

# The Trademark Reporter®



## The Law Journal of the International Trademark Association

Corpora in the Courts: Using Textual Data to Gauge Genericness and Trademark Validity

*Quentin J. Ullrich*

Commentary: Co-Branding with Influencers Is in Fashion and No Longer a Trademark Faux Pas

*Jessica Elliott Cardon*

Book Review (in Spanish): *La Defensa Judicial del Derecho de Marca*.

Diego Chijane.

*Elisabeth Kasznar Fekete*

Book Review (in English): *La Defensa Judicial del Derecho de Marca (Judicial Protection of Trademark Rights)*. Diego Chijane.

*Elisabeth Kasznar Fekete*

Book Review: *Intellectual Property Jurisdiction Strategies: Where to Litigate Unitary Rights vs National Rights in the EU*. Torsten Bjørn Larsen.

*Alfred C. Frawley*

# INTERNATIONAL TRADEMARK ASSOCIATION

*Powerful Network Powerful Brands*

655 Third Avenue, New York, NY 10017-5646

Telephone: +1 (212) 642-1733

email: [wknox@inta.org](mailto:wknox@inta.org)

Facsimile: +1 (212) 768-7796

## OFFICERS OF THE ASSOCIATION

TISH L. BERARD ..... *President*  
DAVID LOSSIGNOL ..... *President Elect*  
AYALA DEUTSCH ..... *Vice President*  
TIKI DARE ..... *Vice President*  
ZEEGER VINK ..... *Treasurer*  
JOMARIE B. FREDERICKS ..... *Secretary*  
ANNA CARBONI ..... *Counsel*  
ETIENNE SANZ DE ACEDO ..... *Chief Executive Officer*

## *The Trademark Reporter* Committee

### EDITORIAL BOARD

#### EDITOR-IN-CHIEF, CHAIR

JESSICA ELLIOTT CARDON

#### STAFF EDITOR-IN-CHIEF

WILLARD KNOX

#### Senior Editors

GLENN MITCHELL

RAFFI V. ZEROUNIAN

FABRIZIO MIAZZETTO

PAMELA S. CHESTEK

BRYAN K. WHEELLOCK

ELISABETH KASZLAR FEKETE

ANDREW J. GRAY IV

#### Senior Staff Editor

BEVERLY HARRIS

#### Editors

TARA AARON

GUY HEATH

MARCOS MERCADO

GIULIA ABBAGNALE

JANET L. HOFFMAN

GEORGE W. MOXON II

BRUCE W. BABER

BARRY R. HORWITZ

AMANDA NYE

JOLIEN BARBIER

BRUCE ISAACSON

BRANDON M. RESS

MARTIN J. BERAN

E. DEBORAH JAY

MICHEL RIJSDIJK

DANIEL R. BERESKIN

DANIELLE JOHNSON

JENNIFER S. SICKLER

SHELDON BURSHTEN

SIEGRUN D. KANE

RINITA SIRCAR

ROBERT H. CAMERON

PETER J. KAROL

RANDEL S. SPRINGER

SUJATA CHAUDHRI

J. MICHAEL KEYES

SARA SULEIMAN

JACQUELINE CHORN

REMCO KLÖTERS

PAUL TACKABERRY

THEODORE H. DAVIS JR.

JENNIFER KWON

IFEYINWA A. UFONDU

ANNE DESMOUSSEAUX

SCOTT J. LEBSON

MARTIN VIEFHUES

MEGHAN DILLON

JOHN LINNEKER

VERENA VON BOMHARD

KAREN L. ELBURG

NELS T. LIPPERT

JEFFREY A. WAKOLBINGER

MATTHEW EZELL

JOEL G. MACMULL

NADALINE WEBSTER

DÉSIRÉE FIELDS

DOUGLAS N. MASTERS

RITA WEEKS

ALFRED C. FRAWLEY

J. DAVID MAYBERRY

JORDAN S. WEINSTEIN

LESLEY MCCALL GROSSBERG

BRYCE J. MAYNARD

JOHN L. WELCH

MICHAEL HANDLER

J. THOMAS MCCARTHY

JIAN XU

NATHAN HARRIS

JOSEPH S. YANG

#### Advisory Board

MILES J. ALEXANDER

ROBERT M. KUNSTADT

PASQUALE A. RAZZANO

WILLIAM M. BORCHARD

THEODORE C. MAX

SUSAN REISS

CLIFFORD W. BROWNING

KATHLEEN E. MCCARTHY

PIER LUIGI RONCAGLIA

LANNING G. BRYER

JONATHAN MOSKIN

HOWARD J. SHIRE

SANDRA EDELMAN

VINCENT N. PALLADINO

JERRE B. SWANN, SR.

ANTHONY L. FLETCHER

JOHN B. PEGRAM

STEVEN M. WEINBERG

ROBERT L. RASKOPF

The views expressed in *The Trademark Reporter* are those of the individual authors and do not necessarily reflect those of INTA.

*The Trademark Reporter* (ISSN 0041-056X) is published electronically six times a year by the International Trademark Association, 655 Third Avenue, New York, NY 10017-5646 USA. INTA, the INTA logo, INTERNATIONAL TRADEMARK ASSOCIATION, POWERFUL NETWORK POWERFUL BRANDS, THE TRADEMARK REPORTER, and [inta.org](http://inta.org) are trademarks, service marks, and/or registered trademarks of the International Trademark Association in the United States and certain other jurisdictions.

# The Trademark Reporter®

## CORPORA IN THE COURTS: USING TEXTUAL DATA TO GAUGE GENERICNESS AND TRADEMARK VALIDITY

*By Quentin J. Ullrich\**

### Table of Contents

I.	Introduction.....	991
II.	Current Methods of Measuring Genericness.....	994
	A. Consumer Surveys .....	995
	1. The Purchase Encounter Survey.....	997
	2. The Classification Survey .....	999
	3. Limitations of Surveys .....	1001
	B. Record Evidence .....	1007
	1. Unsound Record Evidence .....	1008
	2. Limitations of Record Evidence.....	1009
III.	Corpus Linguistics and Corpora .....	1011
	A. Corpora in the Courts .....	1013
	1. Corpus-Based Analyses of Plain Meaning.....	1013
	2. Previous Corpus-Based Analyses in Genericness Litigation .....	1016
	B. The Case for Corpora .....	1021
	1. Large Sample Sizes .....	1022
	2. Replicability and Limited Human Element.....	1023
	3. Cost Effectiveness .....	1024
	4. Probative Value of Usage Over Elicited Understanding.....	1025
IV.	Methodology of Corpus Analysis in Genericness Litigation .....	1027
	A. Corpus Selection.....	1028

---

\* Quentin J. Ullrich studied linguistics at the College of William & Mary in Virginia, where he received Honors for his thesis, from which this article is derived. He lives and works in New York, New York, and can be reached at [qjullrich@email.wm.edu](mailto:qjullrich@email.wm.edu). Many thanks to Professor Laura A. Heymann and Professor Daniel Parker of the College of William & Mary for their helpful guidance. All errors are those of the author, Quentin J. Ullrich.

1. Probative Value of Consumer Language versus Media Language.....	1030
2. Corpus Selection Strategies in Litigation.....	1031
3. When Corpora Cannot Represent the Relevant Consumers.....	1032
B. Capitalization.....	1033
1. Probative Value of Capitalization Rates.....	1033
2. Inclusion of Anomalous Forms in Capitalization Rate.....	1036
3. Manual Calculation of Capitalization Rates .....	1036
4. Interpreting Capitalization Rates .....	1037
5. A Case Study of Capitalization Rates .....	1040
6. Limitations of the Capitalization Rate .....	1042
C. Common Name Frequencies.....	1045
D. Collocation Analysis.....	1047
1. Measurements of Collocation .....	1048
2. Case Studies of Collocation.....	1051
3. Limitations of the Collocation Method.....	1054
V. Conclusion .....	1054

*But the Idols of the Market Place are the most troublesome of all—idols which have crept into the understanding through the alliances of words and names. For men believe that their reason governs words; but it is also true that words react on the understanding; and this it is that has rendered philosophy and the sciences sophistical and inactive. Now words, being commonly framed and applied according to the capacity of the vulgar, follow those lines of division which are most obvious to the vulgar understanding. And whenever an understanding of greater acuteness or a more diligent observation would alter those lines to suit the true divisions of nature, words stand in the way and resist the change.*

—Francis Bacon, *Novum Organum Scientiarum*

## I. INTRODUCTION

In 1620, Francis Bacon held a frustration shared by today’s trademark holders—words can be appropriated by the public to take on meanings unintended by their originators. The modern-day Idols of the Marketplace are genericized former trademarks—words such as “yo-yo,”<sup>1</sup> “trampoline,”<sup>2</sup> and “escalator”<sup>3</sup> that were appropriated by the public as product categories and lost their statuses as U.S. trademarks.<sup>4</sup> This phenomenon has become known as “genericide,” and trademark owners, knowing all too well the perils of losing exclusive use of their valuable brand names, take significant measures to protect them. Such measures include lengthy disclaimers on the use of their trademarks,<sup>5</sup> petitioning dictionary editors to define their marks only as trademarks,<sup>6</sup> running anti-

<sup>1</sup> Donald F. Duncan, Inc. v. Royal Tops Mfg. Co., 343 F.2d 655 (7th Cir. 1965) (finding “yo-yo” generic for “spinning top”).

<sup>2</sup> Nissen Trampoline Co. v. Am. Trampoline Co., 193 F. Supp. 745 (S.D. Iowa 1961) (finding “trampoline” to have become generic for rebound-tumbling gymnastic equipment).

<sup>3</sup> Houghton Elevator Co. v. Seeburger, 85 U.S.P.Q. (BNA) 80 (Comm’r Pats. & Trademarks 1950) (finding “escalator” generic for moving staircases).

<sup>4</sup> This article is U.S. and English-language-focused; however, the methods discussed herein may similarly be applicable to other legal systems and other languages.

<sup>5</sup> *E.g.*, *Guidelines for Using Apple Trademarks and Copyrights*, Apple Legal (2017), [apple.com/legal/intellectual-property/guidelinesfor3rdparties.html](http://apple.com/legal/intellectual-property/guidelinesfor3rdparties.html) (featuring disclaimers regarding “acceptable” and “not acceptable” uses of its trademarks. “Trademarks are adjectives used to modify nouns; the noun is the generic name of a product or service. . . . As adjectives, trademarks may not be used in the plural or possessive form. Correct: I bought two Macintosh computers. Not Correct: I bought two Macintoshes.”).

<sup>6</sup> *See generally* Laura A. Heymann, *The Grammar of Trademarks*, 14 Lewis & Clark L. Rev. 1349 (2010). Of note, in the United States, trademark holders cannot sue dictionary editors, writers, or other consumers for using their marks (either as trademarks or as generic terms), because such usage does not constitute trademark infringement, and because doing so could violate one’s freedom to express an opinion. *See* J. Thomas McCarthy, *McCarthy on Trademarks and Unfair Competition* § 12:28 (Westlaw, 4th ed., 2016). By contrast, in the European Union (EU), dictionary editors can be required to denote a word’s status as a registered trademark in subsequent editions if requested by

genericide advertisement campaigns, using the word “brand” after their marks, and even, as does Apple, preferentially auto-capitalizing their own marks when they are typed on their products.<sup>7</sup> Not all companies may be such careful sentinels of their marks, however, and even when they do engage in trademark policing, sometimes “words stand in the way and resist the change.”<sup>8</sup>

Perhaps to Bacon’s dissatisfaction, U.S. courts look to the “primary significance of the term in the minds of the consuming public”<sup>9</sup> to determine a word’s status as a trademark. In gauging such primary significance, the courts currently give substantial weight to consumer surveys and record evidence of mark usage. Specifically, the consumer survey has been commonly recognized as the preeminent assessment of genericness and is distinguished by the empirical standards to which it is held.<sup>10</sup> Despite surveys’ advantages of direct consumer input, scientific rigor, and well-developed legal precedent, however, they can also be subject to certain shortcomings, namely including expensiveness, limited sample sizes, demand characteristics, and the inability to easily replicate their administration during a proceeding. Contrarily, although record evidence (i.e., documented examples of actual use of a term) can consist of actual consumer use, unaffected by the artificial environment of the survey, such evidence has typically neither been proffered with indication of the representativeness of the sample from which it was drawn nor analyzed with the empiricism with which survey data are.<sup>11</sup>

Given these limitations, this article proposes using corpora, which are electronic databases of natural language, i.e., language appearing organically rather than solicited for purposes of litigation, to measure primary significance. Corpora can consist of

a trademark owner. Regulation (EU) 2017/1001 of the European Parliament and of the Council of 14 June 2017 on the European Union trademark, art. 12.

- <sup>7</sup> For example, when one types “iphone,” a trademark of Apple, on Apple’s Pages word processor, it is automatically changed to “iPhone,” but when one types “iphone” on Microsoft Word, it is not automatically changed to “iPhone.”
- <sup>8</sup> Francis Bacon, *Novum Organum* (1620), in Robert L. Ellis & James Spedding, *The Philosophical Works of Francis Bacon*, at 269 (1905).
- <sup>9</sup> 15 U.S.C. § 1064 (codifying the primary significance test after being used by the court in *Kellogg Co. v. Nat’l Biscuit Co.*, 305 U.S. 111 (1938)).
- <sup>10</sup> Consumer surveys have become, according to McCarthy, “almost de rigeur in litigation over genericness.” (McCarthy, *supra* note 6, § 12:14 (noting that in some instances, a party may be penalized by the court for not introducing into evidence a consumer survey, referring to *Gimix, Inc. v. JS&A Group, Inc.*, 213 U.S.P.Q. 1005 (N.D. Ill. 1982)). Consumer surveys can be administered in many different formats, the most widely recognized of which are the *Thermos* and *Teflon* surveys. McCarthy, *supra* note 6, § 12:14.
- <sup>11</sup> See, e.g., *Trek 2000 Int’l Ltd.*, 97 U.S.P.Q.2d (Each party proffered examples of usage of THUMBDRIVE but did not substantiate their data by indicating it was taken from a representative sample). See *infra* Part II(B) discussing record evidence and its limitations.

books, blogs, social media posts, speech transcripts, and any other source of language use that can be compiled and digitized. Corpora are designed and analyzed in accordance with the principles of corpus linguistics—as with surveys, experts presenting corpus-based evidence should speak to the representativeness of the corpora to the relevant universe of consumers, delineate their empirical methods, and substantiate their analyses with statistical perspective where applicable. Specifically, a corpus-based analysis of a disputed mark should examine the word-initial capitalization rate of the disputed mark in comparison to that of other trademarks and generic terms in the corpus, the frequencies of alternative generic terms and competitors' trademarks, and other linguistic indicators of trademark validity and genericness, including the habitual contexts in which the disputed mark appears.

Simply put, corpora are large compilations of record evidence that can be empirically analyzed. Unlike individual examples of mark usage that may be cherry-picked and unrepresentative of the language of the relevant consuming population, appropriately designed corpora allow for more probative conclusions to be made regarding the public's understanding of the mark in question. Additionally, corpora have distinct advantages over surveys—corpus-based analyses are highly replicable because of the permanence of textual data in comparison to the ephemeral oral survey environment, and they are relatively inexpensive to administer given the ease with which vast quantities of textual data can be harvested from the Internet and already-existent corpora can be analyzed.<sup>12</sup> Most importantly, as corpora consist of naturally occurring language, unaffected by the surveyor, the order of questions, or the respondent's perception of "the right answer," they are uniquely equipped to reflect the actual meaning of disputed marks to consumers.

Part II(A) of this article discusses the methods and limitations of surveys. Part II(B) addresses record evidence, making distinctions between probative and irrelevant forms of record evidence, which inform the proposed corpus methodology. Part III(A) surveys the appearance of corpora in the courts, noting *sua sponte* corpus-based analyses in plain meaning cases and critiquing corpus-based analyses that have been proffered in trademark litigation. Part III(B) argues for the probative value of corpora in comparison to surveys and record evidence, specifically discussing corpus-analyses' large sample sizes, their ease of replication, their cost-effectiveness, and the probative value of usage over elicited

---

<sup>12</sup> E.g., *Corpus of Global Web-Based English*, Brigham Young University (2012–2013), <http://corpus.byu.edu/glowbe/> (a corpus consisting of 1.9 billion words from both blogs and other websites, which can be sorted by country of origin); *Google Books (American) Corpus*, Brigham Young University (1960–2000), <http://googlebooks.byu.edu/x.asp> (a corpus consisting of over 155 billion words from books written between 1960 and 2000).

survey responses. Finally, Part IV delineates an analytical approach for corpus-based analyses of disputed trademarks, discussing how to quantify and interpret capitalization rates, frequencies of related generic and trademarked terms, and the linguistic contexts in which a disputed mark may appear.

## II. CURRENT METHODS OF MEASURING GENERICNESS

The four common sources of evidence used in trademark litigation to capture the elusive concept of genericness are dictionary definitions, testimony of persons in the trade, consumer surveys, and record evidence<sup>13</sup> (competitors' use of the mark, plaintiff's own use of the mark, media usage, and consumer usage).<sup>14</sup> Surveys are currently the most popular form of evidence,<sup>15</sup> generally receiving more credence than other forms of evidence, especially

---

<sup>13</sup> McCarthy, *supra* note 6, § 12:13 (citing, *e.g.*, Dan Robbins & Assocs. v. Questor Corp., 599 F.2d 1009 (C.C.P.A. 1979) (stating that relevant evidence can consist of purchaser testimony, consumer surveys, dictionary listings, trade journals, newspapers, and other publications)). These are not the only sources of evidence proffered in trademark cases, but, as McCarthy notes, they are the most widely used. Another moderately common source of evidence is advertising of the mark, but this article, as well as some courts, do not find it to be particularly probative of the significance of the term in the minds of consumers. *See, e.g.*, Int'l Jensen, Inc. v. Metrosound U.S.A., Inc., 4 F.3d 819 (9th Cir. 1993) (not finding "manufacturer's sales, advertising and promotional activities" to be a "true test" of the secondary meaning of the mark, "Blue Surround."), *quoted in* Lisa L. Ouellette, *The Google Shortcut to Trademark Law*, 102 Cal. L. Rev. 351-407, at 362 n.58 (2014). Among the less conventional sources of evidence that have been presented is an analysis by an expert linguist stating that certain stress patterns in a compound phrase can be indicative of genericness. *See* Stix Prods. v. United Merchs. & Mfrs., 295 F. Supp. 479, 488 (S.D.N.Y. 1968) (not finding the stress pattern test to be probative of genericness in the minds of consumers).

<sup>14</sup> Media usage can include newspaper articles, trade journals, and magazines, among other publications. McCarthy, *supra* note 6, at § 12:13. This article proposes that in addition to these forms, mark usage on blogs, social media, and other sources can also be probative of genericness as long as "generic usage" is rigorously defined and such evidence is proffered using the empiricism of corpus linguistics.

<sup>15</sup> McCarthy, *supra* note 6, at § 12:14.

more than dictionary definitions<sup>16</sup> and testimony of persons in the trade.<sup>17</sup>

### A. Consumer Surveys

Consumer surveys testing for genericness<sup>18</sup> typically fall in two general categories: purchase encounter surveys and classification surveys.<sup>19</sup> Purchase encounter surveys, also known as *Thermos* surveys,<sup>20</sup> ask respondents to answer a series of open-ended questions such as “Can you name any trademarks or brand names that are used on this product?” and “What do you call this type of product?”<sup>21</sup> Classification surveys, also known as *Teflon* surveys,<sup>22</sup> ask respondents to classify words as brand names or common names.<sup>23</sup> To be admissible in court, genericness surveys of any type

---

<sup>16</sup> Dictionary definitions are generally far from probative of whether a disputed term is generic. See McCarthy, *supra* note 6, at § 12:13 (stating that dictionary definitions cannot be conclusive of genericness, “if for no other reason than that this would endow editors of such words with the power to destroy trademarks merely by defining them generically” and citing the Seventh Circuit’s observations that “numerous terms have been found to be generic despite their absence from the dictionary”) (quoting *Liquid Controls Corp. v. Liquid Control Corp.*, 802 F.2d 934, (7th Cir. 1986)). See also Sidney I. Landau, *Dictionaries: The Art and Craft of Lexicography*, 280 (1st ed. 1984) (“The dictionary editor must do battle to include any trademarks, and he is under great pressure to distort the facts of usage by entering all such terms only in capitalized form, even though the record clearly shows they are often written in lower-case letters.”). Dictionaries can be further skewed by the potential overrepresentation of obscure uses. See Ronald R. Butters, *A Linguistic Look at Trademark Dilution*, 24 Santa Clara High Tech. L.J., 507, 509 (“special consideration is given to relatively arcane terms that the general user of a dictionary might nonetheless have occasion to look up”).

<sup>17</sup> The primary downside to such testimony is that witnesses may be subject to bias depending on their relationship to the trademark holder. McCarthy, *supra* note 6, § 12:13.

<sup>18</sup> Surveys are admissible under Federal Rule of Evidence 703. See Fed. R. Evid. 703.

<sup>19</sup> See Itamar Simonson, *Trademark Infringement from the Buyer Perspective: Conceptual Analysis and Measurement Implications*, 13 J. Pub. Pol’y & Mktg. 181, 186 (1994) (broadly categorizing genericness surveys as either classification-based or purchase encounter-based); McCarthy, *supra* note 6, § 12:14 (noting the two most common consumer surveys of genericness, the *Thermos* and *Teflon* surveys).

<sup>20</sup> The purchase encounter-type survey takes its name (“*Thermos* survey”) from the landmark case in which the survey was first introduced—*American Thermos Products Co. v. Aladdin Industries*, 207 F. Supp. 9 (D. Conn. 1962) (relying on purchase encounter survey evidence to find “*thermos*” to be generic).

<sup>21</sup> See generally E. Deborah Jay, *Genericness Surveys in Trademark Disputes: Evolution of Species*, 99 TMR 1118, 1122 (2009).

<sup>22</sup> Classification-type, or *Teflon*, surveys were first introduced in *E. I. DuPont de Nemours & Co. v. Yoshida International, Inc.*, 393 F. Supp. 502 (E.D.N.Y. 1975) (relying upon survey evidence to find defendant had not met its burden in seeking to prove that “TEFLON” is generic).

<sup>23</sup> See Jay, *supra* note 21, at 1122.

must be administered with respect to general principles set forth by the Federal Judicial Center and by precedent.<sup>24</sup>

As with other types of scientific evidence, trademark surveys are admissible and probative to the extent they are administered pursuant to generally accepted scientific and statistical principles.<sup>25</sup> Criteria for surveys include:

- (1) the “universe” was properly defined;
- (2) a representative sample of that universe was selected;
- (3) the questions to be asked of interviewees were framed in a clear, precise and non-leading manner;
- (4) sound interview procedures were followed by competent interviewers who had no knowledge of the litigation or the purpose for which the survey was conducted;
- (5) the data gathered was accurately reported;
- (6) the data was analyzed in accordance with accepted statistical principles; and
- (7) objectivity of the entire process was assured.<sup>26</sup>

That surveys are created and analyzed based upon scientific and statistical principles is one of their primary advantages; unlike record evidence, dictionary definitions, and testimony of individual witnesses, a survey seeks to gauge a mark’s primary significance to the public by directly polling a representative sample, from which an expert witness may draw statistical conclusions about the broader group of consumers.<sup>27</sup> Additionally, one of the advantages of consumer surveys over record evidence, in particular, is the ability for interviewers to communicate directly with respondents to clarify their understanding of the mark of interest.<sup>28</sup> Unlike the analysis of static pieces of record evidence or of corpora, the nature of some survey formats can allow for the interviewer to ask follow-

---

<sup>24</sup> *E.g.*, Federal Judicial Center, Manual for Complex Litigation 116 (5th ed. 1981) [hereinafter “Complex Litigation”], cited by Toys “R” US, Inc. v. Canarsie Kiddie Shop, Inc., 559 F. Supp. 1189 (E.D.N.Y. 1983); Shari Seidman Diamond, *Reference Guide on Survey Research*, in Reference Manual on Scientific Evidence 229 (2d ed. 2000).

<sup>25</sup> Typically, methodological flaws in a trademark survey affect only the survey’s weight and probative value, as opposed to its admissibility, unless such defects are so systemic and serious that they render the survey completely unreliable. *See, e.g.*, Lifeguard Licensing Corp. v. Kozak, 2017 WL 908199, at \*5 (S.D.N.Y. 2017), order *aff’d* 2017 WL 3142072 (S.D.N.Y. 2017) (Certain technical flaws in genericness surveys will not preclude them from being admitted as evidence). *See also* McCarthy, *supra* note 6, at § 32:170.

<sup>26</sup> Toys “R” US, 559 F. Supp. at 1205 (citing Complex Litigation, *supra* note 24, at 116).

<sup>27</sup> Of note, some of the standards to which surveys are held are not dissimilar to those that this article proposes corpus-based analyses should be held. However, this article argues that these principles are easier to follow under a corpus-based approach, as they do not require the administration of questions, and data can be gathered and organized more quickly and more transparently than can survey responses.

<sup>28</sup> This applies only to in-person or to telephone surveys.

up questions or to address perceived misinterpretations of the questions on behalf of respondents.

Despite these benefits, however, surveys of each type are subject to specific shortcomings that can negatively affect their probative value in court, if not their admissibility in the first place. In addition to specific issues associated with the *Thermos* and *Teflon* formats, surveys can be subject to shortcomings, including limited sample sizes, unrepresentative samples, improper environments of administration, measurement errors, expensiveness, and demand characteristics.

### 1. The Purchase Encounter Survey

The purchase encounter, or *Thermos*, survey places the respondent in a hypothetical purchasing encounter and seeks to elicit from the respondent how he or she would ask for a product. A *Thermos* survey typically first asks respondents about their familiarity with a type of product (“are you familiar with the type of product into which you would blow your nose?”).<sup>29</sup> Subsequently, it might ask about the type of store they would seek to purchase the item.<sup>30</sup> Next, it will ask, “if you were going to buy the type of product that you use to blow your nose, what would you ask for—that is, what would you tell the clerk you wanted? Can you think of any other words that you would use to ask for products you use to blow your nose?”<sup>31</sup> The purchase encounter survey may also ask, “Can you think of any trademarks or brand names that are associated with the type of product you use to blow your nose?”<sup>32</sup> The primary advantage of the *Thermos* survey is that, unlike a *Teflon* survey, its questions are open-ended, thereby reducing the degree to which demand characteristics may take effect.<sup>33</sup> Additionally, the purchase encounter survey puts the respondents in a scenario in which they are shoppers, which is the context with which measurements of genericness are most concerned.<sup>34</sup>

The *Thermos* survey’s open-ended questions, however, are also its most significant shortcoming. Open-ended questions cause the survey to be laborious to administer; since respondents answer questions in their own words, survey administrators must review and categorize answers on their own. This is especially the case when respondents give multiple answers that must be analyzed and

---

<sup>29</sup> See McCarthy, *supra* note 6, § 12:15.

<sup>30</sup> *Id.*

<sup>31</sup> See generally Simonson, *supra* note 19, at 186.

<sup>32</sup> See Jay, *supra* note 21, at 1122.

<sup>33</sup> “Demand characteristics” refers to the concept that respondents may interpret the survey’s purpose in such a way that they believe they are expected to provide particular answers. See Simonson, *supra* note 19, at 183.

<sup>34</sup> *Id.*

reported, creating the potential for more observational errors.<sup>35</sup> The open-ended format of the *Thermos* survey also tends to cause it to overestimate generic usage, especially when the mark in question is famous or is used as a generic term frequently.<sup>36</sup> For example, in response to “If you wanted to purchase the type of product into which you blow your nose, what would you tell the clerk you wanted?,” respondents may answer “KLEENEX,” even though they may still understand and use the word as a trademark referring to a specific brand of tissue.<sup>37</sup> It is for this reason that courts have discounted the probative value of the *Thermos* survey in some instances.<sup>38</sup>

Another downside to the *Thermos* format is that even though frequent usage of a word in response to the “What would you ask the clerk for?” question may indicate genericness, the lack of usage of a word in response to this question does not necessarily mean a word is not generic.<sup>39</sup> For example, even if all respondents say they would only ask the store clerk for either a “soda” or a “soft drink,” the word “pop” is still generic for carbonated beverages. Thus, although the *Thermos*-type survey is less prone to demand characteristics than is a *Teflon*-type survey, its open-ended questions may not be regarded as direct enough to ascertain the meaning of a word in the mind of the respondent. For example, asking a respondent “If you were going to buy one of these containers tomorrow—that is, the type that keeps food and beverages hot or cold . . . What would you ask for—that is, what would you tell the clerk you wanted?”<sup>40</sup> is not as direct as simply asking respondents to categorize a word as a generic name or as a brand name. Although this shortcoming can be hedged by asking a follow-up question such as, “Are there any other terms you would use to describe such a container?”<sup>41</sup> it is still the case that the lack of a certain term in the responses does not necessarily imply that such a term is not generic. Likewise, as discussed above, the

---

<sup>35</sup> See *Tea Bd. of India v. Republic of Tea, Inc.*, 80 U.S.P.Q.2d (BNA) 1881 (T.T.A.B. 2006) (finding the coding of survey responses by applicant’s expert witness to have “numerous problems”), cited by Jay, *supra* note 21, at 1130 n.43.

<sup>36</sup> Simonson, *supra* note 19, at 186.

<sup>37</sup> *Id.*

<sup>38</sup> *E.g.*, *E. I. DuPont de Nemours & Co. v. Yoshida Int’l, Inc.*, 393 F. Supp. 502 (E.D.N.Y. 1975) (finding defendant’s and plaintiff’s *Thermos*-type surveys to be ambiguous, because “respondents were, by the design of the questions, more often than not focusing on supplying the inquirer a ‘name’, without regard to whether the principal significance of the name supplied was its indication of the nature or class of an article, rather than an indication of its origin.”) (quoted in Jay, *supra* note 21, at 1126 n.25).

<sup>39</sup> See Jay, *supra* note 21, at 1127.

<sup>40</sup> Vincent N. Palladino, *Assessing Trademark Significance: Genericness, Secondary Meaning and Surveys*, 92 TMR 857, 874 (2002).

<sup>41</sup> *Id.*

presence of a given term in the responses does not imply that the mark is necessarily generic.

## 2. The Classification Survey

The classification, or *Teflon*, survey, which has gained wider respect from the courts,<sup>42</sup> first teaches respondents about the difference between trademarks (or brand names) and generic terms and then tests their ability to differentiate between the two by asking them to categorize certain words, one of which is the mark of interest in the dispute.<sup>43</sup> The primary advantage of the *Teflon* survey is that it approaches the question of genericness very directly by explicitly asking the respondent to categorize words as brand names or as generic names; additionally, it tests<sup>44</sup> the respondents' ability to correctly engage in such classification by having them categorize other words that are not disputed in the given case.<sup>45</sup>

One of the more significant issues with the *Teflon* survey is the potential effect of demand characteristics.<sup>46</sup> Respondents may answer based on what they think a word is *supposed* to mean, or what they perceive to be the interviewer's intended meaning of a word, rather than based on how they primarily use or understand the word. Additionally, regardless of the effects of the interviewer, respondents' answers may still conflict with their habitual usage of the disputed mark. This theory is supported by a study conducted by Itamar Simonson, in which 69% of respondents who indicated they used a term mainly as a common name classified the term as a brand name in a follow-up classification task.<sup>47</sup> In comparison to the

---

<sup>42</sup> See McCarthy, *supra* note 6, at § 12:16. See, e.g., *Booking.com B.V. v. Matal*, 278 F. Supp. 3d 891, 923 (E.D. Va., 2017) (finding “a strong public perception that BOOKING.COM is a brand identifier, as evidenced by the Teflon survey”); *Elliott v. Google, Inc.*, 860 F.3d 1151, 1163 (9th Cir. 2017) (Watford, J., concurring, “In support of its motion for summary judgment, Google produced overwhelming evidence that the public primarily understands the word ‘Google’ as a trademark for its own search engine, not the name for search engines generally. In Google’s consumer survey, 93% of respondents identified ‘Google’ as a brand name, rather than a common name for search engines. In every dictionary in the record, the first entry for ‘Google’ or ‘google’ refers to Google’s search engine.”), *cert. denied* 138 S. Ct. 362 (U.S. 2017).

<sup>43</sup> See McCarthy, *supra* note 6, at § 12:16.

<sup>44</sup> *Teflon*-style surveys have been discounted for simply asking whether the respondent understands the difference between brand names and common names as opposed to actually testing the respondent's understanding of the difference. See, e.g., *Zimmerman v. Nat'l Ass'n of Realtors*, 70 U.S.P.Q.2d 1425 (T.T.A.B. 2004).

<sup>45</sup> See McCarthy, *supra* note 6, § 12:16.

<sup>46</sup> See Hal Poret, *A Comparative Empirical Analysis of Online Versus Mall and Phone Methodologies for Trademark Surveys*, 100 TMR 756, 763 (2010); Simonson, *supra* note 19, at 194.

<sup>47</sup> Simonson, *supra* note 19, at 194. Such a result, indicating majority generic usage but trademark recognition, is likely to support a finding of trademark validity. Although examining only usage, a properly conducted corpus-based analysis considers

*Thermos* survey, the *Teflon* survey may have a tendency to overestimate trademark usage (and underestimate generic usage); the latter is also reflected in Simonson's study, which indicated that respondents of a *Thermos*-type survey categorized marks as generic 83% of the time, while respondents of a *Teflon*-type survey categorized the same marks as generic only 23% of the time.<sup>48</sup>

These results imply that consumers can use a mark generically while still maintaining the ability to recognize the word's function as a trademark in the relevant contexts. However, they also signpost the challenge of reconciling "primary significance" and "majority usage" when interpreting survey results. For example, what if the public uses a mark generically 90% of the time while still capitalizing the mark in their writing because of their recognition of the mark's ability to function as a trademark? A *Teflon* survey forcing respondents to make a binary classification of the mark may be ineffectual, neither fully capturing consumers' majority generic usage nor their recognition of the mark's trademark function, reflected by their capitalization of the term even in generic contexts.<sup>49</sup> Both this scenario and Simonson's findings reveal surveys' shortcomings with respect to comprehensively presenting the public's understanding of a disputed mark.

In addition to demand characteristics, the *Teflon* survey is also prone to circular definitions of "brand name" and "generic"<sup>50</sup> as well as the potential for respondents to guess if they do not have a strong understanding of the word in question.<sup>51</sup> Additionally, the *Teflon* survey's test of respondents' understanding of the difference between generic terms and trademarks may be biased. For example,

---

capitalization rates and other factors that can indicate trademark recognition despite frequent generic use. *See infra* Part IV(B)(1).

<sup>48</sup> *Id.* at 192. While parties to genericness cases could administer both a *Teflon* and *Thermos* survey in order to cover their bases, doing so would be quite expensive; further, any discrepancy between the two surveys' conclusions could raise questions as to the legitimacy of either survey in the first place, as Simonson's study has.

<sup>49</sup> *See infra* Part IV(B), discussing the analysis of capitalization rates.

<sup>50</sup> *See Classic Foods Int'l Corp. v. Kettle Foods, Inc.*, 468 F. Supp. 2d 1181 (C.D. Cal. 2007) (criticizing the defendant's survey for providing a circular definition of "brand name") (cited by Jay, *supra* note 21, at 1133).

<sup>51</sup> The original *Teflon* survey did not include as a possible answer "I don't know." Scholars have, as a result, recommended that *Teflon*-type surveys feature such a response to reduce instances of guessing. *E.g.*, Jay, *supra* note 21, at 1138. However, while the inclusion of such an option in conjunction with the use of test questions may reduce instances of guessing, it does not necessarily eliminate guessing in its entirety; as artificial language environments, all surveys are prone to the risk that respondents may guess on specific questions for one reason or no reason at all. Furthermore, it may be unclear to respondents whether the "I don't know" option should be selected if they do not know whether the word is a brand or if they are not familiar with the term in the first place.

if the test features only fanciful<sup>52</sup> trademarks such as GOOGLE, VIAGRA, and EXXON, the respondent may gain the false impression that all trademarks must be fanciful names. Such an impression could influence how the respondent categorizes the mark in question if it is not also a fanciful mark.

### 3. Limitations of Surveys

Certain problems affect both types of surveys and can either frustrate their administration or diminish their probative value in court.

One of the most critical aspects of surveying is determining the universe, or the group of people whose “perceptions and state of mind are relevant to the issues in the case.”<sup>53</sup> The correct universe consists of potential purchasers of the product whose trademark is being disputed.<sup>54</sup> From that universe, a representative sample is selected and surveyed.<sup>55</sup> There are two significant errors that can be made in this initial step of surveying: the determination of an irrelevant universe of people<sup>56</sup> and the selection of a sample that is unrepresentative of the universe.<sup>57</sup> For a sample to reflect its target universe, it is often necessary to include a qualifying question in the survey. In *Windsurfing International, Inc. v. Fred Ostermann GMBH*, for example, the defendant’s survey measuring the genericness of WINDSURFER featured the following request: “Now, please tell me what kind of products these are,” to which only 58% of the 801 respondents gave answers exhibiting awareness of what a “windsurfer” was.<sup>58</sup> Thus, in circumstances where such qualifying questions are required, coverage errors are avoided, but at the expense of the time and money wasted on surveying the wrong consumers. As other scholars have noted, this could have been avoided in the “WINDSURFER” survey by defining a narrower

---

<sup>52</sup> As opposed to arbitrary, suggestive, or descriptive. See *Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4 (2d Cir. 1976) (setting forth what has become known as the Abercrombie Spectrum of Distinctiveness).

<sup>53</sup> McCarthy, *supra* note 6, at § 32:159.

<sup>54</sup> *Id.*

<sup>55</sup> *Id.*

<sup>56</sup> *E.g.*, *Ty Inc. v. Softbelly’s, Inc.*, 353 F.3d 528 (7th Cir. 2003) (finding 13-18-year-old girls to be an “arbitrary subset of consumers of beanbag stuff animals,” thus rendering “worthless” a survey concerning the mark BEANIES).

<sup>57</sup> *E.g.*, *Water Pik, Inc. v. Med-Sys., Inc.*, 2012 U.S. Dist. LEXIS 81592, at \*31 (D. Colo. June 13, 2012) (finding defendant’s survey is flawed in that it “has an overinclusive survey universe, an unrepresentative sample, replicates unrepresentative market conditions, and contains suggestive questions”). Inconsistencies between a sample and its intended population are called coverage errors. McCarthy, *supra* note 6, at § 32:159 (citing Fed. Judicial Ctr. Reference Manual on Scientific Evidence, 377-78 (3d ed. 2011)).

<sup>58</sup> 613 F. Supp. 933 (S.D.N.Y. 1985), *vacated*, 828 F.2d 755 (Fed. Cir. 1987), *quoted in* Jay, *supra* note 21, at 1135 n.65.

universe than adults between the ages of sixteen and forty in the United States.<sup>59</sup> However, no matter how carefully administered, all surveys are subject to the coverage error created by participant self-selection, namely, that survey respondents are people willing to spend their time taking a survey for free or who are incentivized to do so by the receipt or prospective receipt of a gift.<sup>60</sup> While ascertaining the relevant universe can be one of the more difficult aspects of administering a consumer survey, the ability to do so is also one of the survey's great advantages. Defining a specific population allows survey administrators to ensure that respondents are truly potential consumers of the trademarked product that is of concern to the dispute.<sup>61</sup>

Another significant issue affecting consumer surveys is that of the environment of administration. Surveys are generally administered in the following environments: malls, research facilities, over the telephone, and through the Internet.<sup>62</sup> Although each of these environments can be prone to general shortcomings of survey administration, each environment also poses unique problems to the surveyor.

Mall-intercept surveying involves a team of interviewers who intercept persons at shopping malls based on characteristics such as gender and age until a quota, such as fifty men and fifty women, is filled.<sup>63</sup> Thus, technically, mall surveys can be used only to make approximations about the population of mall shoppers. Further, the quota system of mall surveying does not allow an expert to make accurate projections of the universe at large, since respondents are sometimes selected according to their physical appearances rather than randomly.<sup>64</sup> More problematic, however, is when mall-intercept interviewers themselves are non-expert college students who select friends as respondents in order to fulfill their employment requirements.<sup>65</sup>

---

<sup>59</sup> Jay, *supra* note 21, at 1135.

<sup>60</sup> See Robert H. Thornburg, *Trademark Surveys: Development of Computer-Based Survey Methods*, 4 J. Marshall Rev. Intell. Prop. L. 91, 95 (2004) (noting that self-selection may be a problem with mall-intercept style surveys specifically) (citing *Tyco Indus. v. Lego Sys.*, 1987 U.S. Dist. LEXIS 13193, at \*30 (D.N.J. Aug. 24, 1987)). Corpora are not immune to these issues either. See *infra* Part IV(A) discussing representativeness in corpus selection.

<sup>61</sup> With corpora, on the other hand, there are certain products for which it may be difficult to find instances of language use by potential consumers, even if such words are frequently used. See *infra* Part IV(A)(3).

<sup>62</sup> See McCarthy, *supra* note 6, §§ 32:159-166; Thornburg, *supra* note 60, at 716.

<sup>63</sup> See McCarthy, *supra* note 6, at § 32:165.

<sup>64</sup> *Id.* (stating that for this reason, mall intercept surveys are classified as nonprobability sampling surveys, but noting that most surveys are nonprobability surveys, with the exception of telephone surveys, which have their own downsides). See also Thornburg, *supra* note 60, at 95.

<sup>65</sup> Thornburg, *supra* note 60, at 95.

Central location surveys, whereby respondents are invited to a research facility to take the survey, are an improvement upon mall-intercept surveys, because they allow administrators to target a more specific universe of consumers than mall-shoppers.<sup>66</sup> The downside to central location facilities, however, is that they can be extremely expensive to administer because of the costs of research facility staff, rent, and personnel, which may double the costs associated with a mall-intercept survey.<sup>67</sup>

In contrast to mall-intercept and central location formats, the telephone format is relatively cheap and fast and is widely used in trademark litigation concerning genericness.<sup>68</sup> Generally, telephone surveys ask yes/no questions or provide respondents with a defined set of possible answers from which to choose.<sup>69</sup> Additionally, since telephone numbers can be randomly selected, the telephone format makes it easier to attain a probability sample that is statistically representative of the universe as a whole.<sup>70</sup> While these advantages have made the telephone survey popular in trademark litigation, both the courts and scholars have commented that changing social habits have significantly decreased response rates to these surveys in recent years: increased cell phone usage has made it more difficult for research firms to contact respondents, because such firms may have access only to home phone numbers; caller ID and blocking features have limited the ability to reach landlines; and the increased frequency of two household earners has reduced the number of people who are at home during the day and are willing to be interviewed.<sup>71</sup> On top of these factors, the yes/no, closed-end format of telephone surveys makes it more difficult for interviewers to gauge whether respondents are guessing and whether they are answering with sufficient knowledge.<sup>72</sup>

An increasingly popular environment in which to conduct consumer surveys is the Internet.<sup>73</sup> Advantages of Internet surveys include convenience for the respondent, the ability to quickly and easily survey a larger number of respondents in the relevant universe of consumers,<sup>74</sup> decreased risk of interviewer data-entry

---

<sup>66</sup> *Id.*

<sup>67</sup> *Id.* at 96.

<sup>68</sup> *Id.* The original *Teflon* survey was conducted telephonically. *E. I. DuPont de Nemours & Co. v. Yoshida Int'l, Inc.*, 393 F. Supp. 502, 525 (E.D.N.Y. 1975).

<sup>69</sup> Thornburg, *supra* note 60, at 96

<sup>70</sup> McCarthy, *supra* note 6, at § 32:164.

<sup>71</sup> *Id.* (also citing the more general phenomenon of an increasing number of people who are simply unwilling to spend time on surveys).

<sup>72</sup> Thornburg, *supra* note 60, at 96.

<sup>73</sup> McCarthy, *supra* note 6, at § 32:165.25.

<sup>74</sup> *Id.*; see also Thornburg, *supra* note 60, at 120-121.

mistakes and biasing<sup>75</sup> as a result of computerized data collection, and reduced expenses.<sup>76</sup> One major downside of the Internet survey in comparison to other types of surveys is the inability to ask follow-up questions.<sup>77</sup> Additionally, Internet respondents, to the ignorance of survey administrators, could consult outside resources in order to answer questions<sup>78</sup> or take the survey in the company of peers who may affect their answers.<sup>79</sup>

Regardless of the environment in which the survey takes place, surveys can be prone to wording-related factors and personnel issues that can affect respondents' answers. Whether taking the survey in a mall, in a supermarket, over the phone, or over the Internet, respondents can be affected by what they have just seen or heard in respect to the mark in question.<sup>80</sup> Further diminishing the probative value of surveys are question order<sup>81</sup> and wording effects.<sup>82</sup> Since the order of questions can affect respondents' answers,<sup>83</sup> the courts have at times criticized surveys that do not randomize the order of questions across respondents.<sup>84</sup> Likewise,

---

<sup>75</sup> Biasing in verbal survey environments can include reading "questions different, with different stress and intonation, creating the impression that there is a 'correct' answer." McCarthy, *supra* note 6, at § 32:165.25.

<sup>76</sup> Thornburg, *supra* note 60, at 122.

<sup>77</sup> McCarthy, *supra* note 6, at § 32:165.25.

<sup>78</sup> See Poret, *supra* note 46 (but arguing that such practice is unlikely given that most respondents wish to quickly complete surveys rather than pick the "correct" answers).

<sup>79</sup> Other objections to Internet surveys, such as the supposed inability to determine whether respondents are taking the survey seriously, have been raised and addressed by Internet survey experts. See generally Himanshu Mishra & Ruth M. Corbin, *Internet Surveys in Intellectual Property Litigation: Dovesyay, No Proveryay*, 107 TMR 1097 (2017) (discussing methods of detecting and reducing cheating, faking, and inattention in online survey responses).

<sup>80</sup> *Id.* § 32:171 (also noting that interviewees may have just purchased the product in question). These issues, of note, are not entirely absent from corpus-based measurements of genericness. Users of a mark on a blog, website, or newspaper article may also have recently seen an advertisement for or purchased the mark in question. The benefit, however, of this type of usage is that it does not take place in an artificial environment, as does a survey.

<sup>81</sup> Question order effects are also known as context effects. Jay, *supra* note 21, at 1141.

<sup>82</sup> *Id.* (citing Howard Schuman & Stanley Presser, *Questions & Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context* (1996)).

<sup>83</sup> Jay, *supra* note 21, at 1141.

<sup>84</sup> *Classic Foods Int'l Corp. v. Kettle Foods, Inc.*, 468 F. Supp. 2d 1181, 1193 (C.D. Cal. 2007) ("KFI's survey, however, has several flaws in its methodology. It uses a circular definition for what a brand name is, it fails to screen subjects to see if they understood what a brand is, and it fails to identify or account for any order bias that might have been produced."), cited by Jay, *supra* note 21, at 1141 n.87. Cf. *March Madness Ath. Ass'n v. Netfire, Inc.*, 310 F. Supp. 2d 786, 803 (N.D. Tex. 2003) (finding convincing the respondent's survey that rotated the sequence of terms across respondents), cited by Jay, *supra* note 21, at 1141 n.87.

leading<sup>85</sup> and ambiguously worded questions<sup>86</sup> have also undercut the probative value of surveys proffered in cases involving genericness claims. In addition to these problems, there is the perhaps more concerning risk of personnel and tabulation problems affecting non-Internet surveys.<sup>87</sup> Surveyors may err by recording inconclusive responses as unqualified, using unprofessional interviewers, or failing to accurately record and code responses.<sup>88</sup> Because of the ephemeral nature of the oral survey environment, such misclassification may go unchecked at trial if responses are not audio recorded.

For two types of marks in particular, it can be especially challenging for surveys to adequately measure genericness: polysemous marks referring to both a trademark and the common name of that trademark's respective product class,<sup>89</sup> and unique products.<sup>90</sup>

An example of polysemy: consumers may understand and use the mark JEEP both as a brand name for a specific producer of cars and as a generic name for a type of car.<sup>91</sup> Likewise, consumers may understand and use the mark KLEENEX both as a brand name for a specific producer of facial tissues and as a generic name for all facial tissues.<sup>92</sup> Therefore, it does not make sense to ask consumers

<sup>85</sup> See *In re Minnetonka, Inc.*, 212 U.S.P.Q. (BNA) 772, 780 (T.T.A.B. 1981) (heavily criticizing the applicant's survey for its use of a "flagrantly leading" question. Specifically, to respondents who responded with the mark, "Soft Soap," to a question inquiring about common names for a product, the survey posed a follow-up question asking "Do any companies also use [Soft Soap] as their brand name?," which the court noted "strongly implies that the respondent was mistaken in having indicated a brand name (in answer to Question 2) as a common name for the product category and that he or she is now being given an opportunity (tactfully) to correct this error.").

<sup>86</sup> *Tea Bd. of India v. Republic of Tea, Inc.*, 80 U.S.P.Q.2d (BNA) 1881 (T.T.A.B. 2006) (finding that applicant's survey asking "What is Darjeeling tea?" "was not even designed to show that the term is generic.").

<sup>87</sup> See McCarthy, *supra* note 6, at § 32:171 (discussing many issues that can hinder surveys' probative value).

<sup>88</sup> See, e.g., *Revlon Consumer Products Corp. v. Jennifer Leather Broadway, Inc.*, 858 F. Supp. 1268, 1276 (S.D.N.Y. 1994) (According "no weight" to plaintiff's likelihood of confusion survey due to misclassification of responses).

<sup>89</sup> See Ralph H. Folsom & Larry L. Teply, *Surveying "Genericness" in Trademark Litigation*, 78 TMR 1, 6 (1988) (noting that genericness surveys do not account for "an important third category of consumers: those consumers who regard a particular mark as simultaneously functioning as a source-significant word and as the common descriptive name of a product class.").

<sup>90</sup> *A.J. Canfield Co. v. Honickman*, 808 F.2d 291, 302 (3d Cir. 1986) (doubting a survey's ability to measure consumers' understanding of a mark identifying a product for which there is only one producer because "generic marks signifying goods produced by only one manufacturer may function both as generic terms, signifying the product genus, and as brand names, indicating continuity of source."), cited by Jay, *supra* note 21, at 1162 n.197.

<sup>91</sup> Folsom & Teply, *supra* note 89, at 6.

<sup>92</sup> When interpreting "Can you please give me a 'Kleenex?'," one may provide a PUFFS-brand tissue and understand the term in a generic sense, while still maintaining the

to categorize such terms as trademarks or as generic names exclusively. Neither a classification as “generic” nor as “trademark” would adequately capture the nuanced significance of the term to respondents who recognize both meanings. As a solution to this issue, some scholars have proposed incorporating into the *Teflon* survey questions asking respondents “whether they use each term mainly as a common name, mainly as a brand name, equally as a common and brand name, or not at all.”<sup>93</sup> A major problem with this proposed method is that people are not very good at observing their own speech, let alone explaining it to surveyors.<sup>94</sup>

Additionally, surveys have a difficult time measuring the genericness of marks for unique products.<sup>95</sup> When originally presented with a fanciful name for a new product, such as “escalator,” consumers have no need to differentiate between the brand name for the product and the generic name for the product, because there is only one brand name associated with the product. As the court noted in *A.J. Canfield Co. v. Honickman*, a survey asking respondents to categorize a mark for a unique product as a brand name or as a common name is asking respondents to make a “false dichotomy.”<sup>96</sup> A few solutions have been proposed to address this issue; most notably, Palladino has suggested asking respondents hypothetical questions such as “If this type of product were made by more than one company, would you associate [the word at issue] with the product of one, or more than one, company?”<sup>97</sup> However, such a question would require respondents, who are already engaged in the artificial language environment of a survey, to place themselves in a hypothetical situation, potentially rendering answers even more divorced from the reality of their understanding and usage.<sup>98</sup>

The crux of the matter is this: surveys are unnatural linguistic environments, and therefore, prone to extraneous factors affecting

---

ability to differentiate between KLEENEX- and PUFFS-brand tissues at a store. Thus, a *Teflon*-type classification of “Kleenex” may force the respondent to make a false dichotomy.

<sup>93</sup> Simonson, *supra* note 19, at 187. See also Folsom & Teply, *supra* note 89, at 26 (proposing a survey that measures the extent to which the mark “performs hybrid functions”).

<sup>94</sup> See Natalie Schilling, *Sociolinguistic Fieldwork 77* (1st ed. 2013) (citing studies indicating that respondents may overreport or underreport certain instances of language usage such that their responses to elicitation tasks do not match their levels of language usage in conversational data. “[R]espondents may not be consciously aware of their usage patterns or know how to express them (e.g. ‘How often do you use the form ‘fixin’ to?’); in addition, they may purposely over- or under-report their use of particular forms.”).

<sup>95</sup> Jay, *supra* note 21, at 1162.

<sup>96</sup> 808 F.2d 291, 302 (3d Cir. 1986).

<sup>97</sup> Palladino, *supra* note 40, at 882.

<sup>98</sup> Even Palladino comments that surveys using such a methodology could perhaps be flawed, because it seeks, to some extent at least, to gauge public reaction to a hypothetical situation. *Id.* at 882.

the perceived understanding of marks they seek to elicit. As McCarthy states:

[T]here are three unavoidable constraints that render all surveys less than perfect reflections of actual human behavior and perception: (1) Persons never behave exactly the same in a survey setting as in real life. The mere fact of being observed and questioned itself compromises “normal” behavior; (2) There are never infinite resources of time and money with which to conduct a survey and explore all of the permutations and combinations; (3) Sometimes, respondents simply do not know the answer to the question that the law seeks to resolve.<sup>99</sup>

On top of these factors, surveys tend to measure consumers’ elicited understanding of a disputed mark’s meaning as opposed to consumers’ actual usage of a mark. Actual, naturally occurring instances provide a better gauge of a word’s meaning, because they are less constrained by the artificial elements to which survey responses are exposed and can be assumed to reflect consumers’ intended meanings.<sup>100</sup> Record evidence, which comprises the foundation from which corpora are built, captures such naturally occurring usage.

### ***B. Record Evidence***

Record evidence—recorded usage of a mark by the media, the plaintiff, competitors, or others—is frequently accepted by the court as evidence of genericness or lack thereof.<sup>101</sup> Evidence of trademark/generic usage can include initial capitalization<sup>102</sup> of the

---

<sup>99</sup> McCarthy, *supra* note 6, § 32:178 (citing Michael Rappeport, *Litigation Surveys: Social Science as Evidence*, 92 TMR 957, 961 (2002)).

<sup>100</sup> See *infra* Part III(B)(4) for discussion of the probative value of actual usage versus that of elicited understanding.

<sup>101</sup> Numerous courts have found the usage of disputed terms in a generic sense to be probative of genericness. *E.g.*, *Murphy Door Bed Co. v. Interior Sleep Sys., Inc.*, 874 F.2d 95 (2d Cir. 1989) (finding generic usage in newspapers and magazines to be “a strong indication of the general public’s perception that Murphy bed connotes something other than a bed manufactured by the Murphy Co.” (i.e., that “murphy bed” is generic)), *quoted in* McCarthy, *supra* note 6, § 12:13. See also Rita M. Irani, *The Importance of Record Evidence to Categorize Marks As Generic, Descriptive, or Suggestive*, 83 TMR 607 (1993) (noting an uptick in the reliance on record evidence facilitated by the “information explosion of the 1980s” as well as criticism by some courts of consumer surveys).

<sup>102</sup> See, *e.g.*, *Door Sys., Inc. v. Pro-Line Door Sys., Inc.*, 83 F.3d 169, 171 (7th Cir. 1996) (“If the term ‘door systems’ appeared in a standard dictionary in lower case, this would be powerful evidence that the term was generic”), cited by Heymann, *supra* note 6, at 1341 n.106; *Nat’l Nonwovens, Inc. v. Consumer Prods. Enters., Inc.*, 397 F. Supp. 2d 245, 250 (D. Mass. 2005) (finding uncapitalized instances of “wool felt” in Nexis and Google search results to be probative of genericness); *In re Country Music Ass’n*, 100 U.S.P.Q.2d (BNA) 1824 (T.T.A.B. 2011) (“[I]n the English language, initial capitalization of a term or phrase is generally used to designate a brand name, as opposed to a generic term.”). See

mark and sentential context.<sup>103</sup> Since word-initial capitalization in American-English<sup>104</sup> orthography is conventionally used to indicate proper names, capitalization is a useful measurement of the status of a mark as generic or as a trademark.<sup>105</sup> Likewise, when writers use a plural word indicating “producers” after a mark, it can be inferred that they are referring to a generic product rather than to a specific brand name associated with just one producer.<sup>106</sup> For example, use of the phrase “ipod companies” or “kleenex brands” implies that “ipods” and “kleenexes” are products made by more than one producer, indicating that they are generic. While these types of usage are strong indicators of whether the language user intends to use the mark in question as a trademark or as a generic term, there are other forms of record evidence, accepted by some courts, that this article and other scholars contend do not have cognizable probative value.

### 1. Unsound Record Evidence

In many instances, courts have found usage of a mark as a verb or as a noun to be probative of genericness.<sup>107</sup> Certain scholars of

---

*infra* Part IV(B)(1) for an extended discussion of why capitalization makes sense for the analysis of disputed marks.

<sup>103</sup> See, e.g., *In re Energy Prods. of Idaho*, 13 U.S.P.Q.2d 2049, at \*1 (T.T.A.B. 1989) (citing as an example of generic usage of the mark WASTE-TO-ENERGY, the following phrase from an April 26, 1988, *Los Angeles Times* article: “Henley Group gained \$2.50 to \$24.875 as it announced a new entity, in combination with Waste Management and its Wheelabrator Technology division, to form a ‘billion-dollar’ waste-to-energy company.”). In the *LA Times* quote cited as evidence of genericness, the mark is not only in all lowercase letters, but it also precedes the word, “company” and is preceded by “a,” implying that the mark is generic for all “waste-to-energy” companies. See *infra* Part (IV)(D), discussing how sentential context evidence can be made more empirical through the collocation analysis of corpus linguistics.

<sup>104</sup> U.S. trademark law is concerned exclusively with U.S. consumers’ usage and understanding of marks. Thus, foreign publications are not probative of genericness in the eyes of U.S. courts. See, e.g., *UGG Holdings, Inc. v. Severn*, 2005 WL 5887187, at \*6 (C.D. Cal., 2005) (Finding the UGG mark’s generic status in Australia to be irrelevant to a determination of its trademark status in the U.S. and noting that the doctrine of foreign equivalents is inapplicable when the disputed term is not a word from a foreign language); *In re Consol. Cigar Corp.*, 13 U.S.P.Q.2d 1481 (T.T.A.B. 1989) (With regard to supposed evidence of generic use of the mark WHIFFS, “[e]ven if considered, this evidence, since it comprises foreign publications, would be of little moment in determining the instant issue.”).

<sup>105</sup> See generally Adam Kilgarriff, *Corpus Linguistics in Trademark Cases*, 36 J. Dictionary Soc’y N. Am. 100 (2015) (stating that capitalization is conventionally used in American English to denote proper names and is therefore a fair gauge of a mark’s “name-likeness,” or degree to which a term is generic or is a trademark); Heymann, *supra* note 6 (“If consumers understand the word to be the proper name of a brand of product or service, the word serves as a trademark.”).

<sup>106</sup> See *supra* note 103, discussing generic usage of WASTE-TO-ENERGY.

<sup>107</sup> See, e.g., *Birtcher Electro Med. Sys. v. Beacon Labs., Inc.*, 738 F. Supp. 417, 420 (D. Colo. 1990) (finding use of the disputed terms “as nouns as opposed to adjectives” to be evidence of genericness), cited by Heymann, *supra* note 6, at 1344 n.117; *In re Treeradar*,

trademark law and linguistics,<sup>108</sup> however, argue that consumers can use trademarks as verbs, as nouns, and as metaphors without the mark becoming generic.<sup>109</sup> This argument has been corroborated by some courts. In *Elliot v. Google, Inc.*, the Ninth Circuit held that even if the public uses “google” as a verb to refer to the act of using a search engine generally, the fact that the public recognizes GOOGLE in the context of brands of search engines as a trademark renders the term not generic.<sup>110</sup> This finding is in keeping with one’s intuition that GOOGLE is a very strong trademark, even though consumers might frequently use the term as a generic verb to refer to searching the Internet. Likewise, when one uses iPhone as an inflected noun in asking a friend, “How many iPhones do you have?,” the consumer likely still understands “iPhone” as referring to Apple’s iPhone brand of smartphone, especially if the “p” in the term is capitalized. Trademarks also can be used as metaphors without becoming generic; something can be called the “Cadillac of its class” and someone can be said to have a “Napoleon complex” without CADILLAC losing its significance as a trademark for cars and without “Napoleon” losing its significance as someone’s name.<sup>111</sup> This article, thus, agrees with and advocates for the idea that trademarks can be used as nouns or verbs, can be inflected, and can be used as metaphors without becoming generic.<sup>112</sup>

## 2. Limitations of Record Evidence

Record evidence, however, even when it consists of usage that is truly indicative of genericness, is generally not presented with the same consideration of scientific and statistical principles with which

---

Inc., 2011 WL 3212252, at \*7 (T.T.A.B. 2011) (“In one instance the term is used in the manner of a verb (‘TreeRadar for tree decay’) which indicates generic use rather than trademark use.”).

<sup>108</sup> *E.g.*, Heymann, *supra* note 6, at 1318 (“To say that all uses of a trademark inconsistent with its prescribed form will cause that trademark to lose its original associations reflects a narrow view of how language functions in society.”).

<sup>109</sup> *Id.*

<sup>110</sup> 860 F.3d 1151, 1159 (9th Cir. 2017) (“Even if we assume that the public uses the verb ‘google’ in a generic and indiscriminate sense, this tells us nothing about how the public primarily understands the word itself, irrespective of its grammatical function, with regard to internet search engines.”). *See infra* Part IV(B)(1), discussing why capitalization of verbs can be probative of trademark validity even when verb usage itself is not determinative of genericness.

<sup>111</sup> Heymann, *supra* note 6, at 1336, 1338.

<sup>112</sup> Even if courts do not recognize the usage of trademarks as verbs and nouns as sufficient evidence of genericness (and they should not), it may still be in the economic interests of trademark holders to avoid using their marks in such ways themselves and to enjoin others to only use their trademarks as capitalized adjectives modifying generic nouns. Advertisements discouraging certain uses of a trademark by themselves create increased awareness that the term is a trademark.

surveys are administered.<sup>113</sup> The record evidence proffered by parties to a genericness suit often includes fewer than 100 examples of usage and sometimes includes only five to twenty examples indicating generic or trademark usage.<sup>114</sup> When such evidence involves capitalization, the capitalization rates of all words in the sample from which the examples were chosen are generally not provided.<sup>115</sup> For example, one defendant in *Official Airlines Guides, Inc. v. Goss*<sup>116</sup> presented as evidence sixty-four examples of generic usage in newspapers and magazines as well as thirty-nine such examples in government publications; however, the defendant did not make an effort to show that the usage was a representative sample or was significant with respect to the capitalization norms in those mediums.<sup>117</sup> Even when thousands of instances of record evidence are provided, such evidence, when unsubstantiated by the representativeness of the sample from which it is taken, does little to support a claim of genericness or of trademark validity.<sup>118</sup> Such lack of statistical perspective is what currently handicaps record evidence in comparison to survey evidence.

Another concerning shortcoming of record evidence as it has historically been presented is that it has consisted primarily of publications in the media rather than of consumer writing or speech in natural environments.<sup>119</sup> The problem with using media usage as evidence in genericness cases is that such usage might be biased toward trademark usage and, therefore, unrepresentative of the language of the consuming public, as a result of conservative editorial policies and letters received from producers.<sup>120</sup> Despite

---

<sup>113</sup> For surveys, courts have cited several authorities on these principles. See *supra* note 24 for authorities on principles of surveying in trademark law.

<sup>114</sup> See, e.g., See also Irani, *supra* note 101 (citing, e.g., *In re Energy Prods. of Idaho*, 13 U.S.P.Q.2d 2049 (T.T.A.B. 1989)).

<sup>115</sup> See *infra* Part IV(B)(4), discussing how to establish capitalization rate norms against which to compare the capitalization rate of the mark in question.

<sup>116</sup> 6 F.3d 1385 (9th Cir. 1993) (relying upon record evidence in newspapers, magazines and government publications to find that the term “travel planner” is generic).

<sup>117</sup> *Id.* Such examples, where several pieces of record evidence are provided but unqualified as to their statistical representativeness are fairly common in the case law. See, e.g., *Baroness Small Estates, Inc.*, 2012 WL 4763149, at \*3 (Finding several, contradictory examples of usage insufficient to show genericness).

<sup>118</sup> See, e.g., *San Diego Comic Convention v. Dan Farr Productions*, 2017 WL 4227000, at \*7 (S.D. Cal., 2017) (granting plaintiff’s motion to exclude defendant’s expert report, partly on the basis that the expert’s record evidence, featuring over 1,800 pages of documents, “was generated from a narrow field of documents.”).

<sup>119</sup> See McCarthy, *supra* note 6, § 12:13; Irani, *supra* note 101.

<sup>120</sup> Ralph H. Folsom & Larry L. Teply, *Trademarked Generic Words*, 89 Yale L.J. 1323, 1353 (1980) (arguing that record evidence from media does not necessarily provide “a truly representative sampling of media usage” and that such record evidence “may reflect not the public’s understanding of the trademarked word, but rather the media’s own editorial policies—policies that inevitably reflect the existence of numerous trademark-policing programs designed to pressure the media into ‘proper’ trademark usage”). See

these issues, record evidence has been well received by some courts in its current form;<sup>121</sup> making the analysis and presentation of record evidence more empirical, however, has the potential to drastically increase its probative value. Corpora, properly constituted and analyzed, allow for such empiricism.

### III. CORPUS LINGUISTICS AND CORPORA

Corpora are large collections of textual data that can be used to analyze patterns of language use. Corpora have a variety of applications—linguists use corpora of historical language to examine changes in the use of words over time,<sup>122</sup> political scientists use corpora to investigate patterns of political discourse,<sup>123</sup> and lexicographers use corpora to create dictionary definitions.<sup>124</sup> The data comprising a corpus can be harvested from transcripts of oral language,<sup>125</sup> books,<sup>126</sup> the Internet,<sup>127</sup> and a host of other sources. Corpora can be generalized, consisting of tens of millions of words from a variety of sources, or they can be specialized, consisting of data from specific sources such as Wikipedia<sup>128</sup> or from specific speakers such as upper-level college students.<sup>129</sup> The data may be raw such that the corpus essentially resembles a very large text document, or the data may be annotated for parts of speech,

---

*infra* Part IV(A)(1), however, discussing why media usage, although not preferable, can still be useful for measuring genericness.

- <sup>121</sup> See, e.g., *In re Volvo White Truck Corp.*, 16 U.S.P.Q.2d 1417, 1421 (T.T.A.B. 1990) (lack of record evidence from publications influenced TTAB decision to reverse the PTO's refusal to register the mark INTEGRAL SLEEPER), cited by Irani, *supra* note 101, at 610 n.27.
- <sup>122</sup> *Changes*, Corpus.BYU.Edu, Brigham Young University, <https://corpus.byu.edu/coca/help/changes.asp> (discussing, for example, researching the trend of words with the suffix “-gate,” indicating “scandal,” in the 1990s and 2000s).
- <sup>123</sup> See generally, e.g., Rebecca Willis, *Taming the Climate? Corpus analysis of Politicians' Speech on Climate Change*, 26 *Environmental Politics* 212 (2017) (Using corpora to investigate politicians' discussion of climate change, noting an avoidance of words implying irreversibility and a material social impact); Robert H. Thornburg, *Trademark Survey Evidence: Review of Current Trends in the Ninth Circuit*, 21 *Santa Clara Computer & High Tech. L.J.* 715, 717 (2005).
- <sup>124</sup> See, e.g., *From Corpus to Dictionary*, Macmillan Dictionaries (2013), <http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/> (discussing how Macmillan uses corpora to create its dictionaries).
- <sup>125</sup> See, e.g., *The Buckeye Speech Corpus*, Ohio State University (2005), <http://buckeyecorpus.osu.edu/>.
- <sup>126</sup> See, e.g., *Google Books Corpus*. Google Books Ngram Viewer (2017), <https://books.google.com/ngrams>.
- <sup>127</sup> See, e.g., *Corpus of Global Web-Based English*, Brigham Young University (2012–2013), <http://corpus.byu.edu/glowbe/>.
- <sup>128</sup> See *The Wikipedia Corpus*, Brigham Young University (2014), <http://corpus.byu.edu/wiki/>.
- <sup>129</sup> See *Michigan Corpus of Upper-Level Student Papers*, University of Michigan (2010), <http://eli-corpus.lsa.umich.edu/>.

syntactic structure, or other information.<sup>130</sup> Whether annotated or not, corpora can be used to measure the frequency of usage of a word in the language, the rate at which a word is capitalized,<sup>131</sup> and the rate at which other words juxtapose a term in question,<sup>132</sup> among other features.<sup>133</sup> In addition to allowing one to quantify information about a word's use, corpora allow the expert to qualitatively analyze the sentential context of a word in question.<sup>134</sup>

While the first corpus, the *Brown Corpus*, was created over fifty years ago,<sup>135</sup> the recent explosion of information freely available on the Internet has made it incredibly easy to create large corpora, even ones that target specific universes of language users. For example, Mark Davies created his immensely large (1.9 billion-word) *Corpus of Global Web-Based English (GloWbe Corpus)* by harvesting textual data from 1.8 million web pages, including those of online publications and blogs.<sup>136</sup> Likewise, the *BYU News on the Web (NOW) Corpus*, composed of text from online magazines and newspapers, currently features greater than six billion words and is growing at a rate of more than one hundred million words per month.<sup>137</sup> If the numerous already-existent corpora are deemed inadequate for the corpus-linguist's analysis of interest, it is easy to automatically construct corpora using corpus-creation websites

---

<sup>130</sup> Mary Ann Marcinkiewicz, *Building a Large Annotated Corpus of English: The Penn Treebank*, 19 *Computational Linguistics*, 313 (1993).

<sup>131</sup> Corpora composed of speech cannot be used to measure capitalization rates; most corpora, however, are composed of written language, which, of course, can be used for this purpose.

<sup>132</sup> Words that habitually juxtapose the word being analyzed are called high-frequency collocates. See generally Gena R. Bennett, *Using Corpora in the Language Learning Classroom*, 9 (2010). For example, high frequency collocates for the word, "pie," might include "apple," "pumpkin," and "sweet potato" (i.e., "apple pie," "pumpkin pie," and "sweet potato pie").

<sup>133</sup> *Id.* See also Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 *BYU L. Rev.* 1915 (2010) (discussing four unifying characteristics of corpus linguistics: the empirical analysis of actual patterns of natural language use, large and principled collection of natural texts, the extensive use of computers for analysis, and the incorporation of both quantitative and qualitative techniques) (quoting Douglas Biber, *Corpus-based and Corpus-driven Analyses of Language Variation and Use* in *The Oxford Handbook of Linguistic Analysis* 159 (2009)).

<sup>134</sup> Concordance lines are the contexts in which instances of a word appear in the corpus; they can be sentences, a defined number of words appearing before and after the term of interest, or even an entire paragraph. See generally Bennett, *supra* note 132, at 17.

<sup>135</sup> *Id.* at 2 ("The first computer-based corpus, the Brown corpus, was created in 1961 and comprised about 1 million words.").

<sup>136</sup> *Corpus of Global Web-Based English*, Brigham Young University (2012–2013), <http://corpus.byu.edu/glowbe/>.

<sup>137</sup> *News on the Web Corpus*, Brigham Young University, <http://corpus.byu.edu/now/> (last visited Aug. 18, 2018).

such as Sketch Engine<sup>138</sup> or to do so manually using available web-scraping and text harvesting software.<sup>139</sup>

### A. Corpora in the Courts

In law, corpora can be especially useful for judges, attorneys, expert witnesses, and legal scholars concerned with patterns of language use or with the meaning of certain words. As the literature and some courts have already noted,<sup>140</sup> corpora are specifically useful for ascertaining plain meaning<sup>141</sup> and, as this article further discusses, for evaluating disputed trademarks.

#### 1. Corpus-Based Analyses of Plain Meaning

While judges have often consulted dictionaries, precedent, or their intuition, corpus-based analyses are becoming increasingly popular for discovering plain meaning.<sup>142</sup> Stephen Mouritsen, a corpus-linguist and legal scholar, and Justice Thomas Lee, Associate Justice of the Utah Supreme Court, have been among the most significant proponents of using corpora for such purposes.<sup>143</sup>

Most notably, the Supreme Court in *FCC v. AT&T, Inc.*<sup>144</sup> referenced an amicus curiae brief that incorporated a corpus-based analysis of the word “personal.”<sup>145</sup> At issue in the case was the plain meaning of the word “personal”; specifically, because corporations

<sup>138</sup> *Create Your Corpus*, Sketch Engine (2017), <https://www.sketchengine.co.uk/quick-start-guide/create-your-corpus-lesson-4/>.

<sup>139</sup> Web Scraper (2017), <http://webscraper.io/>.

<sup>140</sup> *See, e.g.*, *State v. Rasabout* 356 P.3d 1258 (Utah 2015) (Lee, J., concurring in part and concurring in the judgment, employed a corpus-based analysis to conclude the word “discharge” typically refers to a single burst of a firearm). *See infra* Part III(A)(1) for further discussion of the application of corpora to plain meaning.

<sup>141</sup> The plain meaning rule holds that statutes are to be interpreted using the ordinary meaning of the language, unless such an interpretation would be absurd or cruel. *See Plain Meaning Rule Law and Legal Definition*, U.S. Legal (2016), <https://definitions.uslegal.com/p/plain-meaning-rule/>.

<sup>142</sup> Mouritsen notes that surveys can also be used to gauge the plain meaning of words, but he identifies, as this article does, several flaws inherent to surveys, including context effects whereby the respondent’s answer is influenced by other information on the questionnaire. Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*. 13 Colum. Sci. & Tech. L. Rev. 156, 186 (2011).

<sup>143</sup> *See* Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 Yale L.J. 788-1105 (2018); Mouritsen, *The Dictionary Is Not a Fortress*, *supra* note 133; Mouritsen, *Hard Cases and Hard Data*, *supra* note 142.

<sup>144</sup> *FCC v. AT&T Inc.*, 562 U.S. 397 (2011).

<sup>145</sup> *See* Mouritsen, *Hard Cases and Hard Data*, *supra* note 142, at 158 (quoting Transcript of Oral Argument, *FCC v. AT&T, Inc.*, 562 U.S. 397 (2011) (“JUSTICE GINSBURG: Mr. Klineberg, you have read the brief of the Project on Government Oversight, where they give dozens and dozens of examples to show that, overwhelmingly, ‘personal’ is used to describe an individual, not an artificial being. And it is the overwhelming use of ‘personal.’”).

are “artificial persons,” could the “personal privacy” exemption in the Freedom of Information Act be applied to corporations?<sup>146</sup> The amicus curiae brief submitted by the Project on Government Oversight used three corpora to examine whether the words “personal” or “privacy” are commonly used in reference to corporations or businesses.<sup>147</sup> Based on searches in the *Time Corpus of American English*, the *Corpus of Contemporary American English (COCA)*, and the *Corpus of Historical American English*, the brief listed the nouns that “personal” modifies with the highest frequency and the possessive pronouns that most frequently modify “privacy.” The brief’s study found that instances of phrases containing “personal” “can be used only with regard to human beings”<sup>148</sup> (i.e., “personal integrity,” “personal communication,” “personal responsibility,” etc. are used only in reference to humans). Likewise, the study found that almost all possessive pronouns modifying “privacy” indicate that the word is used in reference to humans (i.e., instances of “my privacy,” “our privacy,” “his privacy” were drastically more frequent than instances of “its privacy”).<sup>149</sup> Although the Court’s opinion did not explicitly reference the corpus study, Justice Ginsburg cited the amicus brief during the proceedings.<sup>150</sup>

In addition to *AT&T*, Justice Lee of the Utah Supreme Court has administered *sua sponte* corpus-based analyses to gauge plain meaning.<sup>151</sup> In *Baby E.Z. v. T.I.Z.*,<sup>152</sup> the first time formal corpora were referenced in a reported opinion,<sup>153</sup> it was disputed whether the term “custody,” as used in the Parental Kidnapping Prevention Act, applied to adoption proceedings in addition to divorce-related proceedings. Addressing this issue of plain meaning, Justice Lee reviewed 500 randomized sample sentences from the *Corpus of Contemporary American English*, which indicated that “[t]he word ‘custody’ is approximately ten times more likely to collocate with the

<sup>146</sup> *FCC v. AT&T Inc.*, 562 U.S. 397 (2011).

<sup>147</sup> Brief for the Project on Government Oversight, the Brechner Center for Freedom of Information, and Tax Analysts as Amici Curiae, *FCC v. AT&T Inc.*, 562 U.S. 397 (2011).

<sup>148</sup> *Id.*

<sup>149</sup> *Id.*

<sup>150</sup> See *supra* note 145.

<sup>151</sup> 2015 UT 72, 356 P.3d 1258 (Utah 2015); *State v. J.M.S.*, 280 P.3d 410 (Utah 2011); *Baby E.Z. v. T.I.Z.*, 266 P.3d 702 (Utah 2011).

<sup>152</sup> 266 P.3d 702, 724 (Utah 2011).

<sup>153</sup> Though not explicitly referencing corpora or necessarily employing the methods of corpus linguistics, the majority in *Muscarello v. U.S.*, in 1998, found the word “carries” as in “carries a firearm” to encompass not only carrying firearms on one’s person but also conveying firearms in a vehicle by relying upon databases of newspapers and Lexis/Nexis searches. 524 U.S. 125, 129 (1998) (“We looked for sentences in which the words ‘carry,’ ‘vehicle,’ and ‘weapon’ (or variations thereof) all appear. We found thousands of such sentences, and random sampling suggests that many, perhaps more than one-third, are sentences used to convey the meaning at issue here, i.e., the carrying of guns in a car.”).

word ‘divorce’ than with the word ‘adoption’ in contemporary usage.”<sup>154</sup> More recently, in *State v. Rasabout*, in which it was debated whether twelve gunshots supported one or twelve counts of “unlawful discharge of a firearm,” Justice Lee administered a *sua sponte* corpus-based analysis of the word “discharge,” which indicated that the term was almost exclusively linked to a single bullet as opposed to a burst of shots.<sup>155</sup> However, Justice Lee’s reliance on corpora was criticized by his colleagues. In *Rasabout*, the majority expressed concern that such *sua sponte* analyses were not subject to scientific review or cross examination,<sup>156</sup> and in *T.I.Z.*, the majority questioned whether Justice Lee’s analysis necessarily precluded the relevance of the word “custody” to adoption hearings.<sup>157</sup> These criticisms are not necessarily well taken, however, as the dictionary definitions of “discharge,” cited by the majority in *Rasabout*, are also not subject to scientific review,<sup>158</sup> and as Justice Lee notes,<sup>159</sup> corpus data can more easily be challenged in a petition for rehearing.<sup>160</sup> Despite conflicting judicial opinions on corpus-based studies of plain meaning,<sup>161</sup> these analyses at least

---

<sup>154</sup> 266 P.3d 702, 724 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment).

<sup>155</sup> 356 P.3d 1258, 1271-90 (Utah 2015) (Lee, J., concurring in part and concurring in the judgment).

<sup>156</sup> 356 P.3d 1258, 1264 (Utah 2015) (“We decline to adopt his approach because, among other reasons, it is unfair to the parties and it attempts scientific research that is not subject to scientific review.”)

<sup>157</sup> 266 P.3d 702, 725 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment) (“Both the majority and Justice Durrant in his separate concurrence object to my reliance on linguistic data from an electronic corpus in analyzing the comparative usage of different possible meanings of the term *custody* in the PKPA, contending that such analysis is ‘of little analytical or persuasive value.’”).

<sup>158</sup> This was noted by Justice Durrant in his concurring opinion. *State v. Rasabout*, 356 P.3d 1258, 1264 (Utah 2015) (Durrant, J., concurring in part and concurring in the judgment, noting that cited dictionaries and other tools for statutory interpretation have also not been argued by the parties, but citing Justice Lee’s own criticism of Justice Posner’s *sua sponte* attempt at corpus analysis to argue that such research may be “fraught with the potential for error” and therefore best left to the parties to raise themselves).

<sup>159</sup> This is noted by Justice Lee. *State v. Rasabout*, 356 P.3d 1258, 1283 (Utah 2015) (“A judge who presents a transparent corpus analysis opens the curtain in a manner allowing the parties to review and analyze his approach, and even challenge it on a petition for rehearing.”).

<sup>160</sup> While this article agrees that corpora can be useful tools for judges in gauging plain meaning, it does, particularly in the context of trademark law, recommend corpus-based analyses be introduced by the parties rather than the justices.

<sup>161</sup> Corpora were also consulted *sua sponte* by the Michigan Supreme Court in *People v. Harris*, with two justices disagreeing with majority’s analysis. 885 N.W.2d 832, 839 (Mich. 2016) (Finding that the word “information” may describe statements that are either true or false by citing the instances in the *Corpus of Contemporary American English* wherein “information” was “used in conjunction with adjectives suggesting it may be both true and false.”) (Markman, J., concurring in part and dissenting in part, found the majority’s corpus analysis warranted the opposite conclusion, that “information” generally refers to only true information, on the grounds that the vast

bring transparency and rigor to the examination of plain meaning, and their presence in opinions demonstrates the growing recognition of the corpus as a valuable analytical tool for studying language and meaning.<sup>162</sup>

## 2. Previous Corpus-Based Analyses in Genericness Litigation

While there is little legal scholarship and judicial guidance on corpus linguistics in trademark litigation, corpus-based evidence has been proffered, albeit not always methodically, by expert witnesses in a few genericness cases. Most notably, in *Scooter Store, Inc. v. SpinLife.com, LLC*, the court briefly opined on the plaintiff's citation of nine instances of "SCOOTER STORE" in BYU's various corpora.<sup>163</sup> Specifically, the plaintiffs proposed that, as all nine were related to its products, such evidence was indicative of the mark's validity. Given the limited sample size, the court rightly found that the data were insufficient to fulfill the burden of creating uncertainty of the term's genericness.<sup>164</sup> Promisingly, however, the court noted that corpus data is "in the category of [e]vidence of consumer recognition," which can be "relevant to assessing the strength of the mark."<sup>165</sup>

---

majority (99.44%) of instances of "information" in the corpus were "unmodified by any of these adjectives related to veracity.")

<sup>162</sup> In addition to the *sua sponte* corpus analyses discussed above, corpus analyses have also appeared in expert witness reports pertaining to issues of plain meaning. See, e.g., Expert Declaration of Dr. Judith Levi in Support of Whirlpool Corporation's Memorandum of Law Opposing LG's Motion for Preliminary Injunction, 2008 Misc. Filings LEXIS 9212 (supporting Whirlpool's claim that the word "steam" can have multiple meanings, including "a visible phenomenon that manifests at temperatures below (and often well below) one hundred degrees Centigrade.").

<sup>163</sup> *Scooter Store, Inc. v. SpinLife.com, LLC*, 2011 WL 6415516, at \*8 (S.D. Ohio, 2011). The specific corpora from which the instances were pulled are not specified.

<sup>164</sup> *Id.* With so few examples, it is difficult to make inferences about the larger population of relevant consumers. See *infra* Part III(B)(1) discussing sample sizes and Part IV(A) discussing representativeness in corpus design.

<sup>165</sup> *Id.* (quoting *Citizens Banking Corp. v. Citizens Fin. Group, Inc.*, 320 F. App'x 341, 347 (6th Cir. 2009)). Additionally, corpora briefly appeared in an opinion pertaining to the proposed exclusion of expert testimony. In *Leadership Studies, Inc. v. Blanchard Training and Development, Inc.*, the court found the defendant's expert report did not constitute improper rebuttal, partly on the basis that the report justifiably sought to rebut plaintiff's expert witness report, which featured corpus-based analysis of the disputed mark SITUATIONAL LEADERSHIP. Specifically, the court stated: "Although the reports prepared by Dr. Butters and Ms. Harper apply linguistic methods and cite various research tools to analyze how Situational Leadership is perceived by actual or potential consumers, their reports do not cite any direct evidence from specific individuals who would qualify as actual or potential consumers. Mr. Poret's report directly rebuts and contradicts the opinions of Dr. Butters and Ms. Harper based on 'a survey among actual and prospective consumers of the relevant type of services. . . ." 2018 WL 1989554, at \*14 (S.D. Cal. 2018).

Perhaps the most rigorous use of corpora in an expert report,<sup>166</sup> although it went unmentioned by the court, is that of linguist Oliver Northrup in *Moroccanoil, Inc. v. Marc Anthony Cosmetics, Inc.*<sup>167</sup> The plaintiff retained Northrup to gauge whether the term “MOROCCANOIL” was a brand name or a generic term for a category of hair treatment products.<sup>168</sup> Since the *Corpus of Contemporary American English* contained fewer than 50 instances of the term, Northrup created his own corpus composed of 586,317 (over 30 million words worth of) Amazon product reviews.<sup>169</sup> To ensure that the reviews reflected the language of U.S. consumers during the time frame of the existence of the MOROCCANOIL trademark (since 2007), Northrup filtered the corpus down to 155,763 product reviews that had U.S. location tags and were written during and after 2007.<sup>170</sup> In this sub-corpus, there were 451 instances of “[M/m]oroccan[ ][O/o]il” (case insensitive, including instances with and without spaces).<sup>171</sup> In 62% of these instances, either both the “M” in “Moroccan” and the “O” in “Oil” were capitalized, or the single word “Moroccanoil” was capitalized, generally indicating trademark usage.<sup>172</sup> In addition to considering the capitalization rates, Northrup also manually reviewed and coded<sup>173</sup> 300 of the 451 instances as either trademark, descriptive, or generic uses, finding that, out of the 260 instances that were classifiable, 60+/-6.0% were unambiguous trademark uses.<sup>174</sup> The defendant’s expert witness, meanwhile, criticized Northrup’s methodology for using an invalid sample.<sup>175</sup> Specifically, the defendant’s expert testified that Northrup’s corpus of Amazon product reviews was “not representative of the full context in which the words [“MOROCCANOIL”] might appear (e.g., beauty

---

<sup>166</sup> See Expert Report of Oliver B. Northrup, 2014 Misc. Filings LEXIS 4411.

<sup>167</sup> 2014 U.S. Dist. LEXIS 184585 (C.D. Cal. Oct. 7, 2014).

<sup>168</sup> Expert Report of Oliver B. Northrup, 2014 Misc. Filings LEXIS 4411.

<sup>169</sup> *Id.* at \*10.

<sup>170</sup> *Id.* at \*13.

<sup>171</sup> When there were multiple instances of “MOROCCANOIL” in a single product review, Northrup randomly selected one of the instances to include in his analysis to avoid overrepresenting any single product review. *Id.* at \*16.

<sup>172</sup> *Id.* at \*17. An additional 27% were instances of “Moroccan oil”; however, the status of these instances are more ambiguous as a result of the typical capitalization of adjectival references to countries (e.g., American, Canadian, Moroccan, etc.). *Id.*

<sup>173</sup> See *infra* Part IV(B)(3), discussing the analysis and coding of concordance lines in corpus-based analyses.

<sup>174</sup> Expert Report of Oliver B. Northrup, 2014 Misc. Filings LEXIS 4411 \*26 (“with 95% confidence in this interval”).

<sup>175</sup> *Moroccanoil, Inc. v. Marc Anthony Cosmetics, Inc.*, 2014 U.S. Dist. LEXIS 184585, at \*9 (C.D. Cal. Oct. 7, 2014) (“Dr. Steckel is a qualified expert in the field of statistics, and his opinions are adequately based on accepted scientific methodology and his experience as an expert in that field.”).

magazines, trade press, and other e-commerce websites).<sup>176</sup> This article agrees that Northrup’s corpus may have been biased by the possibility that most of its instances of “MOROCCANOIL” occurred in reviews of the Moroccan oil product itself, which would have overrepresented trademark usage.<sup>177</sup> Despite this shortcoming, Northrup’s report is groundbreaking in that it was conducted with substantial analytic rigor in comparison to other linguists’ corpus-based analyses, which have generally neither delineated forms of usage with as much granularity<sup>178</sup> nor utilized corpora with as many instances of the disputed term as Northrup did in *Moroccan oil*.

For example, in *Bon Vivant Catering, Inc. v. Duke University*,<sup>179</sup> in which the validity of the mark REFECTORY CAFÉ, in the context of dining facilities, was in dispute, an expert linguist retained by the defendant cited 108 instances of “refectory” in the *Corpus of Contemporary American English* as evidence that the mark was generic.<sup>180</sup> The expert’s testimony, however, did not mention the capitalization rate of the mark in the context of dining areas, only mentioning: “there are certainly many references to a dining area in a religious institution (e.g., Tibetan monastery, Catholic convent) or in a school setting (e.g., boarding school, institute).”<sup>181</sup> Other examples of corpus analyses in genericness litigation include that of Apple’s expert in its fight for the claimed mark APP STORE<sup>182</sup> and that of Solid 21’s expert in its fight for the

<sup>176</sup> Expert Report of Joel H. Steckel, PhD, 2014 Misc. Filings LEXIS 4413, at \*70 (further stating that “In addition, Dr. Northrup relies on a skewed sample of product reviews prepared by customers and not including industry participants, professionals, or advertisers, which given the availability of particular products on Amazon is likely to be centered on reviews of the Moroccan oil products specifically.”).

<sup>177</sup> *Cf. Eagle Snacks, Inc. v. Nabisco Brands, Inc.*, 625 F. Supp. 571, 577 (D.N.J. 1985) (discounting plaintiff’s evidence consisting of 4,000 consumer letters (sent to the plaintiff), wherein 80% to 90% of instances of “HONEY ROAST” were capitalized, because such letters were likely sent with the plaintiff’s product in hand, biasing their usage).

<sup>178</sup> Including, for example, margins of error.

<sup>179</sup> 2016 U.S. Dist. LEXIS 72524 (M.D.N.C. June 3, 2016).

<sup>180</sup> Expert Report of Sandra Ferrari Disner, PhD, 2015 Misc. Filings LEXIS 6487.

<sup>181</sup> *Id.* at \*7. Despite this and despite that the court made no comment on the corpus analysis, the court did not grant summary judgment in favor of the plaintiff due to the existence of genuine issues of material fact as to the validity of the trademark. *Bon Vivant*, 2016 U.S. Dist. LEXIS 72524, at \*50.

<sup>182</sup> *See* Declaration of Robert A. Leonard, PhD, in Support of Apple Inc.’s Motion for Preliminary Injunction, 2011 Misc. Filings Lexis 4952; Declaration of Dr. Robert A. Leonard in Opposition to Opposer Microsoft Corp.’s Motion for Summary Judgment, 2008 Misc. Filings LEXIS 13600. Apple eventually abandoned its case, resulting in APP STORE becoming generic. *See* Neil McAllister, *Apple Surrenders in ‘App Store’ Trademark suit against Amazon*, *The Register* (July 9, 2013). As of September 16, 2018, “App Store” is still registered. APP STORE, Registration No. 77525433. However, Amazon now uses “Appstore” generically. *Amazon Appstore for Android*, Amazon, <https://www.amazon.com/gp/feature.html?ie=UTF8&docId=1000626391>. *See* McCarthy, *supra* note 6, § 17:8 (noting that marks can become generic through abandonment if the

claimed mark “Red Gold.”<sup>183</sup> In both reports, however, the experts testify that the mark in question appears fewer than forty times in the chosen corpus, the *Corpus of Contemporary American English*, limiting the probative value of such analyses. And although Apple’s expert provided an analysis of 1,000 instances of “APP STORE” from a Nexis search, he did not delineate the method by which he categorized such instances as “related to Apple’s App Store service.”<sup>184</sup>

A slightly more rigorous assessment was proffered in *Elliot v. Google*.<sup>185</sup> The defendant’s expert linguist, although not explicitly discussing corpora, cited 973 instances of “GOOGLED” or “GOOGLING”<sup>186</sup> in Nexis search results, of which over 90% were capitalized, and rigorously defended his contention that such data should be probative of trademark validity.<sup>187</sup> Specifically, Nunberg provided useful guidance for interpreting the 90% capitalization rate given the medium from which it was drawn, explaining that one should expect “a higher proportion of capitalization in the largely edited prose of Nexis than on the Web.”<sup>188</sup> Nunberg’s report, however, could be further bolstered by a comparison of the capitalization rates for the word “GOOGLE” and other proper nouns in the data, the inclusion of margins of error, and more discussion of how representative the data is of the relevant population.

In addition to these corpus-based analyses, some scholars have proposed and some expert witnesses have administered corpus-like studies of marks using search engines. This method, most notably proposed by Lisa Larrimore Ouellette,<sup>189</sup> involves analyzing the order in which search results (instances of usage) appear.<sup>190</sup>

---

owner fails to police competitors’ use of the mark such that the public also starts using the mark generically).

<sup>183</sup> Declaration of Expert Ronald R. Butters, PhD in Support of Opposition to Defendants’ and Counter-claimant’s Motion for Summary Judgment, 2014 Misc. Filings Lexis 116. Issuing no opinion on the corpus-based analysis, the court concluded that RED GOLD was generic and dismissed with prejudice Solid 21’s claims that defendant infringed upon the RED GOLD mark. *Solid 21, Inc. v. Breitling USA Inc.*, 2011 U.S. Dist. LEXIS 81004, at \*14 (C.D. Cal. July 19, 2011).

<sup>184</sup> See Declaration of Dr. Robert A. Leonard in Opposition to Opposer Microsoft Corp.’s Motion for Summary Judgment, 2008 Misc. Filings LEXIS 13600, \*12.

<sup>185</sup> 45 F. Supp. 3d 1156 (D. Ariz. 2014), *aff’d*, 860 F.3d 1151 (9th Cir. 2017).

<sup>186</sup> See *infra* Part IV(B)(2), discussing why the capitalization rate of verbal forms of a disputed mark can be relevant for gauging genericness.

<sup>187</sup> Report of Geoffrey Nunberg, Ph.D., 2013 Misc. Filings LEXIS 1211, at \*85. See *infra* Part IV(B)(1), discussing the probative value of capitalization.

<sup>188</sup> Nunberg stated that he found a lower, 83%, capitalization rate in a search of “randomly accessed web documents.” *Id.*

<sup>189</sup> See Ouellette, *supra* note 13.

<sup>190</sup> Searched terms, for example, of the MICROTHIN mark (registered by Kimono condoms), included “MICROTHIN,” “MICRO-THIN,” “MICROTHIN condoms,” and “MICRO-THIN condoms.” See Ouellette, *supra* note 13, at 374 (citing *Church & Dwight Co. v. Mayer*

Specifically, Ouellette's work indicates a correlation between trademark usage in the top ten Google search results and trademark validity, suggesting that if trademark usage of a searched mark appears two times or fewer in the top ten Google search results, then that mark is more likely to be generic and unprotectable, and if trademark usage appears eight to ten times, then it is more likely to be a strong or famous mark.<sup>191</sup> Even if this correlation holds true in all circumstances, the primary issue with proffering search results in court is noted by Ouellette: "The exact algorithm used to produce Google results is a complex trade secret."<sup>192</sup> Since the process by which search engines order results is not transparent and, therefore, such analysis is neither replicable<sup>193</sup> nor able to be cross-examined, it would seem challenging for courts to recognize ordering as probative of genericness.<sup>194</sup> As other scholars of linguistics and the law have noted, companies optimize their advertisements and web pages so that they appear toward the top of search results, rendering such analyses biased in that they may not be reflective of consumers' recognition of the mark in question.<sup>195</sup> Although the analysis of search engine results suffers from these shortcomings, litigants have proffered search engine results as evidence in genericness cases, in some circumstances resulting in criticism from the court.<sup>196</sup>

Despite the presentation of corpus-based and corpus-like evidence in such genericness suits, there is a limited amount of

---

Labs., Inc., 868 F. Supp. 2d 876 (N.D. Cal. 2012), vacated in part, 2012 WL 1745592 (N.D. Cal. May 16, 2012)).

<sup>191</sup> *Id.* at 379.

<sup>192</sup> *Id.* at 363.

<sup>193</sup> This was noted by Justice Lee in *Rasabout*. 356 P.3d 1258, 1280 (Utah 2015) ("... there is no guarantee that the same search performed at another time on another computer will generate identical results.") (citing Adam Kilgarriff, *Googleology Is Bad Science*, 33 *Computational Linguistics* 147 (2007)).

<sup>194</sup> Such a claim that the search order itself is probative of genericness would be subject to rejection as hearsay unless an expert provided specific insight into the respective search engine's algorithm. *See* Rule 802. The Rule Against Hearsay, FRE Rule 802.

<sup>195</sup> *See, e.g.*, Kilgarriff, *supra* note 105, at 111 ("Much of the advertising has been written not so that people will read it but so the search engine's robots (Google's in particular) see it, and this supports the Search Engine Optimization (SEO) for the product."); McCarthy, *supra* note 6, § 20:126.50 ("Search engine results pages 'have little probative value' because of the lack of context to show how the search term was used on the listed web pages").

<sup>196</sup> *See, e.g.*, *In re Bayer Aktiengesellschaft*, 488 F.3d 960 (Fed. Cir. 2007), the trademark examiner cited a report on the top ten Google search results for "ASPIRINA," which was criticized by the court. 488 F.3d at 966-67 (not finding the examiner's evidence of Google search results probative, largely due to their giving "little context to discern how a term is actually used on the webpage"). While this article agrees that the search results themselves, i.e., the links with which one navigates Google's interface, may generally provide little context, the primary issue with presenting only the ten top Google search results is that it is quite unclear whether those particular results are representative of the average consumer's use of the mark in question.

literature on such evidence. The most notable piece of scholarship<sup>197</sup> on corpora in genericness litigation is that of corpus linguist Adam Kilgarriff, whose article in the *Journal of the Dictionary Society of North America* discusses the usefulness of corpora in trademark disputes.<sup>198</sup> Specifically, Kilgarriff's article contributes the concept of capitalization-based measures of a disputed mark and the importance of comparing a mark's capitalization rate to the capitalization norms in a language.<sup>199</sup>

### *B. The Case for Corpora*

Corpora are simply compilations of record evidence, which, as previously discussed, is already frequently accepted by the courts. The principal advantage of corpora is that when they are designed to be representative of the language of the population of interest, neither generic nor trademark uses should be disproportionately likely to be included in the analysis, unlike typical record evidence, which may be handpicked so as to overrepresent one kind of usage. When corpora are large enough, they can be used to make meaningful, quantitative measurements of usage of a disputed mark. Further, corpora are analyzed in accordance with principles of corpus linguistics (e.g., capitalization rates of a specific mark are not taken out of context but rather are compared with the capitalization rates of all language in a given dataset; data are gathered from sources containing usage by or usage indicative of the understanding of potential consumers of the product whose mark is in question, etc.).

In addition to these characteristics, corpus-based analyses have other distinct advantages that may capture the attention of courts, including the relative ease of their replication, the ability to capture language in natural rather than artificial environments, and cost effectiveness. Even if a court does not recognize these potential advantages of corpora to the extent this article does, corpora should

---

<sup>197</sup> One other scholar, linguist Pi-Chan Hu, employed a limited corpus-based analysis in an evaluation of the mark MISS WU. Hu used the Corpus of Contemporary American English to evaluate the mark, despite the fact that there was only one instance of "MISS WU" in the corpus, and concluded that the USPTO registered MISS WU because "Wu" was a rare surname in the United States. See Pi-Chan Hu, *A Linguistic Study of the Distinctiveness of a Trademark*, 3 NTUT J. of Intell. Prop. L. & Mgmt. 1-20 (2014). Additionally, although not pertaining to the analysis of generic trademarks specifically, an article by Rebecca Tushnet on trademark dilution uses corpora to measure the frequency with which proposedly famous marks are used in a language. See Rebecca Tushnet, *Gone in Sixty Milliseconds: Trademark Law and Cognitive Science*, 86 Tex. L. Rev. 507 (2008).

<sup>198</sup> See Kilgarriff, *supra* note 105.

<sup>199</sup> *Id.* at 105.

still at least be admissible, in that, at the very least, they are just compilations of record evidence.<sup>200</sup>

### 1. Large Sample Sizes

The immense sizes of current corpora and the ability to quickly and inexpensively harvest large quantities of data with which to create new corpora are among the most significant advantages of the method.<sup>201</sup> Further, the sheer quantity of textual data, including that from different periods, allow for corpus-based analyses of mark usage taking place at moments other than the present, should such an analysis be necessary.<sup>202</sup> In contrast, it can be quite expensive and time consuming to survey a statistically significant<sup>203</sup> number of people, let alone to gather survey data from a time period other than the present moment.

Even surveys that include “relatively large samples,”<sup>204</sup> such as that influencing the court in *E.I. DuPont* (1,031 respondents),<sup>205</sup> can be easily surpassed in sample size by corpora. Among just three corpora, Brigham Young University’s *NOW Corpus*,<sup>206</sup> *Corpus of Contemporary American English (COCA)*,<sup>207</sup> and *The Wikipedia*

---

<sup>200</sup> Because corpus-based analyses must be conducted in accordance with scientific principles as surveys are, they would not be admissible without expert testimony. *See, e.g., M2 Software, Inc., a Delaware corporation v. Madacy Entertainment, a corporation*, 421 F.3d 1073, 1087 (9th Cir. 2005) (finding district court to have properly rejected M2 Software’s likelihood of confusion survey on the grounds that the survey designer did not qualify as an expert).

<sup>201</sup> *See Douglas Biber, Representativeness in Corpus Design*, 8 *Literary & Linguistic Computing* 243, 248 (1993) (discussing generally the inverse relationship between sample sizes and standard errors).

<sup>202</sup> *See, e.g., Classic Foods Intern. Corp. v. Kettle Foods, Inc.*, 468 F. Supp. 2d 1181, 1188 (C.D. Cal., 2007) (The relevant date for the determination of genericness, in evaluating whether plaintiff Classic Foods infringed upon Kettle Foods’ KETTLE mark, is the date (1999) that Classic Foods first entered the market using the disputed mark).

<sup>203</sup> *E.g., In re Softspikes, Inc.*, 2012 T.T.A.B. LEXIS 107 (T.T.A.B. 2012) (referring to a survey asking 25 respondents to classify “softspikes” as a generic term or as a trademark, “the sample size is so small and the universe is so narrowly defined that the results are not meaningful.”); *Zimmerman v. Nat’l Ass’n of Realtors*, 70 U.S.P.Q.2d (BNA) 1425, 1435 (T.T.A.B. 2004) (referring to the petitioner’s survey, “The low response rate (~10%) negatively affects the reliability of the survey, as the sample may well no longer be representative of the relevant consuming universe. The final number of survey subjects (96) was too low to accord much weight to the study results.”).

<sup>204</sup> Jay, *supra* note 21, at 1154.

<sup>205</sup> McCarthy, *supra* note 6, § 12:16 (stating that the *Teflon* survey “was conducted of 514 men and 517 women on the evening of June 4, 1973, in 20 cities”), *quoted in* Jay, *supra* note 21, at 1154 n.152.

<sup>206</sup> *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (531 results, last visited Aug. 18, 2018).

<sup>207</sup> *Corpus of Contemporary American English*, Brigham Young University (1990–2015), <http://corpus.byu.edu/coca/> (498 results).

*Corpus*,<sup>208</sup> for example, there are over 1,837 instances of the word “TEFLON” (case insensitive) in American English. Such a cursory<sup>209</sup> search can be conducted in seconds using BYU’s website; featuring more corpora drastically increases the number of results.<sup>210</sup> A search of “TEFLON” (case insensitive) on Google Book’s Ngram Viewer, for example, yields over ten thousand results for the years between 1960 and 2010.<sup>211</sup> In comparison to the majority of surveys, which do not feature samples as large as that proffered in *E.I. DuPont*, corpora are even more robust.<sup>212</sup>

Despite this, survey sample sizes are rarely so small as to render their results meaningless, and, furthermore, while corpus results may be plentiful, they may often yield results that must be excluded from analysis.<sup>213</sup> For example, an analysis of the APPLE trademark must exclude instances referencing the fruit, and for the analysis of any mark’s capitalization rate, sentence-initial instances must be removed. In some circumstances, especially for certain polysemous marks, it may either be difficult to exclude such instances from the analysis, or the removal of such instances may result in a small remaining sample, potentially rendering the corpus analysis less useful than a survey. If corpus analyses can yield large samples for analysis, however, they doubly benefit from ease of replicability as well as from low cost of administration.

## 2. Replicability and Limited Human Element

As a result of the permanence of textual data, corpus-based analyses can be easily replicated<sup>214</sup> and may be conducted by just

---

<sup>208</sup> *The Wikipedia Corpus*, Brigham Young University (2014), <http://corpus.byu.edu/wiki/> (808 results).

<sup>209</sup> This search was conducted for the purpose of exemplifying the size of corpora rather than the methodology by which a rigorous corpus analysis is conducted. Factors such as the date of publication of the words in the corpus are important and relevant to corpus-based studies.

<sup>210</sup> Such an increase would be marginally offset by uses of the word “TEFLON” outside of the context of synthetic resin.

<sup>211</sup> *Google Books (American) Corpus*, Brigham Young University (1960–2000), <http://googlebooks.byu.edu/x.asp> (11,271 results). One downside of using the *Google Books Corpus* is that it is so large (155 billion words) that it is nearly impossible to download to a computer for analysis. Nonetheless, one can still use the Google Books Ngram Viewer webpage to count both uppercase and lowercase instances of a word, which is still very valuable for the analysis of a mark, especially given the sheer size of the *Google Books Corpus*. Google Books Ngram Viewer, <https://books.google.com/ngrams>.

<sup>212</sup> In *Zimmerman*, for example, the survey proffered by the respondent, National Association of Realtors, featured only 204 responses. *Zimmerman v. Nat’l Ass’n of Realtors*, 70 U.S.P.Q.2d 1425, 1435 (T.T.A.B. 2004).

<sup>213</sup> See *infra* Part IV, discussing the methodology of corpus-based analyses, including circumstances where corpus results must be excluded from analysis.

<sup>214</sup> See *State v. Rasabout*, 356 P.3d 1258, 1281 (Utah 2015) (Lee, J.) (“COCA is also completely transparent, and it generates search results that are easily replicable.”).

one qualified expert. While the human element of some surveys is beneficial for asking follow-up questions and gaining clarification of the respondent's understanding, the questioning, recording, and sometimes coding of all the individual interviews comprising a survey may require significant human resources and can expose surveys to personnel- and tabulation-related errors.<sup>215</sup> This issue is compounded by the difficulty in replicating an individual survey interview to ensure that the respondent's answers were recorded or coded correctly. The capitalization rate of a word in a corpus, however, can be re-calculated with great ease by opposing counsel, opposing experts, and courts in order to ensure that no human errors were made. Even manual classifications of instances of a term in a corpus can be easily replicated by the opposing party or by the court; a dataset of language usage, unlike the spoken words of an interview, are not ephemeral. Generally, the only circumstances under which additional personnel could be required for corpus-based analyses is if it is necessary to manually analyze and classify a very large number of instances of the disputed mark in the corpus.<sup>216</sup> But even these more personnel-heavy corpus analyses are at least replicable such that there is no reason for tabulation-related errors to go unchecked at trial.

### 3. Cost Effectiveness

It is no secret that surveying is generally very expensive, costing anywhere from \$30,000 to more than \$100,000.<sup>217</sup> Considering that survey experts can demand hourly rates of over \$450 and require support staff billing over \$200 per hour, in addition to the cost of resources such as call centers and research facilities, this estimate is unsurprising.<sup>218</sup> Corpus-based analyses, on the other hand, do not require multiple interviewers, a large team of support staffers, or a

---

<sup>215</sup> See *supra* Part II(A)(3), discussing personnel and tabulation related errors.

<sup>216</sup> Moroccanoil's expert witness, Oliver Northrup, for example, manually reviewed and coded 300 instances of the disputed term as either trademark, descriptive, or generic uses. Expert Report of Oliver B. Northrup, 2014 Misc. Filings LEXIS 4411, \*22 ("Given these variables which an automated approach could easily overlook, I manually classified uses of the term in a random sample of 300 reviews containing it."); see *supra* Part III(A)(2). Considering Northrup's highly detailed manual classification of 300 instances, there would likely have to be significantly more instances for it to be necessary to retain additional experts/personnel for the purposes of coding individual uses.

<sup>217</sup> See generally, e.g., Robert C. Bird & Joel Steckel, *The Role of Consumer Surveys in Trademark Infringement: Empirical Evidence from the Federal Courts*, 14 U. Pa. J. Bus. L. 1013 (2012). See also Poret, *supra* note 46, at 757 (noting the prevalence of instances in the case law where expensive, time-consuming surveys were administered and but given little weight by the court due to methodological errors).

<sup>218</sup> See Robert H. Thornburg, *Trademark Survey Evidence: Review of Current Trends in the Ninth Circuit*, 21 Santa Clara Computer & High Tech. L.J. 715, 717 (2005).

research facility. Rather, a computer and a qualified expert are sufficient for such analyses.<sup>219</sup>

The replicability of corpus-based studies and the convenience of harvesting textual data also contribute to their cost-effectiveness. When mistakes in a corpus study are brought to light, one can easily review the data and correct them. As a result of the ephemeral nature of an oral survey interview, however, it is much more costly to reconcile certain flaws, because it may be necessary to conduct an entirely new survey. The availability and permanence of textual data also make it possible to efficiently target a specific universe of consumers, such as purchasers of hiking equipment, by either using an already existent corpus of hiking blogs and magazines or by creating such a corpus. By contrast, with surveys, it may be necessary to expend time asking respondents a qualifying question such as, “Have you or do you plan on purchasing hiking equipment?” to ensure that the respondent is in the universe of potential consumers.<sup>220</sup>

#### 4. Probative Value of Usage Over Elicited Understanding

Given that surveys measure elicited understanding of a mark while corpora and record evidence measure usage of a mark, the question that remains is which measure provides a better gauge of the “primary significance of the term in the minds of the consuming public”?<sup>221</sup> To the extent trademark law is concerned with the “primary significance,” or *meaning*,<sup>222</sup> of words, perhaps it should follow the guidance of lexicographers,<sup>223</sup> who examine usage, often

---

<sup>219</sup> Of note, one expert linguist testified to a rate of \$250 per hour. See Expert Report of Oliver B. Northrup, 2014 Misc. Filings Lexis 4411 (“I am being compensated in this case at an hourly rate of \$250.”).

<sup>220</sup> In *Windsurfing Int’l*, discussed *supra* in Part II(A)(3), recall that it was necessary for surveys to ask such a qualifying question, “Now, please tell me what kind of products these are?” to ensure that respondents were familiar with the product and could be potential purchasers. *Windsurfing Int’l, Inc. v. Fred Ostermann GMBH*, 613 F. Supp. 933 (S.D.N.Y. 1985), *vacated*, 828 F.2d 755 (Fed. Cir. 1987), *quoted in* Jay, *supra* note 21, at 1135 n.65.

<sup>221</sup> 15 U.S.C. § 1064.

<sup>222</sup> “Meaning” is frequently cited by the courts in litigation concerning genericness. See, e.g., *OBX-Stock, Inc. v. Bicast, Inc.*, 558 F.3d 334, 339-40 (4th Cir. 2009) (“[W]ords that do not identify goodwill attached to products or product sources but rather are used for their common meaning or meanings not indicative of products and product sources.”), *quoted in* McCarthy, *supra* note 6, § 1:27.

<sup>223</sup> Lexicographers have earned the respect of some courts as scientists concerned with the meaning of words. See, e.g., *Mead Johnson & Co. v. Abbott Labs.*, 201 F.3d 883, 886 (7th Cir. 2000) (“So far as we can tell, however, never before has survey research been used to determine the meaning of words, or to set the standard to which objectively verifiable claims must be held. Dictionaries themselves are a form of survey; lexicographers determine how words have been used in both scholarly and popular texts. But philologists and others who contribute to dictionaries devote their lives to discovering

through corpora, to define words.<sup>224</sup> Lexicographers create dictionaries in a descriptive manner, in that they seek to reflect the actual usage of language rather than to prescribe language use to their readers.<sup>225</sup> Since users of language “share an interest in effective and efficient communication,”<sup>226</sup> instances of language use can be assumed to be correct and indicative of language users’ understanding of a word.

A useful thought experiment for considering whether usage or elicited understanding should be the gauge of genericness is the case of a word for a product made by only one producer.<sup>227</sup> Asking consumers the name they would request in a purchase encounter does not fully capture whether the word is generic, because the brand and the product are essentially one and the same for unique products; likewise, asking consumers to classify the mark in question as a generic product or as a trademarked brand name is problematic, because it forces to consumers to make what may be a false dichotomy.<sup>228</sup> Further, as a result of demand characteristics associated with such a question, respondents may feel influenced to give what they perceive is the correct answer in the eyes of the interviewer.<sup>229</sup>

A more accurate way to back into the meaning of the word to the consuming public is to examine consumers’ actual usage of the term<sup>230</sup>—how frequently do consumers capitalize the mark in question compared with baskets of other proper and common nouns in the given medium? Do they frequently use the word to unambiguously identify a type of product rather than to signify the proper name of a specific brand? What other words do consumers use to denote the type of product? Such usage-based measurements

---

usage and interpreting nuance. It would be a bad idea to replace the work of these professionals with the first impressions of people on the street, especially because consumers’ sketchy understanding of science means that survey results are apt to present firms with unrealistic demands for verification.”).

<sup>224</sup> Lexicographers have used corpora since at least 1984. Landau, *supra* note 16, at 280 (citing the *Corpus of Present-Day Edited American English*, among others, as valuable for ascertaining the way language is used for creating definitions).

<sup>225</sup> See, e.g., Landau, *supra* note 16, at 32 (“All dictionaries based on usage—and all competently done dictionaries must be based on usage—are descriptive. Prescription is impossible to distinguish from bias.”).

<sup>226</sup> Heymann, *supra* note 6, at 1323.

<sup>227</sup> See *supra* Part II(A)(1), discussing why it can be challenging for purchase encounter surveys to adequately gauge the genericness of marks for unique products.

<sup>228</sup> “There is no reason why a given term, such as ‘teflon,’ cannot coexist as both a brand and a product (or something else) in consumers’ minds.” Timothy Green, *Trademark Hybridity and Brand Protection*, 46 Loy. U. Chic. L.J. 75, 110 (2014).

<sup>229</sup> See *supra* Part II(A)(2), discussing how classification surveys may be prone to demand characteristics.

<sup>230</sup> The ideal corpus consists of the language of representative consumers. See *infra* Part IV(A).

reveal more about the consumer's (perhaps subconscious) understanding of the meaning of a mark than do survey questions that require respondents to make a binary distinction that may not accurately represent their nuanced understanding of the mark. Unlike surveys, which capture the aggregate of all respondents' obligatorily discrete classifications of a mark, corpora can measure the full spectrum of consumer use of a mark, providing a more comprehensive picture of public consensus on the meaning of the word. For example, an individual respondent may, when forced to choose, classify a mark as generic but still be able recognize the term as a brand differentiator in a purchasing situation. Likewise, a respondent may categorize a word as a trademark but predominantly recognize the mark as generic because of a lack of alternative generic terms for the given product.

On top of the factors discussed above, with usage-based evidence, one can generally be assured that the recorded language occurred naturally and, therefore, reflects the mindset of the speaker. Contrarily, the artificial environments in which surveys are conducted can affect respondents' answers through demand characteristics, question ordering, and other factors to which the language reflected in record evidence and corpora is generally not exposed. This is not to say that usage is not exposed to any external factors; published articles are written in accordance with editing policies that may have strict requirements regarding the use of trademarks; likewise, autocorrection algorithms may artificially affect usage by capitalizing words that the writer specifically did not intend to capitalize. But the difference between these examples and that of a survey environment is that with a survey, there is always someone, the surveyor, to whom the respondents may tune their answers.

#### IV. METHODOLOGY OF CORPUS ANALYSIS IN GENERICNESS LITIGATION

The first step of a corpus-based analysis is choosing an existing corpus or creating a new corpus featuring or reflecting the language of consumers and potential consumers. Once an appropriate dataset is chosen, there are a few methods by which to analyze its contents. Automated<sup>231</sup> methods include calculating the capitalization rate of the disputed term, analyzing collocates of the disputed term, and examining frequencies of alternative generic terms and competing brands' trademarks. In addition to these methods, it may be useful, especially for polysemous marks, to use concordance lines to manually categorize uses of a mark as trademark, generic, or

---

<sup>231</sup> "Automated" means it is may be less necessary to read and classify manually each individual instance of usage.

extraneous.<sup>232</sup> After selecting or creating a corpus, this article recommends first examining the capitalization rate of a term; if the capitalization rate alone is inconclusive of the mark's status,<sup>233</sup> it becomes especially important to survey the existence and frequency of alternative common names for the product in question and to conduct an analysis of words that habitually precede or follow the mark in question.

### A. Corpus Selection

The selection of the appropriate corpus is similar to the selection of an appropriate universe for consumer surveys<sup>234</sup>—the ideal corpus consists of the language of individuals who are potential purchasers of the product whose trademark is being disputed,<sup>235</sup> and it does not overrepresent the language of one individual or institution. The corpus data should also have been created (typed/written) between the date on which the original trademark owner entered the market and the current date of the litigation.<sup>236</sup>

Since most products' purchasers are likely to constitute a subset of the population whose language use exists in some form on the web, whether in blogs or in specific publications, it is quite feasible either to find or create a corpus representing the language of potential consumers.<sup>237</sup> For the mark REALTOR, for example, an expert could create a corpus consisting of language harvested from magazines such as *The Real Estate Professional* and blogs such as *Market Leader Blog*.<sup>238</sup> As real estate professionals, users of the mark REALTOR in such contexts would reasonably be purchasers

<sup>232</sup> See *supra* note 134 on concordance lines.

<sup>233</sup> A capitalization rate between 40% and 70% may be regarded as inconclusive. See *infra* Part IV(B)(3), discussing how to interpret capitalization rates.

<sup>234</sup> See generally James C. Phillips & Jesse Egbert, *Advancing Law and Corpus Linguistics: Importing Principles and Practices from Survey and Content-Analysis Methodologies to Improve Corpus Design and Analysis*, 2017 BYU L. Rev. 1589 (2017) (discussing surveying techniques relevant to corpus analyses).

<sup>235</sup> The ideal universe of language users for corpus analyses is the same as the appropriate universe for a survey measuring usage of that mark, i.e., potential purchasers of the product whose trademark is being disputed. McCarthy, *supra* note 6, § 32:159.

<sup>236</sup> See McCarthy, *supra* note 6, § 12:17.50 (“Three possibilities for the critical date when genericness is to be tested are (1) the date when the proponent of trademark rights first began the alleged trademark usage; (2) just before the date when the accused infringer first entered the market; and (3) the current date of the litigation. If survey has been taken at a date different from that found to be the critical date, this should only affect the weight, not the admissibility, of the survey.”).

<sup>237</sup> For more information on ensuring representativeness when creating corpora, see Phillips & Egbert, *supra* note 234 and Biber, *supra* note 201.

<sup>238</sup> *Zimmerman v. Nat'l Ass'n of Realtors*, 70 U.S.P.Q.2d 1425, 1435 (T.T.A.B. 2004) (finding real estate brokers and agents to be an appropriate universe for respondent's survey, because they are “actually the purchasers or prospective purchasers of membership in respondent and the services provided by respondent”).

or potential purchasers of membership in the National Association of Realtors.

For products or services that are or can be habitually consumed by the majority of the speaking public, it may not be necessary to use or create a specific corpus. For example, the universe of potential consumers for a trademarked product such as KLEENEX tissues could be defined as “people who blow their noses and use facial tissues when they do so,” which is obviously so broad that anyone using the word is very likely a consumer or potential consumer of KLEENEX products. Other trademarked products, such as COCA-COLA soft drinks, CHAPSTICK lip balm, THUMBDRIVE USB flash drives, PING-PONG table tennis, BUBBLE WRAP cushioning product, and BAND-AID adhesive bandages also fit this bill. Given the relatively low price of such products and the fact that anyone may feasibly purchase them, mere usage (and, therefore, implied awareness) of these words very likely indicates that the speaker (or writer) is a potential consumer of the product. For these types of marks, it may be appropriate to use more general corpora, such as the *GloWbe’s blog-specific U.S. sub-corpus*, which contains over 130 million words from over 48,000 different blogging websites.<sup>239</sup>

Regardless, all proffered corpus-based analyses must be supported by expert testimony discussing their representativeness, method of design, and margins of error<sup>240</sup> associated with calculations.<sup>241</sup> Specifically, experts should be prepared to address whether the corpus is a probability or non-probability sample,<sup>242</sup> how findings are limited if the corpus is a non-probability sample,<sup>243</sup> and the number of distinct texts from which analyzed instances are drawn (e.g., 90% of instances of a mark could be capitalized, but all of those instances could theoretically come from a single book or text that is included in the corpus).<sup>244</sup>

---

<sup>239</sup> *Corpus of Global Web-Based English*, Brigham Young University (2012–2013), <http://corpus.byu.edu/glowbe/>.

<sup>240</sup> See Biber, *supra* note 201, at 248 (discussing sampling size and the calculation of standard error).

<sup>241</sup> See *Daubert v. Merrell Dow Pharms., Inc.*, 113 S. Ct. 2786, 2790 (U.S. Cal. 1993) (setting forth the “Daubert Standard” regarding the admissibility of expert testimony, noting specifically that such testimony should include the “known or potential error rate”).

<sup>242</sup> Most if not all corpora would be categorized as non-probability samples, primarily because corpus designers would not have access to spoken language. See Phillips & Egbert, *supra* note 234, at 1596.

<sup>243</sup> See *id.* (discussing how findings from non-probability samples cannot be “generalized beyond the scope of what the sample actually represents”).

<sup>244</sup> This can be mitigated by counting only one instance for each source included in a corpus (e.g., if a corpus consists of 10 books and a total of 100 instances of the searched term that are unevenly distributed among the books, the predominant usage in each book would be equally weighted). See Phillips & Egbert, *supra* note 234, at 1605.

## 1. Probative Value of Consumer Language versus Media Language

Generally, it should be noted that since most people do not publish in newspapers, academic journals, or magazines, corpora consisting primarily of such published language should be appropriately discounted when considering their representativeness of a broader target population.<sup>245</sup> In addition to the fact that professional writing may not always be indicative of consumer usage or consumer understanding,<sup>246</sup> writing in such publications may be influenced by editorial policies structured to avoid irritating trademark owners.<sup>247</sup>

These shortcomings can partially be offset by the proposition that professional writing such as that of the media is so widely read that it influences consumers' understanding and usage of the word, such that professional usage and consumer usage of a term should not differ substantially. Additionally, it can be argued that publishers have an interest in communicating clearly with their readers, which includes using trademarked words in ways that are sensible to their audiences. While some scholars have argued that media and even other, non-purchase-encounter uses should not be considered probative of genericness,<sup>248</sup> the ways consumers use language, even if not at the cashier's counter, is strongly indicative of their understanding of the term and their ability to use the mark in purchasing encounters. Taking this concept to the extreme, if there is not a single instance of a term indicating trademark usage in a corpus that is perfectly representative of American English communication, then it is reasonable to conclude that the typical consumer is not equipped to use the trademarked sense in a purchasing encounter.

Thus, although the ideal corpus consists of the natural language of non-media consumers, corpora consisting of media and other published language could still be considered probative. At their very least, media-based corpora, such as BYU's *NOW Corpus* (over 800

---

<sup>245</sup> See Phillips & Egbert, *supra* note 234, at 1603.

<sup>246</sup> See Irani, *supra* note 101, at 612 ("What is troubling about the tendency to focus on newspapers and publications of general circulation is the significant logical leap from printed publications to a conclusion about consumer understanding.").

<sup>247</sup> See, e.g., Kilgarriff, *supra* note 105, at 103 (noting that professional writing "is sometimes less good evidence of the language than amateur writing" due to professionals' adherence to style guides obliging them to capitalize registered trademarks).

<sup>248</sup> See, e.g., Deven R. Desai & Sandra L. Rierson, *Confronting the Genericism Conundrum*, 28 Cardozo L. Rev., 1789, 1836 ("noncommercial uses of a trademark are, however, poor barometers of the consumer's perception of the mark in commercial contexts"). While Desai and Rierson argue that people do not decide how to use a word by "by blindly mimicking the way a reporter used the word in *Newsweek* or *The New York Times*," the same could be said of the language of anyone else. However, the difference between a single individual's language in a purchasing encounter and the language of an author for *The New York Times* is that significantly more people engage with the latter.

million words from U.S. news outlets since 2010), are simply large compilations of media-based record evidence, which has been frequently accepted by the courts.<sup>249</sup> For this reason, if no existing corpora feature the language of the relevant consumer group, and if it is unfeasible to create such a corpus, the case law implies it may be acceptable to use a generalized corpus composed of textual data from books, magazines, and newspapers for a corpus-based analysis of genericness.<sup>250</sup> Corpora consisting of language from social media, blogs, and other mediums in which the average consumer more frequently produces language, however, are generally better equipped to serve the purposes of corpus-based analyses of most disputed marks. Such corpora are more representative of the consuming population and consequently should hold more weight in court.

## 2. Corpus Selection Strategies in Litigation

An important consideration for attorneys and experts is how an opposing party may use a corpus to benefit its own argument. For example, it may favor the party arguing for genericness to use a broader corpus featuring the language of many different types of consumers as opposed to a specific corpus featuring the language of a narrower set of sources that better reflects the universe of potential purchasers of the product whose mark is being disputed. Thus, in a case such as *Windsurfing International*,<sup>251</sup> the plaintiff, arguing for the validity of WINDSURFER as a mark for watercraft, hypothetically, might use a corpus composed specifically of language from a water sports blog, whose authors may be more familiar with the different brands of “sailboards”<sup>252</sup> and know that WINDSURFER was a trademark. The defendant, contrarily, might use a corpus composed of the language of a wider population of consumers, who may not be as familiar with water sports brands and who have perhaps only heard of the predominant “sailboard” name, “windsurfer,” such that they assume “windsurfer” was a

---

<sup>249</sup> See *supra* Part II(B).

<sup>250</sup> Courts' have historically accepted usage from newspapers and magazines as evidence of genericness or validity. See, e.g., *McCarthy*, *supra* note 6, § 12:13; *Official Airlines Guides, Inc. v. Goss*, 6 F.3d 1385 (9th Cir. 1993). Where courts have criticized the presentation of such evidence for being either cherry-picked or from a narrow range of documents (see *supra* note 118), corpora make up by being compiled and analyzed in accordance with scientific and statistical principles. As with the proffering of other forms of usage in genericness cases, hearsay objections may be overcome, as such materials would be submitted to demonstrate their content and the fact of publication rather than the truth of any underlying information conveyed therein. See *McCarthy*, *supra* note 6, §§ 20:125-126.50.

<sup>251</sup> 613 F. Supp. 933 (S.D.N.Y. 1985), *vacated*, 828 F.2d 755 (Fed. Cir. 1987).

<sup>252</sup> *Id.* at 954 (“WSI maintains that ‘sailboard’ is the generic name for the craft and that the generic terms for the sport and sportsman are ‘boardsailing’ and ‘boardsailor,’ respectively.”).

generic term for the product. Similarly, in reference to the litigation concerning Apple's APP STORE mark, a corpus composed of the language of *Macworld* readers, who would conceivably be quite familiar with Apple's trademark, would likely have been far more favorable to demonstration of trademark usage than would BYU's broader *GloWbe Corpus*.<sup>253</sup>

Another significant factor to be considered is whether the corpus was created for the purposes of the trial in the first place. In creating a corpus, an expert could potentially cherry-pick, or be accused of cherry-picking, the textual data comprising it so that it favors a certain pattern of usage. For example, recall that Northrup's corpus of Amazon reviews in *Moroccanoil* may have been biased because most capitalized instances of "MOROCCANOIL" were likely to occur in reviews of the Moroccanoil product specifically.<sup>254</sup> Thus, one significant benefit of using an already existent corpus is that it reduces the chance of bias in favor of one party. Such examples illustrate the significant influence corpus-selection can have on the outcome of analyses and on their probative value in the eyes of the court. For this reason, this first step in a corpus-based analysis is extremely important.

### 3. When Corpora Cannot Represent the Relevant Consumers

A significant limitation to the corpus-based approach is that there are certain marks for which it may be difficult to find textual data reflecting the appropriate universe of language users, even if such words are frequently used. This is the case for products not purchased by the public, such as escalators.<sup>255</sup> Merely using and being aware of the term "escalator" does not qualify one as a potential consumer of the product. For example, a search of "escalator" (case insensitive) on BYU's *GloWbe Corpus* yields 338 instances of the term in U.S.-based blogs;<sup>256</sup> however, it is unlikely that many of these bloggers are potential purchasers of escalators. For a word like "escalator," then, it may be quite difficult to find textual data produced by potential purchasers, and a survey might

---

<sup>253</sup> Such hypothetical corpus-based analyses were offered in neither *Windsurfing International* nor the APP STORE litigation.

<sup>254</sup> Expert Report of Oliver B. Northrup, 2014 Misc. Filings LEXIS 4411. *See also supra* note 176 (noting criticism of Northrup's report by the defendant's expert witness).

<sup>255</sup> Folsom & Teply, *Surveying "Genericness," supra* note 89 at 15 ("If a product—for example, 'escalators'—is not sold to the public at large, mere familiarity with the product will not suffice to place the respondent in the relevant consuming universe. Further questions must be asked to ascertain whether the respondent may be involved in purchasing escalators"). *See also* Haughton Elevator Co. v. Seeburger, 85 U.S.P.Q. (BNA) (Comm'r Pat. & Trademarks 1950) (finding "escalator" generic for moving stairways).

<sup>256</sup> *Corpus of Global Web-Based English*, Brigham Young University (U.S. only 2012–2013), <http://corpus.byu.edu/glowbe/>.

be the most viable source of evidence.<sup>257</sup> Since it may, in instances such as that of “escalator,” be difficult to ensure that language users are potential consumers of the trademarked product in dispute, corpus-based analyses are generally best suited for products and services that can be habitually consumed by the majority of the speaking public.

### **B. Capitalization**

As discussed in Part II(B), capitalization is a strong indicator, recognized by the courts,<sup>258</sup> that writers recognize a term as a proper name, or, as relevant here, a trademark.<sup>259</sup> Fortunately, it is very easy to measure capitalization rates in corpora by using simple regular expression<sup>260</sup> searches or using a graphical interface such as Sketch Engine software.<sup>261</sup> Before further delineating the capitalization method, it is important to note that measurements of capitalization should exclude sentence-initial capitalization and should generally include only instances of the mark that refer to the trademark and product-category in question (i.e., instances of “APPLE” in the context of fruit should not be included).<sup>262</sup>

#### **1. Probative Value of Capitalization Rates**

While the courts have recognized capitalization, or the lack of capitalization, as probative in genericness suits, it is important to discuss further the value of such evidence and why it should be given special weight, considering its centrality to a corpus-based analysis. Capitalization generally is a deliberate action signposting the writer’s recognition of a term as a proper name, of referring to a specific named entity rather than to a general one.<sup>263</sup> This is the

---

<sup>257</sup> Assuming it is possible to specifically target and survey potential purchasers of escalators (i.e., a mall intercept or generic telephone or Internet survey would likely not capture the relevant population).

<sup>258</sup> See *supra* note 102.

<sup>259</sup> See Kilgarriff, *supra* note 105 (“When we learn to write, we learn—and internalize—the conventions of the written language. For English, one of these conventions is “if it is a name, capitalize it.”).

<sup>260</sup> A regular expression is a string of characters used for searching files containing textual data. Regular expression searches can be made using the command-line utility known as EGREP (“Extended Global Regular Expressions Print”). See *generally Regular Expressions Quick Start*, Regular-Expressions.info (2017), <http://www.regular-expressions.info/quickstart.html>.

<sup>261</sup> See *supra* note 138.

<sup>262</sup> See *infra* Part IV(B)(2), however, discussing how metaphors can still be included in capitalization measurements.

<sup>263</sup> See *generally* Kilgarriff, *supra* note 105. Of note, this article is concerned with corpora consisting of typed textual data, although some corpora may be composed of recorded spoken language, which could be constructed through manual transcription or using

case even when the mark is not being used as an adjective modifying a common noun, as in “Kleenex facial tissues”; rather, even when a mark is being used as a verb (“I Googled” something”) or as an inflected noun (“I used Kleenexes”), capitalization should still be considered indicative of consumers’ brand awareness.<sup>264</sup>

While mere use of a mark as a verb should not be dispositive as to the genericness of a disputed mark,<sup>265</sup> the capitalization of such usage should be appropriately weighed. As verbs and common nouns are in all lowercase in English, the capitalization of marks in such contexts provides insight into consumer recognition of the trademarked sense of the word.<sup>266</sup> The converse, albeit, would not hold: lack of capitalization of verbal use does little to prove genericness, since verb use occurs outside the context of “product source” and “product category,” the analysis of which is required for a finding of genericness.<sup>267</sup> For example, people may use “google” indiscriminately to indicate the act of searching the Internet using any search engine or use “xerox” to refer to the indiscriminate act of copying documents, while still using the nounal and adjectival forms of “Google” and “Xerox” discriminately as trademarks to differentiate product sources. While the Ninth Circuit in *Elliot* did not specifically articulate such a view regarding capitalization, the decision does not necessarily rule out that verb use can ever be relevant to a determination of trademark validity; specifically, the majority stated:

Contrary to our colleague’s suggestion, we do not hold that generic verb use is “categorically irrelevant.” However,

---

speech recognition tools. Naturally, spoken language corpora cannot be used to evaluate capitalization rates.

<sup>264</sup> This argument is partially corroborated by the district court’s opinion in *Elliot*, where it was found that the mark GOOGLE could distinguish a particular brand of service, even if it is being used as a verb. *Elliot v. Google Inc.*, 45 F. Supp. 3d 1156, 1162 (D. Ariz. 2014) (“Verb use of a trademark is not fundamentally incapable of identifying a producer or denoting source. A mark can be used as a verb in a discriminate sense so as to refer to an activity with a particular product or service, e.g., ‘I will Photoshop the image’ could mean the act of manipulating an image by using the trademarked Photoshop graphics editing software developed and sold by Adobe Systems. This discriminate mark-as-verb usage clearly performs the statutory source-denoting function of a trademark.”), *aff’d*, 860 F.3d 1151 (9th Cir. 2017).

<sup>265</sup> *Elliott v. Google, Inc.*, 860 F.3d 1151, 1157 (9th Cir. 2017) (“verb use does not automatically constitute generic use.”).

<sup>266</sup> See Report of Geoffrey Nunberg, PhD, 2013 Misc. Filings LEXIS 1211 (“Searches show that when used as a verb, Google is overwhelmingly capitalized in news stories (90 percent of the time) and on the web (83 percent). Thus people are highly aware of the association of the verb with the Google trademark.”).

<sup>267</sup> See 860 F.3d 1151, 1156 (9th Cir. 2017) (“... a trademark only becomes generic when the ‘primary significance of the registered mark to the relevant public’ is as the name for a particular type of good or service irrespective of its source.”). This article contends that in the relevant context, wherein the mark is used to refer to a name or type of product, capitalization can be probative of both validity and genericness when compared with the habitual capitalization rates in the medium.

evidence that a mark is used in a generic sense in one particular setting cannot support a finding of genericide when it is unaccompanied by evidence regarding the primary significance of the mark as a whole.<sup>268</sup>

Under the Ninth Circuit's guidance, this article contends that the discriminate (i.e., generally capitalized) use of a verb,<sup>269</sup> can be appropriately weighed as probative of trademark validity. Likewise, the capitalization of marks in metaphorical contexts can also be considered indicative of trademark validity. In examples such as "the Cadillac of its class" and "put a Band-Aid on the situation,"<sup>270</sup> the capitalization of "Cadillac" and "Band-Aid" indicates the consumer recognizes the terms as proper names, even though they are used analogically.

Therefore, a corpus search made for the purposes of calculating capitalization rates should delineate usage of the term as an inflected noun, as a verb, and as a metaphor. For the word "GOOGLE," for example, a corpus analysis of capitalization should include the rates for forms such as "GOOGLES," "GOOGLED," and "GOOGLING." The capitalization of such inflections and derivations conveys information about writers' understanding of "GOOGLE" as a term that should be denominated as a proper name as a result of its being a trademark and, therefore, is useful for the consideration of its status as a mark.<sup>271</sup> Similarly, for an analysis of the mark ROLLS ROYCE or CADILLAC, metaphorical uses such as "the Rolls Royce of computers" or "the Cadillac of beers" should be included.

---

<sup>268</sup> *Id.* at 1160 (partially responding to J. Watford's concurring opinion, stating that "verb use could potentially be relevant in deciding whether a trademark has become the generic name for a type of good or service.").

<sup>269</sup> So long as the use of the verb is related to use of the product itself, as in "Googling."

<sup>270</sup> See generally Heymann, *supra* note 6, at 1335 ("And yet, reasonable speakers not only engage in such metaphorical wordplay often but can do so without losing sight of the mark's original naming function.").

<sup>271</sup> The capitalization rates of each form, however, should be listed separately from (and in addition to) a single capitalization rate incorporating all forms of usage to increase the transparency and replicability of such analyses. In *Elliot*, for example, an expert would be well served to have delineated the capitalization rates of each of the four distinct meanings of the word "GOOGLE" that were deemed relevant by the court. 45 F. Supp. 3d at 1173-74 ("The word google has four possible meanings in this case: (1) a trademark designating the Google search engine; (2) a verb referring to the act of searching on the Internet using the Google search engine; (3) a verb referring to the act of searching on the Internet using *any* search engine; and (4) a common descriptive term for search engines in general. The '502 and '075 marks are subject to cancellation *only* if the fourth meaning is the *primary* significance of the word google to a majority of the consuming public."), *aff'd*, 860 F.3d 1151 (9th Cir. 2017).

## 2. Inclusion of Anomalous Forms in Capitalization Rate

In addition to considering inflections, derivations, and metaphors, a well-informed corpus-based analysis should also consider how to treat variations such as the use of spaces and hyphens in composite marks, capitalization in marks such as iPod, and misspellings. These are important to include in an analysis, because consumers may employ such variations while still having a strong understanding of mark's status as a trademark. For example, consumers who type "Facebook" as "Face-Book" or "FaceBook" likely still have a very strong recognition of FACEBOOK as a trademark, as evidenced by their capitalization of their misspellings.<sup>272</sup> Similarly, consumers who spell "iPod" as "Ipod" or "IPod" and consumers who spell "Mattress Firm" as "Matress Firm" or "Mattres Firm" are nevertheless likely to have a trademark understanding of "iPod" and "Mattress Firm," as signposted by their capitalization of the terms.<sup>273</sup> Thus, a corpus-based analysis is more comprehensive when it includes spelling variations, especially in cases involving composite marks.<sup>274</sup>

## 3. Manual Calculation of Capitalization Rates

Sometimes it is necessary to manually classify the term of interest using concordance lines, especially when it is a polysemous mark such as APPLE or DELTA.<sup>275</sup> While it may be possible to

---

<sup>272</sup> Courts have at times recognized the equivalency of various forms of composite marks. *See, e.g.,* *Loctite Corp. v. Nat'l Starch & Chem. Corp.*, 516 F. Supp. 190, 193 n.1 (S.D.N.Y. 1981) ("For clarity, throughout this opinion we have used the term 'Super Glue' as a standard form of reference meant to stand for all of the different forms the term could take, e.g., 'Super glue,' 'Superglue,' 'Superglue,' 'Super Glue,' 'SUPERGLUE,' and the same forms without quotation marks.").

<sup>273</sup> The courts have at times recognized the equivalence of spelling variations in the context of finding genericness, lending credence to this argument that misspellings should at least be included in a corpus-based analysis. *See, e.g., In re Dial-A-Mattress Operating Corp.*, 240 F.3d 1341, 1347 (Fed. Cir. 2001) ("But the difference in spelling between 'M-A-T-R-E-S-S' and 'M-A-T-T-R-E-S' is immaterial, and the Director conceded the legal equivalence of 'mattress' and 'mattress' for the purposes of genericness.").

<sup>274</sup> Oliver Northrup's corpus analysis of MOROCCANOIL is an excellent example of the importance of including spelling variations; his report included "Moroccanoil," "moroccanoil," "MoroccanOil," "moroccanOil," "Moroccan Oil," "moroccan Oil," "Moroccan oil," and "moroccan oil." The capitalization rate of each of these terms was listed separately, even though it was noted that the most common variants were "Moroccan Oil," Moroccan oil," and, the official form, "Moroccanoil." Expert Report of Oliver B. Northrup, 2014 Misc. Filings LEXIS 4411, \*16. The only potentially useful forms that Northrup did not include are those containing hyphens, as in "Moroccan-oil" and "Moroccan-Oil." Otherwise, his analysis was very comprehensive.

<sup>275</sup> Expert linguist Robert Leonard, for example, noted in his report on the disputed mark PODS that the word has so many different meanings that a corpus-based analysis of the term would have been very difficult to administer, in comparison to a mark like APP STORE, for which nearly every result "was an instance of an online application marketplace and [was] therefore relevant to the genericness of that phrase." Declaration of Dr. Robert Leonard, 2014 Misc. Filings LEXIS 4360, \*8.

automatically identify the relevant uses of a mark using collocation data to classify the contexts in which a mark is used,<sup>276</sup> the only way to ensure that only the relevant instances are being counted is by reviewing concordance lines. This risk is most significant for descriptive, suggestive, or arbitrary marks; however, even fanciful marks may have homonyms that are irrelevant to the case at hand; the fanciful mark KODAK, for example, is also the name of an unincorporated coal town in Kentucky.<sup>277</sup> Irrelevant homonyms, such as “Kodak” in Kentucky, should not be counted in corpus-based analysis of capitalization; only uses of KODAK in the context of cameras and technology companies should be considered. To avoid this risk, experts should conduct a survey of potential homonyms of the mark in question using dictionaries, Wikipedia, and web searches to ensure that the analysis excludes terms that are irrelevant. To be certain that such irrelevant uses are excluded, the diligent corpus-based analysis includes a manual classification of uses taken from a random sample of the entire set of uses in a corpus.

#### 4. Interpreting Capitalization Rates

After calculating a capitalization rate for the mark of interest, including each of its applicable inflections and derivations, it is important to ask what such a rate means in the context of not only the English language but also of the specific medium from which the corpus is constructed—that is, what does it mean for a mark to be capitalized a certain percentage of the time given the habitual capitalization rates of words in the English language and in the specific corpus being analyzed?<sup>278</sup> As a baseline of comparison, it is useful to examine analyses of the normal capitalization rates in English that have already been conducted. One of the more robust analyses of this sort is that of Kilgarriff, who generated the distribution of English nouns in relation to their frequency of capitalization using the *UK Web as Corpus* (the “*ukWaC*”),<sup>279</sup> which

---

<sup>276</sup> For example, while it could be reasonably inferred that the majority of instances of “APPLE” on electronics blogs refer to the computer and software brand, an expert can only know for certain by manually reviewing the contexts in which the instances appear.

<sup>277</sup> *Kodak, Kentucky*, Wikipedia (2017), [https://en.wikipedia.org/wiki/Kodak,\\_Kentucky](https://en.wikipedia.org/wiki/Kodak,_Kentucky).

<sup>278</sup> Kilgarriff establishes the important concept of comparing a mark’s capitalization rate to the respective capitalization rates of proper names and common names in a language, but he does not explicitly discuss measuring the capitalization norms for different media, considering that words may be capitalized at different rates on a medium such as Twitter than on a medium such as *Times Magazine*. See Kilgarriff, *supra* note 105.

<sup>279</sup> The *UK Web as Corpus* (the “*ukWaC*”) is an English language database constructed from websites with the “.uk” domain and contains approximately two billion words in total. Adriano Ferraresi et al., *Introducing and evaluating ukWaC, a very large web-derived corpus of English*, Proceedings of the 4th Web as Corpus Workshop, Language Resources and Evaluation (2008).

contained 82,965 unique nouns and 412 million instances of those nouns.<sup>280</sup> His analysis indicates that nouns can generally be categorized in one of two camps: nouns that are capitalized nearly 100% of the times they are used (proper names), and nouns that are capitalized less than 30% of the times they are used (common names).<sup>281</sup> From this, Kilgarriff concludes that “If a word has under forty percent or over eighty percent capitalisation in a corpus such as ukWaC (and has just one sense dominating the distribution, and with some further conditions as explored below), it can generally be identified as, respectively, generic or namelike.”<sup>282</sup>

Kilgarriff’s qualification, “in a corpus such as ukWac,” is very important and deserves further explanation. The capitalization norms for different media, even within American English, may significantly differ from each other. For example, a 70% capitalization rate in news publications should be weighted differently than a 70% capitalization rate on Twitter or on Facebook. As has been noted, professional journalists adhere to editorial policies and write more formally than do consumers who post on many social media websites and blogs. Twitter-users and authors on other types of blogs, on the other hand, are unlikely to capitalize any proper name at as high of a rate as do magazine and newspaper writers.<sup>283</sup> Thus, the ideal corpus-based analysis calculates the capitalization rates of other terms in the specific corpus being used, so that the capitalization rate of the mark in question can be considered in the context from which it is derived. This type of benchmarking is precisely what has been absent from record evidence as it has historically been reported. Several, or even thousands, of instances of a term without such benchmarking are wholly insufficient for a finding either way.

Although the court noted in *Elliot* that it cannot rely on “grammatical formalism” such as capitalization to determine “what

<sup>280</sup> Kilgarriff was able to identify nouns, because the *ukWaC* is tagged for part-of-speech. He also was able to identify that words such as “baby” and “babies” were instances of the same word, because the corpus was *lemmatized*. Kilgarriff, *supra* note 105, at 107-08. Lemmatization is the process by which forms of a word are identified as instances of an underlying basic form. In corpus linguistics, these instances are frequently referred to as “tokens,” and the underlying types are frequently referred to as “types.” See Christian Lehmann, *Lemmatization*, Linguistic Methodology (2013), [http://www.christianlehmann.eu/ling/ling\\_meth/ling\\_description/lexicography/index.html?http://www.christianlehmann.eu/ling/ling\\_meth/ling\\_description/lexicography/lemmatization.html](http://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/index.html?http://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/lemmatization.html).

<sup>281</sup> Kilgarriff, *supra* note 105, at 108.

<sup>282</sup> *Id.*

<sup>283</sup> An 80% capitalization rate in an informal medium such as Twitter, therefore, should be considered more indicative of trademark status than an 80% capitalization rate in a more formal medium such as a collection of magazine articles. Or, contrarily, a 40% rate in a more formal medium should be considered more indicative of genericness than a 40% rate in a less formal medium, since the capitalization rates of proper nouns are likely to be higher in general in a more formal medium.

a speaker has in mind when using a registered trademark,”<sup>284</sup> an analysis showing, for example, a mark to be capitalized 5% of the time compared with a 90% average for proper nouns does much to support an inference that the term is not proper noun by public consensus. The *Elliot* decision does not seem to rule out the efficacy of such evidence; notably, the court mentioned that “[d]ocumented examples of generic use might support a claim of genericide if they reveal a prevailing public consensus regarding the primary significance of a registered trademark.”<sup>285</sup> Such evidence was not probative in *Elliot* largely because the examples proffered by the plaintiff were solely verb uses, which do not necessarily imply genericness.<sup>286</sup> Further frustrating its probative value, the record evidence included neither an indication of the representativeness of the sample from which uses were drawn nor any kind of standard against which to compare such uses.<sup>287</sup>

In addition to comparing the disputed mark’s capitalization rate to the capitalization rates of all words in a corpus, an expert can compare it to the capitalization rates of a subset of trademarks and a subset of generic marks. For example, an analysis of the mark FLORIDA’S NATURAL could compare the capitalization rate of the disputed mark to the capitalization rates of other beverage brand names (such as TROPICANA and SIMPLY ORANGE) and beverage common names (such as “orange juice” and “root beer”). If the disputed mark is capitalized at a rate comparable to generic terms in the corpus, that is indicative of that mark’s being generic; likewise, if the disputed mark is capitalized at a comparable or higher rate than its competitor’s marks, that is indicative of its status as a trademark.

The discussion of various rates of capitalization raises the subsequent and very important question of “At what capitalization rate can it be concluded that a mark is generic?” or, conversely, “At what capitalization rate can it be concluded that a word is a trademark?”<sup>288</sup> As with surveys, there is not an answer to this

---

<sup>284</sup> 860 F.3d 1151, 1162 (9th Cir. 2017).

<sup>285</sup> *Id.* at 1160.

<sup>286</sup> *Id.* at 1161 (“Elliott’s admissible examples are only examples of verb use. To repeat, verb use does not automatically constitute generic use.”).

<sup>287</sup> *Id.* The court noted: “if the parties offer competing examples of both generic and trademark use, this source of evidence is typically insufficient to prove genericide.” *Id.* The issue of interpreting “competing examples” is specifically what corpus-based analyses seek to address in selecting a relevant sample and benchmarking calculations.

<sup>288</sup> For reference, in the context of surveys, in *King-Seeley Thermos Co.*, 75% usage of a term as a generic name has been found to indicate that the disputed mark is generic. See *King-Seeley Thermos Co. v. Aladdin Indus.*, 321 F.2d 577 (2d Cir. 1963). And, in *Ty Inc.* 60% recognition of a term as a trademark has been found to indicate that the disputed mark is not generic. *Ty Inc. v. Softbelly’s, Inc.*, 353 F.3d 528 (7th Cir. 2003). Thus, a capitalization rate lower than 25% may be argued to be indicative of genericness, while a capitalization rate higher than 60% may be argued to be indicative of trademark status.

question that could apply to each and every mark that is analyzed. As discussed above, it is important to consider the disputed mark's capitalization rate in the context of the medium, or media, from which it is derived.

For marks whose capitalization rates are neither as low as most generic terms nor as high as most trademarked terms, and fall generally into the 30% to 70% capitalization range, it becomes important to consider whether alternative generic words exist for the relevant product category. For example, consider the following hypothetical scenario: if a disputed mark is capitalized 50% of the time, but there exists no other generic word for the product, then the current owner of the trademark may have an unfair advantage in the market, because competitors would have no common name with which to label their products.<sup>289</sup> For this reason, this article recommends that in addition to capitalization rates, the frequencies of already existent generic terms (if any)<sup>290</sup> should also be considered in a corpus-based measurement of genericness, especially when the capitalization rate is not determinative.

### 5. A Case Study of Capitalization Rates

When a capitalization rate is sufficiently high or low, however, it may be possible to draw a conclusion about the mark in question without analyzing the frequencies of alternative common names or the existence of certain collocates. In at least one instance, a corpus-based analysis of capitalization rates alone indicates that the Trademark Trial and Appeal Board (TTAB) may have erred in finding a mark not to be generic.

In *In re Trek 2000 International Ltd.*,<sup>291</sup> the TTAB considered evidence regarding the genericness of the mark THUMBDRIVE for various computer-related goods, eventually finding that there was not enough evidence to conclude that the mark was generic.<sup>292</sup> Both

---

<sup>289</sup> Consider the court's opinion, written by Justice Posner, in *Ty Inc.*, considering the point at which a mark should be found generic based on the rates at which consumers recognize the mark as a brand. *Ty Inc. v. Softbelly's Inc.*, 353 F.3d 528, 531-32 (7th Cir. 2003) ("To determine that a trademark is generic and thus pitch it into the public domain is a fateful step. It penalizes the trademark's owner for his success in making the trademark a household name and forces him to scramble to find a new trademark. And it may confuse consumers who continue to associate the trademark with the owner's brand when they encounter what they thought a brand name on another seller's brand. (Think of Ostberg's 60 percent—and there would be a problem even if they were only 10 percent.) The fateful step ordinarily is not taken until the trademark has gone so far toward becoming the exclusive descriptor of the product that sellers of competing brands cannot compete effectively without using the name to designate the product they are selling.").

<sup>290</sup> Such terms would likely include those proposed as the generic name by the putative trademark owner.

<sup>291</sup> 97 U.S.P.Q.2d 1106 (T.T.A.B. 2010).

<sup>292</sup> *Id.* (reversing examining attorney's refusal to register on grounds of genericness).

the applicant, Trek 2000, and the examining attorney proffered pieces of record evidence of “THUMBDRIVE” in the media that supported their respective sides.<sup>293</sup> Since both sides proffered pieces of record evidence that contradicted each other, the Board was not able to determine based on record evidence alone that the mark was generic.<sup>294</sup> Notably, the Board quoted a passage from McCarthy’s treatise: “[f]or example, if a survey showed that 75% of the public regarded the word as generic, then that is its ‘principal significance.’ . . . If, on the other hand, 75 percent of the public regards the term as an indication of a single commercial source for certain goods or services, then that term should be protected as a trademark or service mark.”<sup>295</sup> The Board’s inclusion of this quote indicates its inability to conclude “THUMBDRIVE” was generic, because neither side proffered evidence in the form of usage or in the form of surveys indicating that 75% of the public regarded “THUMBDRIVE” as generic. For this reason, the Board’s decision seemed to be most significantly influenced by the fact that competitors did not use the term “THUMBDRIVE” to describe their products.<sup>296</sup>

Were corpus-based evidence of calculation rates proffered at the time, however, the Board might have been able to find that “THUMBDRIVE” was generic. In the *NOW Corpus* of U.S. newspapers and magazine, all 70 non-sentence-initial instances of “THUMBDRIVE” occurring in 2010 (prior to the decision date of November 30) are not capitalized.<sup>297</sup> Similarly, out of 77 non-

---

<sup>293</sup> Additionally, the applicant submitted search results from Merriam-Webster’s online dictionary indicating that it did not have a listing for “thumbdrive.” *Id.* Since then, however, it appears that Merriam-Webster Online has caught wind of the use of “thumb drive” as a generic term, as there is now a definition for the term, reading “a small usually rectangular device used for storing and transferring computer data: flash drive.” The definition makes no mention of THUMBDRIVE being a trademark. Definition of “thumb drive,” Merriam Webster Online (2017), <https://www.merriam-webster.com/dictionary/thumbdrive>.

<sup>294</sup> *Trek 2000 Int’l Ltd.*, 97 U.S.P.Q.2d at \*8 (“At a minimum, the record creates doubt and we are constrained to resolve that doubt in favor of applicant”).

<sup>295</sup> *Id.* (quoting McCarthy, *supra* note 6, § 12:6).

<sup>296</sup> *Id.* (“In circumstances where a coined term used as a trademark is quickly taken up by the public but not by competitors and the stakes are ‘the fateful step’ of full ‘eradication’ of an applicant’s ‘commercial rights,’ the evidentiary burden is heavy indeed. While evidence of competitor use is not required to satisfy this burden, where the record demonstrates both trademark and generic uses, evidence of the lack of competitor use, at a minimum, may create doubt sufficient to tip the balance in favor of registration.”). The problem with this conclusion is that competitors are likely to avoid using the mark out of fear that, were the mark registered, they would be sued; since they have available to them alternative common names such as “USB flash drive,” using “THUMBDRIVE” was likely not worth the potential legal hassle.

<sup>297</sup> *News on the Web Corpus*, Brigham Young University (U.S. only, 2010), <http://corpus.byu.edu/now/>. This capitalization rate is especially meaningful given that the corpus is composed of newspaper and magazine articles, which, as discussed, are more likely to signpost trademarks because of their editorial policies.

sentence-initial instances of the comparable term with a space (“THUMB DRIVE”)<sup>298</sup> occurring in 2010 before November 30, only two are capitalized.<sup>299</sup> In comparison, out of a sample of 100 non-sentence-initial instances, the mark MICROSOFT is capitalized 100% of the time.<sup>300</sup> Considering that these instances of “THUMBDRIVE” in the corpus originate from widely read sources such as *CNET*, *The Wall Street Journal*, *Fox News Online*, and *Wired*, the capitalization rate of 0% in the *NOW Corpus* should be considered very probative of the public’s understanding of the term, especially in comparison to the limited amount of possibly biased record evidence proffered by the applicant and by the examining attorney. This corpus-based evidence alone, which is derived from the eleven months leading up to the *Trek* decision, strongly motivates a finding of the genericness of “THUMBDRIVE,” which the Board failed to make.<sup>301</sup>

## 6. Limitations of the Capitalization Rate

Certain factors can potentially bias a capitalization rate analysis even when the mark’s rate is compared with the capitalization norms of the medium being analyzed. These include the effects of auto-capitalization, the potential overrepresentation of usage by a single author, the capitalized use of a generic term in a composite trademark, the difficulty of knowing the intended usage of a mark, and the potential gap between capitalization and understanding.

Most word processors and messaging programs feature auto-correction tools that modify words as they are being typed; one frequently made “correction” is the capitalization of words that the algorithm recognizes, or has been programmed to recognize, as proper nouns. Auto-capitalization<sup>302</sup> may lead to situations wherein a mark is typed with a capital letter even though the writer does not recognize the word as a trademark—that is, when typing, the writer does not capitalize the term, but, when the term is auto-capitalized, the writer does not retype the word with a lowercase initial letter. This phenomenon can be partially offset by the contention that writers are generally aware the auto-capitalization is taking place and that this awareness could be construed as recognition of the

---

<sup>298</sup> *Id.*

<sup>299</sup> Of the three instances of “THUMB-DRIVE” (with a hyphen), only one was capitalized, but it was in the title of an article in which all words were initially capitalized: “Hackers, Troops Rejoice: Pentagon Lifts Thumb-Drive Ban,” and another sentence in the source document features the word in lowercase: “Soldiers, you are now cleared to use your thumb drives again.” *Id.*

<sup>300</sup> *Id.*

<sup>301</sup> See *Trek 2000 Int’l Ltd.*, 97 U.S.P.Q.2d at \*10.

<sup>302</sup> As mentioned in *supra* note 7, trademark owners may police their marks through auto-correction algorithms that, in some cases, capitalize their own trademarks but do not capitalize the trademarks of other firms.

mark's status as a trademark. More significantly, a given mark's capitalization rate, if potentially biased by auto-correction, can be put into context by comparing it to the rates of other trademarks, which may generally be subject to the same effect. Regardless, however, the effects of auto-correction complicate the results of a corpus-based analysis.

Another shortcoming associated with calculating capitalization rates is the risk that one individual's language is overrepresented in the corpus being analyzed. For example, there may be 1,000 instances of a mark in the corpus, but if 300 of those instances originated from one person, the capitalization rate would be biased in favor of that person's usage. Even when the dataset is not that biased, it is still likely that some individuals' usage is somewhat overrepresented. Thus, although it can be more laborious for the expert, the ideal corpus should feature an equal number of instances of usage from each source. Fortunately, many corpora, including those of Brigham Young University, are annotated for origin, so that, for example, an expert could choose to count only one instance of the mark from each source or, if the mark is used inconsistently in each source, to either count the most predominant usage from each source or count only the trademark usage, since a single instance of trademark usage by one individual is likely sufficient to maintain that the individual recognizes the trademarked sense of the term. Experts creating their own corpora can choose to do the same when harvesting data, as Oliver Northrup did in his report on "MOROCCANOIL."<sup>303</sup>

An additional factor that can potentially bias capitalization rates is the inclusion of generic terms in composite trademarks. It is possible for generic terms to be included in registrable composite trademarks as long as the generic terms are disclaimed by the trademark owner.<sup>304</sup> For example, the word "mart" in a composite trademark such as STEIN MART for retail store services is disclaimed by the trademark owner, meaning that the term "mart" alone is not registrable and may be used by others; nonetheless, it can still be included (and thus capitalized) in a composite term such as "STEIN MART," which as a whole is registrable. Thus, in an

---

<sup>303</sup> Expert Report of Oliver B. Northrup, 2014 Misc. Filings LEXIS 4411, \*16 ("In order to avoid over-representing any particular author, only a single instance was counted per review. In these cases, I randomly determined which instance to tally."). Of note, where a single author uses a term inconsistently, sometimes clearly using the term as a trademark and other times clearly using the term generically, it is most defensible to count the individual's usage as indicative of trademark recognition, since an author cannot use a term as a trademark without recognizing it as such.

<sup>304</sup> See McCarthy, *supra* note 6, § 19.63 ("The purpose of a disclaimer is to permit the registration of a mark that is registrable as a whole, but contains matter that would not itself be registrable alone. Registration of the whole with a disclaimer prevents any false impression that the registrant owns registered rights to the individual part that is disclaimed.").

analysis of the term “mart” alone, instances of “Stein Mart” would have to be excluded from the calculation of the capitalization rate; this can be done manually or by automatically counting instances where the generic term is preceded or followed by a portion of a composite trademark.

In addition to the potential biases of auto-capitalization, overrepresentation, and generic terms in composite marks, the calculation of the capitalization rate is made significantly more difficult when the mark being analyzed has multiple meanings, as most words do. As discussed above in Part IV(B)(3), it is sometimes necessary to manually review the context in which words appear in a corpus to ensure that all instances are relevant to the analysis; a corpus-based analysis of the trademark APPLE, for example, would likely require a manual review of concordance lines to eliminate all instances of the term that are not used to refer to the APPLE electronics and software company or generically to a category of electronics and software products. While it may be easy to identify and exclude some irrelevant instances of the word, such as instances of “APPLE” that refer to the fruit, it may sometimes be more difficult to identify irrelevant uses. For example, the mark KLEENEX is also a slang term for the drug Ecstasy,<sup>305</sup> so instances of “KLEENEX” in phrases such as “I’m going to go buy a pack of kleenex” or “I’m buying some kleenex” could be slightly ambiguous—is the person buying a pack of facial tissues or doses of Ecstasy? If the instance were capitalized, as discussed above, it could be interpreted that the writers recognize the term as a proper name, even if they are using the term metaphorically as a slang term. However, if the term is in lowercase in such instances, the expert may need to also analyze use of the word “ecstasy” to reach a conclusion about the usage—if the expert can conclude that Ecstasy is not purchased in packs or that “Ecstasy” cannot be pluralized, then it could be concluded that the first instance refers to facial tissues and should be included in the capitalization rate, while the second instance refers to Ecstasy and should be excluded from the capitalization rate. Other instances, however, such as “my favorite is kleenex,” may be completely ambiguous such that they need to be excluded from analysis altogether.

The most significant risk, however, associated with the capitalization rate method is the gap that may exist between usage and understanding. Even if a mark is never capitalized, it is possible, though perhaps very unlikely, that the average consumer still recognizes it as a trademark differentiating one producer from others. This risk is significantly reduced when the mark’s capitalization rate is compared with those of other trademarks in

---

<sup>305</sup> Heymann, *supra* note 6, at 1340 (citing Michael Adams, *Slang: The People’s Poetry* 39 (2009)).

the corpus. If the mark's rate is drastically lower than that of other trademarks in the corpus, it would seem highly improbable that the mark is not generic.

### *C. Common Name Frequencies*

While this article recommends first examining capitalization rates, a comprehensive corpus-based analysis should also survey the existence and frequencies of generic names for the product in question. As discussed above in Part IV(B)(3), it is especially important to consider the availability of generic terms when the capitalization rate of the disputed mark is inconclusive (falling in approximately the 40% to 70% range). If a commonly known<sup>306</sup> alternative generic name for the product besides the disputed mark does not exist, it would put competitors at a competitive disadvantage, requiring them to incur the costs associated with promoting a new generic name for the product.<sup>307</sup> Further, allowing one producer to have exclusive rights to the only commonly recognized generic name for a product can effectively give that producer a monopolistic position in the market.<sup>308</sup>

Thus, the courts have recognized the absence of alternative generic names, in addition to other evidence, as an indicator that the disputed mark is generic.<sup>309</sup> Courts, however, are split as to whether the existence of synonyms suggests that the disputed mark is not generic.<sup>310</sup> This article agrees that the lack of another viable

---

<sup>306</sup> By commonly known, it is meant that the word appears at a frequency that is comparable to that of the disputed mark in question. For example, a generic name such as “transparent cellulose sheets” may at least exist in the lexicon, but if it appears at a frequency significantly lower than does “cellophane,” it may be not be regarded as a viable alternative common name.

<sup>307</sup> See Folson & Teply, *Surveying “Genericness,” supra* note 89, at 24.

<sup>308</sup> McCarthy, *supra* note 6, § 12:2.

<sup>309</sup> See *Ty Inc. v. Softbelly’s Inc.*, 353 F.3d 528, 531-32 (7th Cir. 2003) (discussing the absence of viable alternative common names for a product: “Imagine the pickle that sellers would be in if they were forbidden to use ‘brassiere,’ ‘cellophane,’ ‘escalator,’ ‘thermos,’ ‘yo-yo,’ or ‘dry ice’ to denote products—all being former trademarks that have become generic terms. The problem is not that language is so impoverished that no other words could be used to denote these products, but that if no other words *have* emerged as synonyms it may be difficult for a seller forbidden to use one of the trademarked words or phrases to communicate effectively with consumers.”). Cf. *Door Sys., Inc. v. Pro-Line Door Sys., Inc.*, 83 F.3d 169, 171 (7th Cir. 1996) (in reference to the use of a generic term as a trademark: “Competitors would have difficulty informing consumers that they *were* competitors, because they would be unable, without elaborate and possibly confusing paraphrase, to give the name of the product they were selling.”).

<sup>310</sup> Compare, e.g., *S. S. Kresge Co. v. United Factory Outlet, Inc.*, 598 F.2d 694, 696 (1st Cir. 1979) (“The question, however, is not whether a term is more frequently chosen colloquially than any of its synonyms, but whether it still retains its generic meaning.”), with, e.g., *Elliot v. Google Inc.*, 45 F. Supp. 3d 1156, 1172 (D. Ariz. 2014) (“If competitors can accurately describe their products or services without using the mark in question, it suggests the mark is not generic”) (citing, e.g., *Salton Inc. v. Cornwall Corp.*, 477 F. Supp. 975, 986 (D. N.J. 1979)), *aff’d*, 860 F.3d 1151 (9th Cir. 2017)).

common name, in conjunction with other evidence such as low capitalization rates, can be indicative of the disputed mark's genericness, but contends that the existence of synonyms does not necessarily suggest that the mark is not generic.

For example, based on a search of BYU's *NOW Corpus* of U.S. sites, there exist few commonly used alternative common names for the (now) generic term, "app store." "App store" appears over seven thousand times in the corpus, while the total frequency of its alternatives is fewer than 150: "software store" (13 instances), "application store" (39 instances), "digital store" (36 instances), "app marketplace" (48), and "application marketplace" (10 instances).<sup>311</sup> Given the lack of commonly used alternative generic terms and the fact that "app store" is capitalized at a lower rate than other marks,<sup>312</sup> there is reason to consider that "app store" may be generic.<sup>313</sup>

In contrast, in *S. S. Kresge Co. v. United Factory Outlet, Inc.*, multiple high-frequency alternative generic names existed for the disputed mark "Mart," but it was still found to be generic.<sup>314</sup> In fact, the synonyms of "mart" were used far more commonly than "mart" was; according to Google's Ngram corpus of books written in American English, between 1960 (the year in which the senior user of the mark entered the market) and 1979 (the year of the dispute), "mart" only appeared 44,937 times, while "store" appeared 1,139,499 times and "market" appeared 2,874,103 times.<sup>315</sup> Thus, competitors and consumers clearly had viable alternative generic marks with which to describe the relevant product category, but "Mart" was nevertheless found to be generic. This finding, unsurprisingly, is supported by the fact that far fewer than 70% of the instances of "mart" between 1960 and 1979 in the NGram corpus are capitalized.<sup>316</sup> Considering that this rate includes a significant

---

<sup>311</sup> *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (last visited Aug. 18, 2018). Of note, "online marketplace" appears 582 times in the U.S.-only *NOW Corpus*, but this is not a direct synonym for "online application marketplace" or "app store"; "online marketplace" could include sites such as Amazon, which offer a different set of products and services than do "online application marketplaces." *Id.*

<sup>312</sup> Based on a search of U.S. sites on the *GloWbe Corpus* and 2012 data on the *Corpus of Contemporary American English*, "APP STORE" (and its variant spellings) is capitalized 71% of the times that it is used (excluding sentence initial instances). *Corpus of Global Web-Based English* (U.S. sites only, 2012-2013) and *Corpus of Contemporary American English* (2012) [RAW DATA], Brigham Young University. In comparison, prominent trademarks such as MICROSOFT, IPHONE, and PLAYSTATION are all capitalized above 90% of the time in the same set of data. *Id.*

<sup>313</sup> See *supra* note 182, discussing abandonment of the APP STORE mark.

<sup>314</sup> 598 F.2d 694 (1st Cir. 1979).

<sup>315</sup> *Google Books (American) Corpus*, Brigham Young University (1960–2000), <http://googlebooks.byu.edu/x.asp>.

<sup>316</sup> *Id.*

number of capitalized instances of “mart” in WAL-MART and STEIN MART, which are composite trademarks that should not be included in such a calculation,<sup>317</sup> this capitalization rate is indicative of the genericness of “mart.”

Further demonstrating the importance of relying upon capitalization rates in addition to alternative generic frequencies is the possibility of a product having no commonly used generic name. For example, in the market for game systems such as the PlayStation, the generic terms “game/gaming console” and “game/gaming system” appear fewer than 600 times in the *GloWbe Corpus* (U.S. sites), while the trademarks PLAYSTATION, XBOX, and NINTENDO appear over 12,500 times cumulatively.<sup>318</sup>

These examples show that the frequency of alternative generic names should be considered only in conjunction with other linguistic indicators, namely, the rate at which the disputed mark is capitalized. Further, it is important to emphasize that, as mentioned, calculating the frequency of alternative generic names is most important for the marks whose capitalization rates are not completely determinative. If a mark is capitalized 95% of the times it is used, that is a strong indicator it is a trademark; likewise, if it is capitalized only 5% of the times it is used, that is a strong indicator it is generic. If it is capitalized 50% of the times it is used, though, its capitalization rate alone may not be determinative—in a case such as this, a corpus-based analysis indicating that there does not exist a frequently used alternative common name for the product in question would serve to motivate a finding that the disputed mark is generic.

#### *D. Collocation Analysis*

Although it is difficult, given the present capabilities of computational linguistics, to determine with certainty the intended meaning of every instance of a word in a corpus, there are certain usages that can be automatically counted and are very probative of whether or not a term is being used as a trademark. For example, if “KLEENEX” always appeared as “Kleenex brand tissues,” it could be confidently concluded from such usage that KLEENEX is a valid trademark. Likewise, if every instance of “XEROX” were preceded by “Canon” or “Toshiba” (as in “I have a Canon xerox machine, but she has a Toshiba xerox”), then it could be deduced that “XEROX” is generic.

---

<sup>317</sup> The term “mart” in a registered composite trademark such as STEIN MART is disclaimed by the trademark owner and should not be included in a calculation of capitalization rates. See *supra* Part IV(B)(6), discussing how to treat generic terms in composite trademarks.

<sup>318</sup> *Corpus of Global Web-Based English*, Brigham Young University (U.S. only, 2012–2013), <http://corpus.byu.edu/glowbe/>.

In reality, it is, of course, extremely unlikely that words would co-occur with each other 100% of the time. Nonetheless, words can appear next to each other at high enough frequencies that it is possible to deduce a strong and statistically significant association between the two words. For example, although the word “pie” is not preceded by “apple” 100% of the time, it may be preceded by “apple” frequently enough to conclude that the words have a statistically significant association with each other. Such words that habitually co-occur with each other at a frequency greater than chance, such as “apple” and “pie,” are called collocates.<sup>319</sup> If the collocates preceding a disputed mark include other trademarked brand names, that is probative of the mark’s genericness; and if the collocates following the disputed mark include words such as “brand” or are generic terms for the relevant product, that is probative of the mark’s status as a trademark.

### 1. Measurements of Collocation

The degree to which a word is a collocate of mark can be measured using a few different types of calculations,<sup>320</sup> two of which are discussed in this article: the mutual information (MI) score and the t-score. MI scores measure the strength of association between two words,<sup>321</sup> and t-scores measure the confidence with which one can assert that an association between two words is statistically significant (or, non-random).<sup>322</sup> The MI score is useful for automatically searching a corpus for informative collocates, because it gives little value to high-frequency words, such as “the,” that are not useful for analysis;<sup>323</sup> but MI scores must be checked for statistical significance using a t-score before a conclusion can be made about an association between two words. If experts do not wish to survey a corpus for collocates and, rather, have a specific set

---

<sup>319</sup> See *supra* note 132.

<sup>320</sup> It is noted that there is a degree of controversy about the best measurements of collocation and there is further methodological development to be made, especially in their application to trademark law. Nonetheless, this article presents two of the measurements and hopes to inspire future research and consideration of the application of collocation to the analysis of disputed marks. See generally, e.g., Stefan Evert, *Corpora and Collocations* (extended manuscript), 2 *Corpus Linguistics: An International Handbook*, 1212–1248 (2008).

<sup>321</sup> Kenneth Ward Church and Patrick Hanks, *Word Association Norms, Mutual Information, and Lexicography*, 16 *Computational Linguistics*, 22, 23 (1990).

<sup>322</sup> Christopher Manning & Hinrich Schütze, *Foundations of Statistical Natural Language Processing* 181 (2d ed. 1999). See also *A Guide to Statistics: T-Score and Mutual Information*, HarperCollins Publishers Ltd (2008), [https://wordbanks.harpercollins.co.uk/other\\_doc/statistics.html](https://wordbanks.harpercollins.co.uk/other_doc/statistics.html); *Statistics Used in the Sketch Engine*, Lexical Computing Ltd (2014).

<sup>323</sup> For example, BYU’s corpus interface allows one to search for collocates and to rank them by MI score, which the website automatically calculates. *Mutual Information Score*, Corpus.Byu.Edu, Brigham Young University (2017).

of words between which they would like to examine the significance of association, calculating t-scores is most useful.

MI scores are calculated using the following formula:<sup>324</sup>  $MI = \log_2((AB * N) / (A * B * \Delta))$  and t-scores are calculated using the following formula:<sup>325</sup>  $T = (AB - ((A * B * \Delta) / N)) / (AB)^{1/2}$  where: A = the total frequency of the mark being analyzed (the node word), B = the total frequency of the collocate, AB = the frequency of the collocate near the node word, N = the number of words in the corpus, and  $\Delta$  = span, the range of words to the left and/or to the right of the node word whereby the existence of a term in that range should constitute that term's being a collocate.<sup>326</sup> A highly positive mutual information score indicates that the terms tend to frequently occur together, with an MI score of 3 or higher indicating that the words at least moderately act as collocates.<sup>327</sup> The higher a t-score is, on the other hand, the more statistically significant the co-occurrence is. Assuming a normal distribution of word probabilities in a corpus,<sup>328</sup> a t-score higher than 1.96 indicates there is more than 95% confidence that a co-occurrence is not random.<sup>329</sup> A t-score higher than 2.326 indicates there is more than 99% confidence that the co-occurrence is not random.<sup>330</sup>

Since the mutual information formula divides the frequency of co-occurrence by the frequency of the collocate, extremely common words such as “the” and “a” receive a lower MI score, and less common words such as technical terms receive a higher MI score.<sup>331</sup>

<sup>324</sup> See Church & Hanks, *supra* note 321 at 23.

<sup>325</sup> See Manning & Schütze, *supra* note 322 at 165.

<sup>326</sup> Of note, it is important to ensure that the span does not include collocates separated by a comma, so that instances such as “Puffs, kleenex, and Angelsoft” are not counted as instances of “Puffs kleenex,” the latter of which would appear indicative of the genericness of the word “kleenex.”

<sup>327</sup> Richard Xiao, *Making Statistic Claims*, Lancaster University: Corpus-Based Language Studies (2015). As a point of reference, in BYU's *NOW Corpus*, the co-occurrence of “APPLE” and “PIE” (in that order) has an MI score of 9.04. *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (last visited Aug. 18, 2018).

<sup>328</sup> It is accepted that word probabilities (discussed as “frequencies” in this article) are not normally distributed, and other collocation measures, such as the chi-square test, which does not assume a normal distribution, have been suggested for measuring collocation. See Manning & Schütze, *supra* note 322 at 169. However, Manning & Schütze note that t-statistics and chi-square statistics generally do not meaningfully differ for identifying collocations, finding that for the 1990 *New York Times* corpus, the top twenty collocates with the highest t-scores are the same as the top twenty collocates with the highest chi-square scores. *Id.* at 170.

<sup>329</sup> See, e.g., David M. Lane, *Introduction to Statistics: An Interactive e-Book* 320-22 (2013); 1 E.S. Pearson & H. O. Hartley, *Biometrika Tables for Statisticians* 146 (3d ed. 1966).

<sup>330</sup> *Id.* at 146.

<sup>331</sup> Antonia Sánchez Martínez, *Collocation Analysis of a Sample Corpus Using Some Statistical Measures: An Empirical Approach*, University of Murcia: Official School of Languages 763, 767 (2008). See also Manning & Schütze, *supra* note 322, at 181.

For example, in the *NOW Corpus*, “his” and “book” have an MI score of 3.40, while “comic” and “book” have an MI score of 9.93.<sup>332</sup> This calculation is beneficial when surveying a corpus for high-frequency collocates of a disputed mark, as an expert may not want search results to include irrelevant collocates such as certain articles, prepositions, or other common terms that provide little information about the mark in question. Rather, an expert would be interested in more unique, generally lower frequency collocates, such as competitors’ trademarks (preceding the mark) or generic names for product categories (following the mark), which are generally given more weight by the MI calculation.

The downside to the MI score is that it inflates the significance of association of all low-frequency words, even when their co-occurrence with the disputed mark may be completely random.<sup>333</sup> Words appearing only once or twice in the corpus, for example, have extremely high MI scores for the words with which they do happen to co-occur—this is unfavorable for the purposes of analysis, because it results in high MI scores that are not backed by a substantial suite of evidence.<sup>334</sup> In BYU’s *NOW Corpus*,<sup>335</sup> for example, the brand names “SEA-DOO,”<sup>336</sup> “YAMAHA,”<sup>337</sup> “GP800,”<sup>338</sup> and “SEASHARK”<sup>339</sup> are each listed among the top collocates for “JET

---

<sup>332</sup> *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (last visited Aug. 18, 2018).

<sup>333</sup> Martínez, *supra* note 331, at 766.

<sup>334</sup> Manning & Schütze, *supra* note 322, at 182.

<sup>335</sup> The MI score and t-score are functions of the mark’s frequency, the collocate’s frequency, the frequency of their co-occurrence, and the total number of words in the corpus, as discussed above. Each MI score and t-score listed subsequently is calculated using the formulas listed above and the data provided by the *NOW Corpus*.

<sup>336</sup> SEA-DOO is another brand name for personal watercraft. “SEA-DOO” and “JET SKI” had a mutual information score of 15.62, with “SEA-DOO” (with and without a hyphen) preceding “JET SKI” a total of six times and “SEA-DOO” appearing in the corpus a total of 238 times. *News on the Web Corpus*, Brigham Young University, <http://corpus.byu.edu/now/> (last visited Aug. 18, 2018).

<sup>337</sup> Yamaha is also a producer of personal watercrafts, marketing under WAVERUNNER. “YAMAHA” and “JET SKI” had a mutual information score of 8.73, with “YAMAHA” preceding “JET-SKI” three times and “YAMAHA” appearing in the corpus a total of 14,079 times. *Id.*

<sup>338</sup> GP800 is a specific line of personal watercraft produced by Yamaha. “GP800” and “JET SKI” had a mutual information score of 17.84, with “GP800” preceding “JET SKI” twice and “GP800” appearing 17 times in the corpus. *Id.*

<sup>339</sup> “SEASHARK” is the fictional brand name for personal watercraft used in the *Grand Theft Auto* video game series. “SEASHARK” and “JET SKI” had a mutual information score of 19.93, with “SEASHARK” preceding “JET SKI” once and “SEASHARK” appearing in the corpus only twice. *Id.*

SKI,”<sup>340</sup> with MI scores higher than 7.<sup>341</sup> However, some of these words appear in the corpus so few times that it is unclear whether their co-occurrence with “JET SKI” is non-random. For example, the word “SEASHARK” has a very high MI score of 19.93 with “JET SKI,” because, although it appears in the corpus only twice, it co-occurs with “JET SKI” once (50% of the times it is used).

Considering the MI’s score’s tendency to inflate the significance of an association with a low-frequency word, it is important to use a t-score to ensure that collocates with high MI scores are actually statistically significant.<sup>342</sup> For example, despite “SEASHARK”’s high MI score for preceding “JET SKI,” its t-score is only 1.00, indicating there is less than 75% confidence that the co-occurrence was non-random, which is inadequate for drawing a strong conclusion about their relationship. Contrarily, “YAMAHA” and “JET SKI” have a t-score of 2.45, indicating that there is over 99% confidence that their association is not random. These examples exhibit the importance of checking the statistical significance of collocates with high MI scores by calculating t-scores. Only a t-score can be used to support a conclusion that two words have a non-random association with each other. Whereas, the MI score is most useful for conducting an initial screen of collocates for the mark being analyzed.<sup>343</sup>

## 2. Case Studies of Collocation

Even when a word’s co-occurrence with the mark is deemed statistically significant, sometimes co-occurrence manifests itself at such a low frequency in the first place, as in the case of “JET SKI,” that collocation analysis may lack probative value in the eyes of the court. Other examples, however, where the collocate and the mark co-occur at higher raw frequencies, may be more convincing of a mark’s status, especially when capitalization rates are not entirely conclusive.

---

<sup>340</sup> “JET SKI,” (with spaces, without spaces, and with a hyphen), appearing 3,031 times in the corpus (*id.*), is a registered trademark of Kawasaki. *Trademark Electronic Search System*, United States Patent and Trademark Office (Registration date: May 27, 1986).

<sup>341</sup> For each of these calculations, a span of one word to the left of “JET SKI” was used to capture immediately preceding collocates, and spelling variations such as “JET-SKI” and “SEADOO” were included. Of note, for the purposes of making a distinction between the MI score and the t-score, these calculations are based on frequencies derived from the entire *NOW Corpus* (6,042,387,415 words), including non-U.S. sites. A proper analysis of a U.S. trademark should feature language derived from only U.S. websites.

<sup>342</sup> The t-score, unlike the MI score, does not scale the result by the frequency of the collocate; it only measures the chance that the collocation occurred non-randomly. Thus, a word such as “the” would have a high t-score for most nouns in the U.S. language. See Martínez, *supra* note 331, at 766.

<sup>343</sup> On top of the calculation of t-scores and MI scores, further evaluation of the specific contexts of such collocates in the sample would bolster the analysis and would surely not go unchecked by the opposing party at trial.

For example, as discussed above in Part IV(C)(1), “APP STORE” is capitalized 71% of the times it is used,<sup>344</sup> which is rather inconclusive; even though its capitalization rate is lower than that of other trademarks in the corpus,<sup>345</sup> it still is apparent that a majority of writers may recognize the word as a proper noun or as a trademark. However, this capitalization rate may include instances such as “Amazon App Store,” where “APP STORE” is being used generically although it is capitalized—even if the composite terms “Apple App Store” and “Amazon App Store” were both trademarks, if “APP STORE” by itself were deemed generic, it would need to be disclaimed by Apple and Amazon so that they and other competitors could use the term.<sup>346</sup> Thus, evidence indicating that consumers use “APP STORE” following the trademarks of other brands is helpful for disambiguating the capitalization rate:

In BYU’s *NOW Corpus* (U.S. only) the other brand names “TVOS,”<sup>347</sup> “ANDROID,”<sup>348</sup> and “AMAZON”<sup>349</sup> precede “APP STORE” 207 times cumulatively, and they have a combined MI score of 8.24 and t-score of 14.34. That other brand names precede “APP STORE” at a raw frequency greater than 200 and at a level of significance higher than .01 indicates that there is a meaningful linguistic association between the mark and competitors’ brand names. In comparison, the cumulative MI and t-scores for Apple-brand names, including “APPLE,” “iPHONE,” “iPAD,” “iTUNES,” “iOS,” “AAPL,” and “MAC,” preceding “APP STORE” are 9.53 and 35.84, respectively.<sup>350</sup> Thus, although “APP STORE” retains a strong association with Apple, it also has a meaningful and non-random association with other brands. Further, that there are no frequently used alternative common names for “APP STORE”<sup>351</sup> and that consumers feel they need to differentiate between different “app stores” by preceding the term with the names of specific producers strongly indicate that “app store” is the generic term of choice. Thus, in conjunction with the fact that “APP STORE”’s

<sup>344</sup> Based on the *GloWbe Corpus* U.S. sites only and COCA. See *supra* note 312.

<sup>345</sup> *Id.*

<sup>346</sup> See *supra* note 304, discussing disclaimers of generic terms in composite trademarks.

<sup>347</sup> “APP STORE” (with a space, without a space, and with a hyphen) appears 7,370 times in the corpus, “TVOS” precedes it 13 times, “TVOS” appears 278 times in total in the corpus, and the *NOW Corpus* of U.S. sites consists of 1,122,633,101 total words. *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (last visited Aug. 18, 2018).

<sup>348</sup> “ANDROID” appears 43,509 times in the corpus and precedes “APP STORE” 46 times.

<sup>349</sup> “AMAZON” appears 60,145 times in the corpus and precedes “APP STORE” 148 times. *Id.*

<sup>350</sup> *News on the Web Corpus*, Brigham Young University, <http://corpus.byu.edu/now/> (last visited Aug. 18, 2018).

<sup>351</sup> See *supra* Part IV(C), discussing the low-frequency of alternative common names for “app stores.”

capitalization rate is lower than that of other trademarks<sup>352</sup> and that there is a lack of alternative common names for the product category,<sup>353</sup> the collocation-based evidence strongly implies that “APP STORE” is generic and that Apple was smart to abandon its rights to the mark in 2013 as opposed to engage in further litigation over it.<sup>354</sup>

Likewise, the word “PHOTOSHOP” has a capitalization rate that is high but could be further substantiated by analyzing collocates. Out of a random sample of 100 non-sentence-initial instances of “PHOTOSHOP” on the *NOW Corpus* consisting of U.S. newspaper and magazine articles, 91 instances were capitalized;<sup>355</sup> and on the *GloWbe Corpus* of U.S.-based blogs, 76 were capitalized.<sup>356</sup> Considering that blog posts written by individuals are not subject to the editorial standards regarding trademarks that newspapers and magazines are, it is logical that the capitalization rate is lower. However, as the capitalization rates are lower than those of certain other trademarks,<sup>357</sup> it may be helpful to conduct an analysis of collocates of “PHOTOSHOP” to examine the extent to which it collocates with trademark signifying words such as “ADOBE” or “CS6” compared with other brand names such as “MICROSOFT” or “GOOGLE,” which would indicate genericness.

A search of the *NOW Corpus* (U.S. sites only) indicates that there is in fact a significant relationship between “PHOTOSHOP” and brand-indicating terms. Specifically, “ADOBE PHOTOSHOP” appears 153 times, implying an MI and t-score of 13.46 and 12.37, respectively, between the terms. Aggregately, “PHOTOSHOP” has an MI and t-score of 13.84 and 14.83, respectively, with brand-indicating collocates,<sup>358</sup> and there is not a single instance of “PHOTOSHOP” being preceded by the name of another non-Adobe brand.<sup>359</sup> That “PHOTOSHOP” is never preceded in the corpus by another trademark indicates that consumers do not need to use the term generically in the context of differentiating between different

<sup>352</sup> See *supra* note 312, discussing the 71% capitalization rate of “APP STORE.”

<sup>353</sup> See *supra* Part IV(C).

<sup>354</sup> See McAllister, *supra* note 182.

<sup>355</sup> *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (2,886 results, last visited Aug 18, 2018).

<sup>356</sup> *Corpus of Global Web-Based English*, Brigham Young University (U.S. blogs only, 2012–2013), <http://corpus.byu.edu/glowbe/> (3,134 results).

<sup>357</sup> *E.g.*, Out of a random sample of 100 instances of “POWERPOINT” in the *GloWbe Corpus*, 93 were capitalized. *Id.* (1,565 results). Out of a sample of 100 instances in the *NOW Corpus*, 95 were capitalized. *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (1,975 results, last visited Aug 18, 2018).

<sup>358</sup> “ADOBE,” “LIGHTROOM,” “CS2,” “CS3,” “CS4,” “CS5,” and “CS6.” *News on the Web Corpus*, Brigham Young University (U.S. only), <http://corpus.byu.edu/now/> (last visited Aug 18, 2018).

<sup>359</sup> *Id.*

brands of photo-editing software. This, in addition to “PHOTOSHOP”’s significant co-occurrence with brand signifiers and its capitalization rate, strongly implies that the mark is valid.

### 3. Limitations of the Collocation Method

There are a number of shortcomings to using collocations as evidence in genericness cases, the most significant of which is that, although high MI and t-scores may indicate a strong association between a disputed mark and a certain type of usage, they may not be interpreted by the court as capturing the “primary significance of the term in the minds of the consuming public.”<sup>360</sup> As previously mentioned, it is essentially impossible for a word to occur in the same context the majority of the times it is used, let alone 100% of the time. Additionally, as the examples of “APP STORE” and “JET SKI” indicate, there may often be fewer than 100 relevant instances of co-occurrence in a given corpus.

Thus, although a mark’s being preceded by a competitor’s brand name could be said to be statistically significant such that the words have a meaningful association in that context, the court may only recognize such evidence as indicating that 100 consumers (out of millions) think the mark is generic. Further, collocation analysis is significantly more difficult to explain than a simple capitalization rate, potentially diminishing its usefulness in litigation. Evidence that 95% of instances of “KLEENEX” are capitalized, for example, is much easier to explain than evidence that thirty instances of “Puffs kleenex” indicate a statistically significant relationship between “Puffs” and “kleenex” in that context. For these reasons, the collocation method, like the measurement of common name frequencies, is best employed in conjunction with the capitalization rate calculation.

Nevertheless, even if collocation analysis cannot be used alone to ascertain the status of a mark, the statistical methods of the MI and t-score calculations can improve the empiricism with which instances of record evidence are presented. If, for example, counsel wished to present record evidence such as instances of “I purchased Puffs brand kleenexes,” MI and t-scores showing the significance of the co-occurrence of “Puffs” and “kleenexes” could be used to indicate that such instances are meaningful and non-random.

## V. CONCLUSION

In deciding whether to administer a survey or to employ a corpus-based analysis, it is crucial to consider the nature of the mark being analyzed, the type of evidence proffered by opposing parties, and cost-effectiveness. Generally, corpora may be less

---

<sup>360</sup> 15 U.S.C. § 1064.

suited to evaluating marks for products not consumed by the general public or products for which there is insufficient textual data originating from potential purchasers of the product (such as escalators). Contrarily, for a mark commonly used both as a trademark and as a generic term for that respective trademark's product category, a survey asking respondents to make a binary classification of the term may be ineffectual, and a corpus, which can better capture the variance of actual usage, may be preferable. More generally, the marks best suited for corpus-based analyses are those for products habitually consumed by the majority of the speaking public. For these marks, already existent, general corpora can be used for analysis such that it may not be necessary to create a specific dataset, thus saving time and money. In other circumstances, in which either a survey or a corpus analysis could suffice, the corpus-based approach may be preferable because of its ability to capture language in a more natural context, its ease of replicability, and its potential cost-effectiveness.

Regardless, and at the very least, corpora can drastically improve the empiricism with which record evidence is presented. Whenever possible, instances of usage proffered to the court should be substantiated by an indication of the representativeness of the sample from which they were drawn, the methodology by which they were chosen and classified, and how they compare to instances of other trademarks and generic terms from the same sample. Corpora and corpus linguistics can bring such analytical rigor to record evidence and meaningfully enhance its probative value. Some trademark litigants and their experts, discussed in this article, have taken notice of the value in such analyses; however, there is still progress to be made by way of methodological development and judicial recognition of such evidence. Specifically, experts and the courts should reevaluate the probative value of specific forms of record evidence, namely capitalization, in wake of the herein discussed ability to substantiate such metrics with corpus data. In terms of methodological development, future studies directly comparing the application of surveys and corpora could provide more perspective on corpus analyses.

With statistical perspective and the methods of corpus linguistics, this article contends the probative value accorded to record evidence can grow to rival that accorded to surveys. In addition to providing litigants a more accessible way to proffer evidence of genericness or trademark validity, the permanence of the textual data in corpora affords a more transparent judicial process—opposing parties can easily audit and replicate corpus-based analyses, allowing for more productive cross-examinations and responding expert reports. Moreover, trademark owners wishing to discreetly and inexpensively monitor their marks can employ corpus-based analyses using just a computer.

Ultimately, in weighing survey versus corpus-based evidence, jurists will need to grapple with the question of whether elicited “understanding” or usage is the better measure of primary significance. This article, inspired by the practice of lexicographers, who are principally concerned with the meaning of words, argues that the informed analysis of usage has distinct advantages over the elicitation of understanding through surveys—such usage is unaffected by the artificial effects of the survey environment, including question wording and the surveyor’s presence itself, and is not distorted by a binary obligation to label a mark one way or another. As litigants become more comfortable proffering corpora, courts will be forced to opine on this issue. In the meantime, if record evidence is to be acceptable at all, then the corpus-based analysis should be the method by which it is presented to the court.

---