Please enter the word, or words, below:

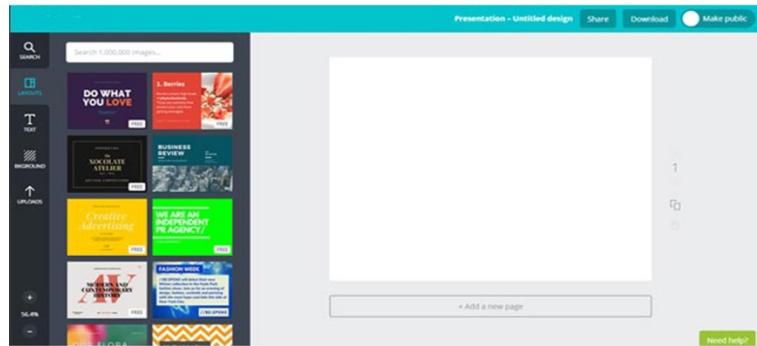
I do not have a word for this



Q7. Please provide the word, or words, you would use to refer to a taxi-like service where you use an app on your phone to request a ride somewhere (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q8. Please provide the word, or words, you would use to refer to a computer program used to make digital slideshow presentations (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q9. Please provide the word, or words, you would use to refer to the disposable object people typically use to wipe or blow their nose (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q10. Please provide the word, or words, you would use to refer to an electronic device connected to a TV used to play video games (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q11. Please provide the word, or words, you would use to refer to the cooking appliance used to cook food over a long period of time (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q12. Please provide the word, or words, you would use to refer to the individual typically hired to help people buy and sell homes (see picture)

Please enter the word, or words, below:

I do not have a word for this



Q13. Please provide the word, or words, you would use to refer to the object that videos could be recorded on and watched with a VCR (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q14. Please provide the word, or words, you would use to refer to the action of digitally altering photos (see picture)

Please enter the word, or words, below:

I do not have a word for this



Q15. Please provide the word, or words, you would use to refer to the over-the-counter drug that is typically white powdery and used to reduce pain (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q16. Please provide the word, or words, you would use to refer to the machine used to provide news anchors or public speakers the written text of what to say (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q17. Please provide the word, or words, you would use to refer to the place where people can go to pay money to wash and dry their clothes (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q18. Please provide the word, or words, you would use to refer to the outdoor equipment typically used by kids or gymnasts to jump around and perform acrobatics (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q19. Please provide the word, or words, you would use to refer to the type of cell phone that is able to be folded but not able to run applications or access Internet (see picture).

Please enter the word, or words, below:

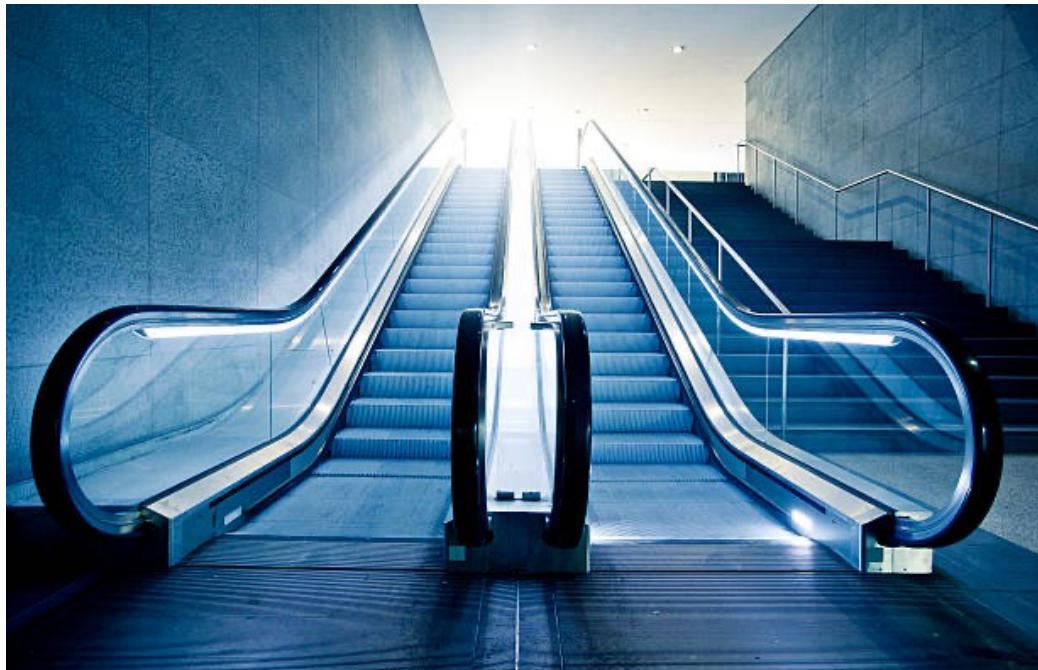
I do not have a word for this



Q20. Please provide the word, or words, you would use to refer to the item typically used to cover cuts and scrapes that is disposable and is meant to stick to skin (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q21. Please provide the word, or words, you would use to refer to the stair-like machinery typically found in malls or parking garages and uses a motor to move the stairs up or down (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q22. Please provide the word, or words, you would use to refer to the plastic-like flooring that is used as a substitute to tile or wood floors (see picture)

Please enter the word, or words, below:

I do not have a word for this



Q23. Please provide the word, or words, you would use to refer to the gasoline-like fuel that is highly flammable and typically used for camping or emergency equipment like lanterns and stoves (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q24. Please provide the word, or words, you would use to refer to video conference computer software (see picture).

Please enter the word, or words, below:

I do not have a word for this



Q25. Please provide the word, or words, you would use to refer to the machine used to make photocopies (see picture).

Please enter the word, or words, below:

I do not have a word for this

Q26. What is your age?

Under 18

18 - 24

25 - 34

35 - 44

45 - 54

55 - 64

Older than 65

Q27. What is your gender?

Male

Female

Non-binary or third gender

I prefer not to answer

Q28. Which country did you grow up in?

Q29. Which state did you grow up in?

Q30. Which country do you currently live in?

Q31. Which state do you currently live in?

Q32. This concludes the survey questions. Thank you very much for your participation!

Would you like to enter a drawing for a chance to win one of four \$25 Amazon gift cards? You will need to provide your name and email address

Yes, I wish to enter the drawing

No, I do not wish to enter the drawing

Q33. In order to maintain the privacy and confidentiality of your responses from this survey, please follow the link below. This will take you to a separate survey that will ask you for your email address. You will only be contacted if you win a gift card.

Amazon Gift Card Drawing [LINK](#)

References

Bellifemine, S. S. (1984). Primary Significance: Proving the Consumer's Perception. *Seton Hall Law Review*, 14(2), 315–355.

Clankie, S. M. (2000). Genericization: A Theory of Semantic Broadening in the Marketplace. *Northern Review*.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>

Collins Sons & Co. Ltd. (1986). *Collins English Dictionary*.

Coverdale, J. F. (1984). Trademarks and Generic Words: An Effect-on-Competition Test. *The University of Chicago Law Review*, 51(3), 868. <https://doi.org/10.2307/1599488>

Corpus of Founding Era American English (COFEA). (2018-). Available at

Cunningham, C. D., & Egbert, J. (2019). *Scientific Methods for Analyzing Original Meaning: Corpus linguistics and the emoluments clauses*. 17.

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.

Davies, M. (2010). *The Corpus of Historical American English (COHA)*. Available online at <https://www.english-corpora.org/coha/>.

Denroche, C. (2014). *Metonymy and language: A new theory of linguistic processing*. Routledge.

Fillmore, C. J., & Atkins, T. S. (1994). Starting where the Dictionaries Stop: The Challenge of Corpus Lexicography. In *Computational Approaches to the Lexicon* (pp. 349–393).

Garceau, M. E. (2020). *Gender Vs. Sex: Defining Meaning in a Modern World through use of Corpora and Semantic Surveys*. Brigham Young University.

Gibson, E., Piantadosi, S., & Fedorenko, K. 2011. Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments. *Language and Linguistics Compass*. 509-524

Gilquin, G., & Gries, S. Th. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, 5(1), 1–26.

Goldfarb, N. (2017). A lawyer's introduction to meaning in the framework of Corpus linguistics. *Brigham Young University Law Review*, 1359–1416.

Gries, S. T. (2009). table.plotter 0.97. A function for R

Hoopes, N. A. (2019). Reclaiming the primary significance test: Dictionaries, corpus linguistics, and trademark genericide. *Tulsa Law Review*, 54(3), 407–440.

Hughes, E. C. (2018). A search by any other name: Google, genericism, and primary significance. *American University Business Law Review*, 7(2), 269–296.

In, S. (2002). Death of a trademark: Genericide in the digital age. *The Review of Litigation*, 21(1), 159–189.

Kövecses, Z., & Radden, G. (1998). Metonymy: Developing a cognitive linguistic view. *Cognitive Linguistics*, 9(1), 37–78. <https://doi.org/10.1515/cogl.1998.9.1.37>

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago press.

Lalonde, A. G., Green, & Gilson, J. (2007). *Gilson on trademarks*. LexisNexis, Matthew Bender.

Lee, T. R., & Mouritsen, S. C. (2018). Judging ordinary meaning. *Yale Law Journal*, 127(4), 788–879. <https://doi.org/10.2139/ssrn.2937468>

List of generic and genericized trademarks. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_generic_and_genericized_trademarks&oldid=1013931263

Longman Group. (1988). *Longman dictionary of the English language*.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 276–282.

Mouritsen, S. C. (2010). The dictionary is not a fortress: Definitional fallacies and a corpus-based approach to plain meaning. *Brigham Young University Law Review*, 2010(5), 1915–1980.

Nerlich, B., & Clarke, D. D. 1992. Outline of a Model for Semantic Change. In Kellermann and Morrissey, 125-141.

Phillips, J. C., Lee, T. R., & Ortner, D. M. (2016). Corpus linguistics & original public meaning: A new tool to make originalism more empirical. *The Yale Law Journal*, 21–32.

Phillips, J. C., & White, S. (2018). The meaning of the three emoluments clauses in the U.S. Constitution: A corpus linguistic analysis of American English from 1760-1799. *South Texas Law Review*, 59, 181–236.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Radden, G. (2005). The ubiquity of metonymy. *Cognitive and Discourse Approaches to Metaphor and Metonymy*, 11–28.

Seto, K. (1999). Distinguishing Metonymy from Synecdoche. *Metonymy in Language and Thought*, 91-120.

Solan, L. M. (2001). Why laws work pretty well, but not great: Words and rules in legal interpretation. *Law & Social Inquiry*, 26, 243–270.

The Cambridge Online Survey of World Englishes. (n.d.) Retrieved November 22, 2021, from http://www.tekstlab.uio.no/cambridge_survey/.

Traugott, E. C., & Dasher, R. B. (2001). *Regularity in semantic change*. Cambridge University Press.

Vander Klok, J. (2014). On the use of questionnaires in semantic fieldwork: A case study in modality. *Proceedings of Language Documentation and Linguistic Theory*, 4, 1–11.

W. & R. Chambers Ltd. (1983). *Chambers 20th Century Dictionary*.