


9-22-2021

## Natural Language and Legal Interpretation

Stephen C. Mouritsen

Follow this and additional works at: <https://brooklynworks.brooklaw.edu/blr>

 Part of the [Courts Commons](#), [Judges Commons](#), [Jurisprudence Commons](#), [Law and Society Commons](#), and the [Science and Technology Law Commons](#)

---

### Recommended Citation

Stephen C. Mouritsen, *Natural Language and Legal Interpretation*, 86 Brook. L. Rev. 533 (2021).  
Available at: <https://brooklynworks.brooklaw.edu/blr/vol86/iss2/8>

This Article is brought to you for free and open access by the Law Journals at BrooklynWorks. It has been accepted for inclusion in Brooklyn Law Review by an authorized editor of BrooklynWorks.

# Natural Language and Legal Interpretation

Stephen C. Mouritsen<sup>†</sup>

## INTRODUCTION

Judges and lawyers often appeal to the “ordinary meaning” of the words in legal texts.<sup>1</sup> Until very recently, claims about the ordinary meaning of words in legal texts have not been informed by evidence of the way that words are used or understood by ordinary people.<sup>2</sup> This is because no such evidence—and no method to gather such evidence—was available. Instead, judges, parties, and scholars have been left to rely on their own linguistic intuitions and dictionaries, both of which are problematic guides to the usage or understanding of ordinary people.

This symposium on Data Driven Interpretation focuses on recent developments that have challenged these traditional limitations: The rise of the use of linguistic corpora in legal interpretation, and recent scholarship evaluating interpretive claims with survey methods. Corpus linguistics is the study of language using large, coded, electronic collections of natural language—language used in natural settings. Survey methods involve the systematic collection of linguistic judgments from survey respondents, particularly through modern survey administration platforms like Amazon’s Mechanical Turk and Lucid.<sup>3</sup>

---

<sup>†</sup> Stephen C. Mouritsen is a shareholder at Parr Brown Gee & Loveless and Adjunct Professor at the J. Reuben Clark Law School at Brigham Young University.

<sup>1</sup> See *The Sussex Peerage*, XI Clark & Finnelly 85, 143 (H.L.) (appeal taken from Eng.) (1844) (Tindal, L.C.J.) (applying the plain meaning rule); *Hill v. E, and W. India Dock Co.*, 9 App. Cas. 448, 465 (H.L.) (1884) (Bramwell, L.J.) (same); Geoffrey P. Miller, *Pragmatics and the Maxims of Interpretation*, 1990 WIS. L. REV. 1179, 1184 (discussing prior analogs to the plain meaning rule).

<sup>2</sup> Craig Hoffman, *Parse the Sentence First: Curbing the Urge to Resort to The Dictionary When Interpreting Legal Texts*, 6 N.Y.U. J. LEGIS. & PUB. POL’Y 401, 401 (2003) (“[J]ust as medical science has progressed since the time of leech treatments, the science of linguistics has progressed . . .”).

<sup>3</sup> Suneal Bedi & David Reibstein, *Measuring Trademark Dilution by Tarnishment*, 95 IND. L.J. 683, 707 n.126 (2020) (“Amazon Mechanical Turk is an internet marketplace that allows businesses (and individuals) to coordinate with humans to perform tasks that computers are currently unable to do. It is a simple and cost-effective way for social scientists to get consumer data on various topics. Marketing, psychology, management, and

For decades, linguists have relied on both corpus methods and survey methods to gather evidence of language usage and understanding. Only recently have these methods been brought to bear on questions of legal interpretation. Using linguistic corpora, interpreters can gather evidence of the way words are used in a given context, in a given speech community, and at a particular point in history. Using survey methods, interpreters can gather survey respondents' judgments about the acceptability of competing interpretations.

The question posed by some participants in this symposium is that if we are trying to understand how the words of the statute are perceived by ordinary people, why not just ask? Certainly, survey scenarios can be constructed to mirror both the background circumstances and the language of the legal text in question. But it is not clear that survey methods can provide us with a clearer picture of ordinary meaning of the words in a statute. Survey responses are gathered in an artificial linguistic setting and are the product of an artificial, metalinguistic task. For these reasons, they may not reflect a respondent's actual language usage or understanding.

Corpus evidence and survey evidence may also disagree, providing inconsistent answers to what appear on the surface to be similar questions. This has led some to conclude that the corpus-based answers to these interpretive questions must be incorrect. But where survey responses and corpus-based evidence differ on similar questions, we may have reason to prefer the corpus-based answer because it is premised on evidence of natural language that was produced in a natural communicative setting. At the very least, where survey-based answers and corpus-based answers to interpretive questions differ, interpreters should carefully interrogate the reason for that difference.

At bottom, both corpora and surveys can provide evidence of linguistic use and understanding that was not previously available to interpreters. This essay will explore some of the comparative advantages and disadvantages of using corpus methods or survey methods in legal interpretation.

Part I will outline some the contours of corpus linguistics, highlighting some of the contributions that corpora can bring to legal interpretation and some of its limitations. Similarly, Part II will outline some of the contours of a survey approach, highlighting some contributions and limitations of the approach.

---

legal scholars have routinely used Amazon Mechanical Turk to recruit respondents for surveys and other empirical studies.”). Lucid, similarly, is a survey respondent aggregator. See LUCID, <https://luc.id/> [<https://perma.cc/RF2N-U9DD>].

Part III will address some circumstances where corpus evidence and survey evidence disagree. Part IV will conclude this paper addressing some avenues for future research.

## I. ORDINARY MEANING AND THE OBJECT OF INTERPRETATION

For all of its ubiquity, there is no generally agreed upon meaning of the phrase “ordinary meaning.”<sup>4</sup> As Richard Posner has observed, “[i]t is ironic that a principle designed to clarify should be so ambiguous.”<sup>5</sup> Courts may refer to the “most common” or “common” sense of a word,<sup>6</sup> or a possible or natural meaning.<sup>7</sup> It is not at all clear that these notions of “ordinariness” carry the same meaning. Nor is it the case that individual judges necessarily use the phrase consistently. When individual judges use the phrase “ordinary meaning,” they may have in mind entirely different concepts of usage or meaning.<sup>8</sup> Indeed, Justice Scalia variously characterized “ordinary meaning” as (1) “what an ordinary speaker of the English language would think it means”, (2) what an “ordinary Member of Congress” would read a legal text, or (3) as an “objectified’ intent—the intent that a reasonable person would gather from the text of the law, placed alongside the remainder of the *corpus juris*,” without explaining how it is that these expressions of ordinary meaning differ.<sup>9</sup>

Justice Scalia was not the only judge to articulate differing conceptions of “ordinary meaning.” In his 1899 essay, *The Theory of Legal Interpretation*, Justice Oliver Wendell Holmes, Jr., famously characterized the task of interpretation as follows:

Thereupon we ask, not what this man meant, but what those words would mean in the mouth of a normal speaker of English, using them in the

---

<sup>4</sup> Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning* 127 YALE L.J. 788, 798 (2018) (“[I]ronically, we have no ordinary meaning of ‘ordinary meaning.’”).

<sup>5</sup> Richard A. Posner, *Statutory Interpretation—In the Classroom and in the Courtroom*, 50 U. CHI. L. REV. 800, 808 (1983).

<sup>6</sup> *Taniguchi v. Kan Pac. Saipan, Ltd.*, 566 U.S. 560, 568–69 (2012).

<sup>7</sup> See *Mont v. United States*, 139 S. Ct. 1826, 1832 (2019) (“[T]he definition of ‘is imprisoned’ may well include pretrial detention.”(emphasis added)); *id.* at 1838 (Sotomayor, J., dissenting) (imprisoned “is most naturally understood in context to mean postconviction incarceration”).

<sup>8</sup> Shlomo Klapper, Soren Schmidt & Tor Tarantola, *Ordinary Meaning from Ordinary People*, U.C. IRVINE L. REV. (forthoming 2021) (manuscript at 14–15).

<sup>9</sup> Miranda McGowan, *Do as I Do, Not as I Say: An Empirical Investigation of Justice Scalia’s Ordinary Meaning Method of Statutory Interpretation*, 78 MISS. L.J. 129, 132–33 (2008) (emphasis omitted).

circumstances in which they were used, and it is to the end of answering this last question that we let in evidence as to what the circumstances were.<sup>10</sup>

Later, in *McBoyle v. United States*, Holmes stated:

When a rule of conduct is laid down in words that evoke [a picture] in the common mind . . . the statute should not be extended . . . simply because it may seem to us that a similar policy applies, or upon the speculation that, if the legislature had thought of it, very likely broader words would have been used.<sup>11</sup>

Perhaps without intending to, Holmes has invoked two very different concepts of meaning. The first seems rooted in language production or usage—what “words would mean in the mouth of a normal speaker of English, using them in the circumstances in which they were used.”<sup>12</sup> The second is rooted in language perception—the picture that the words evoke in the common mind. Judges will often invoke either the usage-based<sup>13</sup> or perception-based<sup>14</sup> notions meaning when appealing to the “ordinary meaning” canon.

With respect to the data-driven approaches to interpretation that are the subject of this symposium, we can gain access to evidence of language production and usage through examination of usage evidence in a corpus. But a picture that appears in the common mind presents a different challenge altogether.

Part of the difficulty is that Holmes’ picture is an abstraction of cognitive process of perception and understanding that is not easily described. This is a limitation the law shares with linguistics. “Linguistics describe languages in terms of specially developed, often over any centuries, for the specific task of describing languages, terms such as ‘noun,’ ‘verb,’ ‘clause,’ ‘passive,’ etc.”<sup>15</sup> These terms, however, “are not part of the vocabulary of

<sup>10</sup> Oliver Wendell Holmes, Jr., *The Theory of Legal Interpretation*, 12 HARV. L. REV. 417, 417–18 (1899).

<sup>11</sup> *McBoyle v. United States*, 283 U.S. 25, 26–27 (1931).

<sup>12</sup> See Holmes, Jr., *supra* note 10.

<sup>13</sup> See, e.g., *S.D. Warren Co. v. Me. Bd. of Env'tl. Prot.*, 547 U.S. 370, 377 (2006) (“In resort to *common usage* under § 401, this Court has not been alone, for the Environmental Protection Agency (EPA) and FERC have each regularly read ‘discharge’ as having its *plain meaning*.” (emphasis added)); *United States v. Lopez*, 590 F.3d 1238, 1248 (11th Cir. 2009) (“When a statutory term is undefined, courts give it its ‘*ordinary meaning*’ or ‘*common usage*.’” (emphasis added)).

<sup>14</sup> See, e.g., *Cuellar v. United States*, 553 U.S. 550, 557–58 (2008) (“This is consistent with the *plain meaning* of ‘money laundering,’ petitioner argues, because that term is *commonly understood* to mean disguising illegally obtained money in order to make it appear legitimate . . . . We agree with petitioner . . . .” (emphasis added)); *Amoco Prod. Co. v. Vill. of Gambell*, 480 U.S. 531, 554 (1987) (“Numerous statements by other legislators reveal a *common understanding*—consistent with the *plain meaning*.” (emphasis added)).

<sup>15</sup> James R. Hurford, *Evolutionary Linguistics: How Language and Languages Got to be the Way they Are*, in ROUTLEDGE HANDBOOK OF LINGUISTICS 18–19 (2019).

psychology, neuroscience or biology.” Instead, they are “abstracted from neural mechanism,” but are “nevertheless intended to be ‘psychologically real’ in the sense that neural correlates will presumably eventually be found for the parts of the linguist’s descriptions.”<sup>16</sup> Just as there are no known “neural correlates” for the concepts of “noun” or “verb,” there are no known neural correlates for “ordinary meaning,” or Holmes’ “picture” that the words evoke in the common mind. If we are looking for evidence of linguistic understanding, it is not necessarily obvious what we are looking for or how we would obtain such evidence.

One proposal is that if we are interested in ordinary people’s understanding of the language of a legal text, we ought to ask ordinary people. Survey methods have a long tradition in both linguistics and the law.

And survey methods can gather information about language perception that cannot be obtained with a corpus. But survey methods, like corpus linguistics, have a number of important limitations. Below, I will examine some of the advantages and disadvantages of using corpora and surveys in legal interpretation.

## II. CORPUS LINGUISTICS AND INTERPRETATION

Language evidence from linguistic corpora can be used to evaluate claims about the meaning of words in a legal text. Corpus linguistics is the study of language using evidence from electronic collections of texts.<sup>17</sup> The corpus-based approach to language study is premised “on the idea that the description of the language cannot be made just from the intuition of the linguist, but that it requires the handling of a set of real language samples.”<sup>18</sup>

Linguistic corpora sample natural language—written text and transcribed speech that were produced in a natural setting—rather than speech or text that are gathered through direct methods of elicitation, like surveys, interviews, or observation.<sup>19</sup> These language samples are often annotated (or tagged) with grammatical information (or metadata).<sup>20</sup>

---

<sup>16</sup> *Id.*

<sup>17</sup> See generally TONY MCENERY & ANDREW HARDIE, *CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE* (2012).

<sup>18</sup> Svenja Adolphs & Phoebe M.S. Lin, *Corpus Linguistics*, in *THE ROUTLEDGE HANDBOOK OF APPLIED LINGUISTICS* 597 (James Simpson eds. 2011) (“[C]orpus linguistics is concerned with language use in real contexts.”).

<sup>19</sup> Thuc Anh Vo & Ronald Carter, *What Can a Corpus Tell us About Creativity?*, in *ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS* 302, 303 (Anne O’Keefe & Michael McCarthy 2010).

<sup>20</sup> Elizabeth Walter, *Using Corpora to Write Dictionaries*, in *ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS* 428, 433.

Over time, examining language evidence from linguistic corpora has become one of the principal methods that linguists use to study language. Wherever there is an intersection between language and technology—in machine learning, speech recognition, or language pedagogy—corpus methods are used to gather and evaluate information about language use.<sup>21</sup> Contemporary lexicographers also use corpus evidence in compiling dictionaries.<sup>22</sup>

Linguistic corpora are designed to be representative samples of the speech or writing of a speech community.<sup>23</sup> They are designed “to actually ‘represent’ a domain of language use with a corpus of texts.”<sup>24</sup> The exponential growth in computing power of recent decades has also seen a sharp rise in the size and availability of sophisticated linguistic corpora.<sup>25</sup>

Corpora allow their users to make observations about language that cannot be made through introspection.<sup>26</sup> And they allow their users to make observations about language that are replicable and falsifiable.<sup>27</sup> The size, design, and purpose of corpora may vary widely.<sup>28</sup> Corpora can be built to represent the

<sup>21</sup> See, e.g., Paul Thompson, *Building a specialized audio-visual corpus*, in ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 93, 102; see also Natalie Kübler & Guy Aston, *Using Corpora in Translation*, in ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 501, 513.

<sup>22</sup> See, e.g., OXFORD DICTIONARY OF ENGLISH xi (3d ed. 2010) (“The general principle on which the senses in the *Oxford Dictionary of English* are organized is that each word or part of speech has at least one core sense or core meaning, to which a number of subsenses may be attached . . . Core meanings represent typical, central uses of the word in question in modern standard English, as established by analysis of the Oxford English Corpus and our other language databases.”).

<sup>23</sup> TONY MCENERY & ANDREW WILSON, CORPUS LINGUISTICS: AN INTRODUCTION 75 (2011); MARCYLIENA H. MORGAN, SPEECH COMMUNITIES: KEY TOPICS IN LINGUISTIC ANTHROPOLOGY 1 (2014) (“Speech communities are groups that share values and attitudes about language use, varieties and practices. These communities develop through prolonged interaction among those who operate within these shared and recognized beliefs and value systems regarding forms and styles of communication.”).

<sup>24</sup> DOUGLAS BIBER & RANDI REPPEN, THE CAMBRIDGE HANDBOOK OF ENGLISH CORPUS LINGUISTICS 1 (2015).

<sup>25</sup> See ELENA TOGNINI-BONELLI, CORPUS LINGUISTICS AT WORK 5 (2001).

<sup>26</sup> See Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Novig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden, *Quantitative Analysis of Culture Using Millions of Digitized Books*, 331 SCIENCE 176, 176 (2011) (“The corpus cannot be read by a human. If you tried to read only English-language entries from the year 2000 alone, at the reasonable pace of 200 words/min, without interruptions for food or sleep, it would take 80 years. The sequence of letters is 1000 times longer than the human genome: If you wrote it out in a straight line, it would reach to the Moon and back 10 times over.”).

<sup>27</sup> See MCENERY & HARDIE, *supra* note 17, at 66 (“As a key goal of corpus linguistics is to aim for replicability of results, data creators have an important duty to discharge in ensuring that the data they produce is made available to analysts in the future.”).

<sup>28</sup> For further discussion of the types of linguistic corpora and the tools they offer, see Lee & Mouritsen, *supra* note 4, at 830–35.



linguistic characteristics of any speech community for which there are existing texts or recorded speech.<sup>29</sup>

Corpora can be used to show how frequently words co-occur in a particular context.<sup>30</sup> And corpora can be used to show concordance or key word in context (KWIC) information. A concordance allows its user to review a particular word or word form in hundreds of contexts, all on the same page of running text.<sup>31</sup> And corpora can be used to show evidence of how frequently a key word occurs, and how frequently a given sense of a key word is used in a particular context, by a particular speech community, and at a particular point in history.

Corpus evidence can also be used to evaluate interpretive questions in the law, as illustrated by the examples below.

#### A. *United States v. Muscarello*

The United States Supreme Court's decision *Muscarello v. United States*<sup>32</sup> has been much discussed in law and corpus linguistics literature.<sup>33</sup> This is likely at least in part because *Muscarello* was the first Supreme Court decision that involved "crudely . . . searching computerized newspaper data bases," in order to "make certain that there is no special ordinary English restriction" on the use of the word in question: "carry."<sup>34</sup> *Muscarello* concerned the interpretation of the phrase *carries a firearm* in the Omnibus Crime Control and Safe Streets Act of 1966, later codified as 18 U.S.C. § 924(c)(1), and whether Congress intended by that term to include the notion of conveyance in a vehicle.<sup>35</sup> *Muscarello* was arrested during a minor drug transaction, but at the time of his

---

<sup>29</sup> A register is "any of the varieties of a language that a speaker uses in a particular social context." *Register*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/register> [<https://perma.cc/6C8U-JCAH>]; see Lawrence M. Solan, *The New Textualists' New Text*, 38 LOY. L.A. L. REV. 2027, 2059 (2005) ("When the legal system decides to rely on the ordinary meaning of a word, it must also determine which interpretive community's understanding it wishes to adopt. This choice is made tacitly in legal analysis, but becomes overt when the analysis involves linguistic corpora because the software displays the issue on a screen in front of the researcher." (emphasis added)).

<sup>30</sup> See TONY MCENERY & ANDREW WILSON, CORPUS LINGUISTICS: AN INTRODUCTION 82 (2011).

<sup>31</sup> Christopher Tribble, *What are Concordances and How are They Used?*, in THE ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 167 (quoting JOHN M. SINCLAIR, CORPUS, CONCORDANCE, COLLOCATION, 32 (1991)).

<sup>32</sup> See generally *Muscarello v. United States*, 524 U.S. 125 (1998).

<sup>33</sup> Neal Goldfarb, *A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics*, 2017 BYU L. REV. 1359, 1397–1415 (2017); Lee & Mouritsen, *supra* note 4, at 845–48; Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915, 1916 (2010); Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311, 1345–47 (2017).

<sup>34</sup> *Muscarello*, 524 U.S. at 129.

<sup>35</sup> *Id.* at 126–27.



arrest he had a handgun locked in his glovebox.<sup>36</sup> A five-to-four majority of the Court concluded that Section 924(c) suggested the conveyance in a vehicle meaning.<sup>37</sup>

The *Muscarello* majority framed the dispute by stating that “[a]lthough the word ‘carry’ has many different meanings, only two are relevant here.”<sup>38</sup> The Court then noted that *carry*’s “first, or primary, meaning” reflects the notion carrying as conveyance (hereafter *vehicle-carry*); while only a “different, rather special” meaning of *carry* reflects the notion of carrying upon one’s person (hereafter *personal-carry*).<sup>39</sup>

The Court appealed to a variety of dictionaries to conclude *vehicle-carry* is *carry*’s ordinary meaning. The Court cited dictionaries not only for the content of their definitions, but for the placement of a given sense in the dictionaries’ overall definition structure.<sup>40</sup> The Court also cited the etymology of *carry* to show that *carry* originally meant *vehicle-carry*.

To begin with, the Court observed that “[t]he Oxford English Dictionary gives as its *first* definition ‘convey, originally by cart or wagon, hence in any vehicle, by ship, on horseback, etc.’”<sup>41</sup> The italicized emphasis on “first” is in the Court’s opinion. The Court also cited the “*first* definition” in Webster’s Third—“move while supporting (*as in a vehicle* or in one’s hands or arms)”<sup>42</sup>—and the “*first* definition” in the Random House Dictionary—“to take or support from one place to another; convey; transport.”<sup>43</sup> In each case, the Court printed “*first*” in italics and asked readers to “[c]onsider first the word’s primary meaning.”<sup>44</sup> The implication is clear: The *first* sense is somehow most likely to be the ordinary meaning.

The trouble is that the dictionaries most cited by the Court, including Webster’s Third New International Dictionary and Oxford English Dictionary do not rank their senses in order of “ordinariness;” they rank them historically with the oldest recorded sense listed first.<sup>45</sup> This attempt to rank senses historically should

---

<sup>36</sup> *Id.* at 127.

<sup>37</sup> *Id.* at 139.

<sup>38</sup> *Id.* at 128.

<sup>39</sup> *Id.*

<sup>40</sup> *Id.* at 127–28, 131.

<sup>41</sup> *Id.* at 128 (quoting 2 THE OXFORD ENGLISH DICTIONARY 919 (2d ed. 1989)).

<sup>42</sup> *Id.* (quoting WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 343 (1986)).

<sup>43</sup> *Id.* (quoting RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE UNABRIDGED 319 (2d ed. 1987)).

<sup>44</sup> *Id.*

<sup>45</sup> WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 17a (1961) [hereinafter WEBSTER’S THIRD] (“The order of senses is historical: the one known to have been first used in English is entered first.”); 1 OXFORD ENGLISH DICTIONARY XXIX (2d ed. 1989) (“[T]hat sense is placed first which was actually the earliest in the language: the others follow in order in which they have arisen.”).

not be understood as an exact enterprise. The ranking of senses is based on the unscientific collection of examples of usage in the citation file maintained by the dictionary's publishing house. These citation files are necessarily incomplete samples of the ways that a word is used.<sup>46</sup> For this reason at least one dictionary, Webster's Third, makes clear that "[s]ometimes an arbitrary arrangement or rearrangement is the only reasonable and expedient solution to the problems of ordering senses."<sup>47</sup> The editors of Webster's Third also counsels that "[t]he best sense is the one that most aptly fits the context of an actual genuine utterance."<sup>48</sup>

The *Muscarello* Court also appealed to the etymology of *carry*, arguing that "[t]he origin of the word 'carries' explains why the first, or basic, meaning of the word 'carry' includes conveyance in a vehicle."<sup>49</sup> The Court cited an etymological dictionary for the proposition that the word *carry* comes from the Latin *carum* meaning *car* or *cart*.<sup>50</sup> But the "meaning of a word in a given period of a given language is a matter of usage," and the word's use "at some earlier period or in some cognate language" does not help in determining its current meaning.<sup>51</sup> This is an error in reasoning so common that it has its own name—the Etymological Fallacy.<sup>52</sup> If words were interpreted according to the meaning they had historically, and before they became a part of the English language, then *December* would quite literally mean *October*,<sup>53</sup> and a *skirt* and a *shirt* would reference the same item of clothing.<sup>54</sup>

In place of citing to dictionaries for information about "ordinariness" that they do not contain, or appealing to structural or etymological information found in dictionaries that

---

<sup>46</sup> SIDNEY I. LANDAU, *DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY* 104 (2d ed. 2001)(citation files that are liable to be "unrepresentative of the language as a whole," and reviewed by editors who "all too often ignore common usages and give disproportionate attention to uncommon ones . . ."); see also DOUGLAS BIBER, SUSAN CONRAD & RANDI REPPEN, *CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE* 26 (Cambridge Univ. Press 1996) ("[C]itation slips represent only those contexts that a human reader happens to notice (in some cases representing only the more unusual uses).")

<sup>47</sup> See WEBSTER'S THIRD, *supra* note 45.

<sup>48</sup> *Id.*

<sup>49</sup> *Muscarello*, 524 U.S. at 128.

<sup>50</sup> *Id.*

<sup>51</sup> HENRY SWEET, *THE PRACTICAL STUDY OF LANGUAGES: A GUIDE FOR TEACHERS AND LEARNERS* 88 (1906).

<sup>52</sup> See *Etymological Fallacy*, WIKIPEDIA, [https://en.wikipedia.org/wiki/Etymological\\_fallacy](https://en.wikipedia.org/wiki/Etymological_fallacy) [<https://perma.cc/NLP4-BPC5>].

<sup>53</sup> See *December*, ONLINE ETYMOLOGY DICTIONARY, <https://www.etymonline.com/search?q=December> [<https://perma.cc/7WKB-XFAN>] (noting that December was the "tenth month of the old Roman Calendar" and comes from the Latin *decem*).

<sup>54</sup> *Compare Shirt*, ONLINE ETYMOLOGY DICTIONARY, <https://www.etymonline.com/search?q=shirt> [<https://perma.cc/84SK-GMEY>] with *Skirt*, ONLINE ETYMOLOGY JOURNAL, <https://www.etymonline.com/search?q=skirt> [<https://perma.cc/8VZS-9QNX>] (noting common origin between both words).

has no bearing on meaning, the corpus can provide a clearer sense of the ways in which the verb *carry* is used, particularly in the circumstances relevant here—where there is an implicit human agent doing the carrying and a firearm as the thing that is carried.<sup>55</sup> The News on the Web (NOW) Corpus can provide a “randomized sample of concordance lines featuring *carry*” that can be reviewed to determine the range of possible meanings of *carry* and how commonly used each sense is in circumstances with similar linguistic features to § 924(c)(1).<sup>56</sup> One way to simplify the examination of this usage evidence is to search in the NOW Corpus for instances of *carry* within a few words of *firearm*.<sup>57</sup> The search can be broadened to include the most common synonyms of *firearm* that can be revealed by looking at the collocates of *carry*. These include words like *gun(s)*, *pistol(s)*, *handgun(s)*, and *rifle(s)*.<sup>58</sup> Of the 271 instances of *carry* reviewed, some 104 instances of *carry* had the *personal-carry* sense while only five instances of *carry* had the *vehicle-carry* meaning.<sup>59</sup>

A search of a historical corpus shows a similar result. Section 924(c)(1) was enacted in 1968. The Corpus of Historical American English (COHA) can provide contemporaneous usage evidence from the decade of the statute’s enactment.<sup>60</sup> The COHA includes a collection of annotated texts from each decade going back to 1810. These samples, however, are comparatively small, somewhere between 10 and 20 million words. That means that much less data is available for *carry* in the COHA. There are some twenty-eight instances of *carry* co-occurring with *firearm(s)*, *gun(s)*, *pistol(s)*, *handgun(s)*, or *rifle(s)* in the COHA. Of these, eighteen were instances of *personal-carry* and two were instances of *vehicle-carry*.<sup>61</sup>

Evidence from both corpora suggest that in contexts similar to Section 924(c)(1), *personal-carry* occurs much more commonly than *vehicle-carry*. This suggests that whatever is meant by the

---

<sup>55</sup> This section is drawn largely from the discussion of *Muscarello* in Lee & Mouritsen, *supra* note 4, at 846–48.

<sup>56</sup> In the context of Section 924(c), *carry* is a transitive verb (or “has a transitive argument structure”), with an implied “human subject and a non-human, inanimate, weapon object.” A search in a corpus can reveal uses of the verb *carry* in similar contexts. *Id.* at 833, 846.

<sup>57</sup> “Such a search can be executed as follows: (1) Select ‘Collocates’ on the NOW Corpus homepage; (2) Enter ‘CARRY\_v’ in the ‘Word/phrase’ field; (3) Enter ‘FIREARM\_n’ in the ‘Collocates’ field; (4) Click on ‘Sections’ and select ‘United States’; (5) Select ‘Sort/Limit’ and set the ‘Minimum’ to ‘FREQUENCY’ and ‘3’; and (6) Click ‘Find collocates.’” *Id.* at 846 n.237; see also NOW CORPUS (NEWS ON THE WEB), ENG.-CORPORA, <https://www.english-corpora.org/now/> [<https://perma.cc/6C3B-32TV>] [hereinafter NOW CORPUS].

<sup>58</sup> See Lee & Mouritsen, *supra* note 4, at 847.

<sup>59</sup> See NOW CORPUS, *supra* note 57.

<sup>60</sup> *Corpus of Historical American English*, ENG.-CORPORA, <https://www.english-corpora.org/coha/> [<https://perma.cc/LCB4-BLMA>] [hereinafter COHA].

<sup>61</sup> *Id.*

“ordinary meaning” of *carry*, it cannot mean that *vehicle-carry* is the most common way that the phrase *carries a firearm* is used.

But corpora can do more than provide evidence of which uses of a word are most common in particular contexts. Examining usage evidence from corpora reveal that “*carry* is used more frequently to talk about acts of personally carrying objects (for instance, in one’s hands or arms, or strapped to one’s back), than about events in which objects are transported or carried in a vehicle.”<sup>62</sup> When speakers use *carry* in the *vehicle-carry* sense, they tend to make explicit reference to the vehicle in question.<sup>63</sup>

The corpus evidence can provide a better picture of the way the phrase *carries a firearm* is used. It can also provide some insight into the circumstances in which the *vehicle-carry* sense is implicated. None of this information can be found in a dictionary nor in an appeal to introspection.

#### B. *Ciner Wyoming, LLC v. Rock Springs Royalty Co.*

Evidence from linguistic corpora can do more than simply show how words are mostly commonly used. Corpora can reveal specialized or technical uses that may not be readily apparent. For example, the case of *Ciner Wyoming, LLC v. Rock Springs Royalty Co.* concerned a mineral lease that was drafted in 1961. The lease provided that the interest rate paid by the lessee was “governable” by the federal rate.<sup>64</sup> The question in the case was whether *governable* always means “capable of being governed,” or whether it can have a mandatory sense of “must be governed?” If *governable* means only “capable of being governed” then the lessor had the discretion to choose a royalty rate higher than the federal rate. But if *governable* also has a mandatory sense—i.e., if *governable* can be used as a synonym of *governed*—then the lessor must accept the federal rate.

Wyoming courts, like other U.S. courts, look to dictionaries to determine plain or ordinary meaning. The trouble is that there is no such definition of *governable* to be found in the dictionaries most commonly cited by courts.<sup>65</sup> The Webster’s Third and the

---

<sup>62</sup> See Goldfarb, *supra* note 33, at 1408.

<sup>63</sup> *Id.* at 1408; see also Solan & Gales, *supra* note 33, at 1346 (“When it comes to meaning, we can say that the unmarked meaning of an expression is the way it would ordinarily be understood without further modification.”).

<sup>64</sup> These facts are adapted, altered, and greatly simplified from the case of *Ciner Wyoming, LLC v. Rock Springs Royalty Co.*, No. C-16-77-L (D. Ct. Sweetwater Cty., Wyo 2016). This case is discussed in Stephen C. Mouritsen, *Contract Interpretation with Corpus Linguistics*, 94 WASH. L. REV. 1337, 1386–89 (2019).

<sup>65</sup> The author was unable to locate any dictionary that defines *governable* with a mandatory sense.

Oxford English Dictionary record only the “capable of being governed” definition.<sup>66</sup> But dictionaries do not record every possible meaning of every word. Indeed, dictionaries often omit uncommon words and uncommon senses.<sup>67</sup> And dictionaries will often fail to include a separate definition for a word formed through derivational affixation, i.e., where a new word (and new part of speech) is created by adding a new suffix (*to govern* + *-able*).<sup>68</sup>

Even though there is no dictionary definition that records what is referred to here as the mandatory sense of *governable*, there are a number of words that are formed with the suffix *-able* that suggest a mandatory sense in a legal context. For example, rent that is “payable” on the first of the month is not merely “capable of being paid.”<sup>69</sup> Income that is “reportable” on a tax filing is not merely “capable of being reported.”<sup>70</sup> Taxpayers are not merely capable of paying taxes on their “taxable” income; they *must* pay their taxes.<sup>71</sup> The *Cambridge Grammar of English* makes this distinction, observing that “[t]he modal meaning [of *-able*] is generally like that of *can*. Sometimes, however, it is stronger, like that of *must* or *will*. Compare, for example, *It is payable at any post office* (‘can be paid’) with *It is payable by 15 June* (‘must be paid’) . . . .”<sup>72</sup>

---

<sup>66</sup> See WEBSTER’S THIRD, *supra* note 45, at 982; *Governable*, OXFORD ENG. DICTIONARY, [www.oed.com/view/Entry/80305?redirectedFrom=governable](http://www.oed.com/view/Entry/80305?redirectedFrom=governable) [<https://perma.cc/99C8-BGKD>].

<sup>67</sup> See Michel et al., *supra* note 26, at 177 (“This gap between dictionaries and the lexicon results from a balance that every dictionary must strike: It must be comprehensive enough to be a useful reference but concise enough to be printed, shipped, and used. As such, many infrequent words [and word senses] are omitted.”); see also BIBER ET AL., *supra* note 46, at 36–41 (demonstrating that general-use dictionaries omit relevant senses); *id.* at 26 (“[C]itation slips [from which lexicographers draft dictionary entries] represent only those contexts that a human reader happens to notice . . .”).

<sup>68</sup> Johan de Caluwe & Johan Taeldeman, *Morphology in Dictionaries*, in A PRACTICAL GUIDE TO LEXICOGRAPHY 114, 121 (Piet van Sterkenburg ed., 2003) (“Restrictions on the inclusion and description of derivations that are regular/predictable in their form-content systems are most clearly visible in paper-based dictionaries. In order to keep the volume of a dictionary within reasonable limits, lexicographers have to use frequency criteria when selecting items for [inclusion in the dictionary].”).

<sup>69</sup> See *Payable*, AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE (5th ed. 2012) (defining *payable* as “[r]equiring payment on a certain date; due,” or “[r]equiring payment to a particular person or entity” (emphasis added)).

<sup>70</sup> See *Reportable*, WEBSTER’S THIRD, *supra* note 45, at 1925 (defining *reportable* as “2: required by law to be reported” (emphasis added)).

<sup>71</sup> See *Taxable*, BLACK’S LAW DICTIONARY 1761 (11th ed. 2019) (defining *taxable* as “1. Subject to taxation”); *Taxable*, 17 OXFORD ENGLISH DICTIONARY 678 (2d. ed. 1989) (defining *taxable* as “1. Liable to be assessed (to a tax, impost, or charge); assessable . . . 2. Liable to be taxed; subject to a tax or duty” (emphasis added)).

<sup>72</sup> RODNEY HUDDLESTON & GEOFFREY K. PULLUM, *CAMBRIDGE GRAMMAR OF ENGLISH* 1707 (2002); see also OXFORD DICTIONARY OF ENGLISH 4 (2010) (While the mandatory sense of *governable* is not reflected in any dictionary, the mandatory sense of *-able* is. The Oxford Dictionary of English defines *-able* as follows: *-able* → suffix forming adjectives meaning: 1 able to be: *calculable*. 2 due to be: *payable*. 3 subject to: *taxable*. 4 relevant to or in accordance with: *fashionable*. 5 having the quality to: *suitable* | *comfortable*.)



So, a mandatory sense of *governable* is possible.<sup>73</sup> The next question is whether any actual example of this use of *governable* can be found.<sup>74</sup> Turning to the corpus evidence, we are faced with a number of challenges. The first is frequency. *Governable* is a very uncommon word, and the contested mandatory sense of *governable* is likely an even less common sense of an uncommon word. An examination of concordance evidence in the Corpus of Contemporary American English (COCA)—the most widely used corpus of contemporary American English<sup>75</sup>—reveals not a single use of a mandatory sense of *governable*. Moreover, the 560 million words that make up the COCA, *governable* occurs only forty-two times. Most of these instances unambiguously refer to the notion of “capable of being governed.”<sup>76</sup> *Governable* appears only three times in the COHA—“the largest structured corpus of historical English”—for the decades of 1950 through 1970.<sup>77</sup> Each of those carry the *capable of being governed* sense.

But the mandatory senses of words formed with *-able* (*reportable*, *payable*, *taxable*) tend to appear in legal contexts. So, it makes sense to look for usage evidence in a legal context. An examination of *governable* in a legal context reveals a number of instances where *governable* is used in a mandatory sense. Some of these uses can be revealed simply by searching for *governable* in Westlaw. For example: “[family] status that has long been recognized to be governable by the law of the

---

<sup>73</sup> *Able*, OXFORD DICTIONARY OF ENGLISH 4 (2010). The Oxford Dictionary of English also notes that contrary to popular perception, the morpheme *-able* and the adjective *able*, are etymologically unrelated, even though they have similar meanings. *Id.* (“[O]RIGIN from French *-able* or Latin *-abilis*; originally found in words only from these forms but later used to form adjectives directly from English verbs ending in *-ate*, e.g. *educable* from *educate*; subsequently used to form adjectives from verbs of all types (influenced by the unrelated word *ABLE*), e.g. *bearable*, *saleable*.”).

<sup>74</sup> See WEBSTER’S THIRD, *supra* note 45 (“The system of separating by numbers and letters reflects something of the semantic relationship between various senses of a word. It is only a lexical convenience. It does not evaluate senses or establish an enduring hierarchy of importance among them. The best sense is the one that most aptly fits the context of an actual genuine utterance.”).

<sup>75</sup> See *Corpus of Contemporary American English*, ENG.-CORPORA, <https://www.english-corpora.org/coca/> [<https://perma.cc/F2RA-LRBX>] [hereinafter COCA] (“The Corpus of Contemporary American English (COCA) is the only large, genre-balanced corpus of American English. COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English that we have created . . . [which] offer unparalleled insight into variation in English.”).

<sup>76</sup> *Id.*

<sup>77</sup> See COHA, *supra* note 60 (“The Corpus of Historical American English (COHA) is the largest structured corpus of historical English. It is related to many other corpora of English that we have created . . . [which] offer unparalleled insight into variation in English . . . COHA contains more than 475 million words of text from the 1820s-2010s (which makes it 50-100 times as large as other comparable historical corpora of English) and the corpus is balanced by genre decade by decade. The creation of the corpus results from a grant from the National Endowment for the Humanities (NEH) from 2008-2010.”).

domicile.”<sup>78</sup> Family status is not *capable of being governed* by the law of the domicile, it is *governed* by the law of the domicile. Or “the contract . . . is thus governable by the statute of frauds, not the statutes of wills.”<sup>79</sup> The contract is not *capable of being governed* by the statute of frauds, it must be. There are a number of examples of this usage found in Westlaw.<sup>80</sup> Similar usage examples can be found at LawInsider.com, a database that collects contracts included in public filings SEC document filings.<sup>81</sup> Many of these examples come from the period near in time to the drafting of the mineral lease at issue.

The examples above indicate that—as with *payable*, *reportable*, and *taxable*—*governable* has a mandatory sense that appears to occur in legal contexts, but that is not recorded in dictionaries.

### C. *Some Advantages and Limitations of Corpora*

These examples illustrate some of the uses of corpora in gathering evidence of language usage and evaluating claims about meaning. Corpora can make available information about language use that is not found in a dictionary and may not be accessible via introspection. As illustrated by the *Muscarello* example, corpora can show not just how often a given sense is used comparatively but can also show the use of that word in contexts similar to those of the legal text in question.<sup>82</sup> As in the *Ciner* case, corpora can show differences in language use across different speech communities and registers.<sup>83</sup> And comparative corpora can reveal differences in language use across speech communities. Moreover, corpora can show language use at a given point in history and changes through time. In

---

<sup>78</sup> Woodward v. Stewart, 243 A. 2d 917, 921 (RI 1968).

<sup>79</sup> Eberle v. Ohlheiser, No. HHDCV126029172, 2012 WL 5201312, at \*6 (Conn. Super. Ct. Sept. 27, 2012).

<sup>80</sup> Lee v. St. Joe Paper Co., 371 F.2d 797, 797 (2d Cir. 1967) (“In this diversity suit governable by the law of the State of New York . . . .”); Alexandria, A & FT. S. R. Co. v. Johnson, 59 P. 1063, 1065 (Kan. 1900) (“[T]he contract . . . having . . . been made in Louisiana, and therefore governable by the laws of that state . . . .”); Columbia Lumber & Millwork Co. v. De Stefano, 95 A.2d 914, 917–18 (N.J. 1953) (“[T]heir subject matter is constitutionally governable by rules promulgated by this court . . . .”); Bonwit Teller v. Dist. 65, 142 A.2d 193, 197 (Pa. 1958) (“[T]he conduct . . . herein is solely governable by the law of the Commonwealth of Pennsylvania.”); *In re* Beuckmann, 203 Misc. 447, 448 (N.Y. Sur. Ct. 1953) (“The situation here encountered is governable by the rule established in the case of Hatch v. Bassett.”); Corrigan v. O’Reilly, 107 A.2d 322, 323 (R.I. 1954) (“[T]he transaction between the parties was governable by the law applicable to a promissory note.”).

<sup>81</sup> See LAW INSIDER, <https://www.lawinsider.com> [<https://perma.cc/3Z32-W5HJ>]; see also *Loan Financing Agreement*, LAW INSIDER, <https://www.lawinsider.com/contracts/58GFAXs9BGNMdULL8p4Csn/universal-truckload-services-inc/financing-agreement/2015-12-29> [<https://perma.cc/F8ZA-A39X>].

<sup>82</sup> See *supra* Section II.A.

<sup>83</sup> See *supra* Section II.B.



addition, the searches performed in linguistic corpora are replicable and falsifiable. Language claims based on corpus evidence can be evaluated by performing the same search in the same corpus, or by reviewing the same archived information.

Corpora also have a number of limitations. One way in which corpora fall behind other methods for evaluating claims of language usage is with respect to pragmatic information—that is, the social or spatial context of an utterance. While some pragmatic information may be clear from the text, a text-based corpus is likely to possess limited social or spatial information for a given utterance. This is a clear advantage of survey methods, in which the survey architect can include all of the relevant factual information from the case in question and may even make subtle variations to the information presented to test the contours of the respondents' judgments.

Another limitation is that of representativeness. It is not always clear how much of a given speech community's speech or writing is necessary to ensure that the sample in the corpus is representative. Indeed, this is a matter of some debate among linguists.<sup>84</sup> The corollary of this challenge is that some linguistic phenomena, which are important to legal interpreters, may be rare or unusual and poorly represented in the corpus. No such limitation occurs when a survey is used, as survey architects can create survey prompts using any language they choose.

The potential for bias is also a limitation. While it is the case that a corpus may be constructed to represent the linguistic conventions of any speech community—from any region, ethnic or racial background, or socioeconomic background—it is nevertheless the case that many of the most widely used corpora (particularly in the law and corpus linguistics literature) are based on written sources made up mostly of books, newspapers, and magazines. These corpora will necessarily privilege those who write books, newspapers, and magazines, which will create a bias in favor language usage from mostly white and mostly male writers and speakers. This effect is further amplified the more one examines corpus evidence. In this respect, the corpus does not necessarily create the bias, but simply reports linguistic biases that already exist. One caveat is that the corpus can, at least, be used to identify, study, and record such linguistic bias, and comparative corpora can be constructed to examine differences in usage. And bespoke corpora may be constructed with the intention of representing the linguistic conventions of a variety of speech communities.

---

<sup>84</sup> See generally Douglas Biber, *Representativeness in Corpus Design*, 8 LITERARY AND LINGUISTIC COMPUTING 243 (1993).

### III. INTERPRETATION VIA SURVEYS AND EXPERIMENTS

Gathering usage evidence from linguistic corpora is not the only way to examine questions of legal interpretation through evidence of ordinary usage or understanding. One way to gather evidence of language usage is through the administration of surveys or questionnaires. Such surveys have a number of advantages. For example, surveys can provide direct evidence of the linguistic judgments of respondents from a wide range of backgrounds. And surveys can be specifically crafted to match the parameters of a specific question. Unlike the corpus user, who must carefully construct, search, and comb through the results looking for utterances with similar contextual features to the language problem in question, the survey architect can simply construct the survey to carefully track a context similar to that of the interpretative problem under evaluation. In addition, recent advances in survey administration technology, like Amazon's Mechanical Turk and Lucid, render the assembly of a large pool of relevant respondents and the collection of a large quantity of survey responses easier and less costly to obtain.<sup>85</sup>

Advocates have offered a variety of arguments in favor of using surveys in legal interpretation. To begin with, surveys are widely used in trademark cases, so their use is already familiar to many judges.<sup>86</sup> And the administration of surveys is increasingly practical as new technologies have increased access to a large pool of survey respondents.<sup>87</sup> Surveys may also be used to provide real-world evidence of public perception of contractual plain meaning, as well as real-world evidence of what changes in contractual language would result in a clearer expression of the parties intentions.<sup>88</sup> Additionally, surveys may be used to “test whether refinements in the contract language shown to respondents elicit different majority interpretations.”<sup>89</sup> And surveys with respondents from the “intended audience” of the legal text in question may “capture that meaning more accurately than a judge’s imagination.”<sup>90</sup>

Using surveys, it is argued, may “reduce the opportunity for courts to engage in ‘normative interpretation’”<sup>91</sup> Using surveys to interpretate legal texts may “incentiv[ise]” drafters “to draft short, simple, and widely understandable contractual text ex ante in order

---

<sup>85</sup> See sources cited *supra* note 3.

<sup>86</sup> Omri Ben-Shahar & Lior J. Strahilevitz, *Interpreting Contracts Via Surveys and Experiments*, 92 N.Y.U. L. REV. 1753, 1755–56 (2017).

<sup>87</sup> *Id.* at 1759.

<sup>88</sup> *Id.* at 1758.

<sup>89</sup> *Id.* at 1759.

<sup>90</sup> *Id.*

<sup>91</sup> *Id.* at 1760.

to reduce the risk of misinterpretation.”<sup>92</sup> Surveys may also allow drafters to pretest language to be incorporated in the text.<sup>93</sup> In addition, in the private-law context, parties could “reduce ex post battles of surveys” by specifying “an exclusive survey procedure they expect the court to recognize and may even select the survey company in advance.”<sup>94</sup> Another argument is that survey methods offer interpreters “an understanding of how ordinary people approach ordinary meaning that can transcend any particular construction, making judicial applications of the law more realistic, more consistent, and better aimed at fair notice and the rule of law.”<sup>95</sup>

Linguists have relied upon survey methods to gather evidence of language usage for centuries.<sup>96</sup> Survey methods have also been used to offer evidence of linguistic perceptions and attitudes<sup>97</sup> and language variation.<sup>98</sup> And with respect to emerging technologies for gathering survey responses, there is some evidence that these methods, like Amazon Mechanical Turk, may be as reliable as collecting judgments from in-person surveys.<sup>99</sup>

Recent literature has demonstrated ways in which survey methods may be brought to bear on questions of legal interpretation.<sup>100</sup>

#### A. *State Farm Fire & Casualty Co. v. Moore*

In their recent article, *Interpreting Contracts via Surveys and Experiments*, Professors Omri Ben-Shahar and Lior J. Strahilevitz illustrated the use of surveys in a variety of contract cases. To do so, the authors conducted two surveys in which they showed survey respondents a prompt containing one paragraph that presented “the essence of the contested issue and produced the text of the disputed term in the contract.”<sup>101</sup> The respondents

---

<sup>92</sup> *Id.*

<sup>93</sup> *Id.*

<sup>94</sup> *Id.*

<sup>95</sup> See Klapper et al., *supra* note 8, at 6.

<sup>96</sup> Natalie Shilling, *Surveys and Interviews*, in RESEARCH METHODS IN LINGUISTICS 96, 97 (Robert J. Podesva & Devyani Sharma eds., 2014).

<sup>97</sup> See generally Kathryn Campbell Kibler, *Sociolinguistics and Perception*, in 4 LANGUAGE AND LINGUISTICS COMPASS 377 (2010); Jennifer Ha, Paul Warren & Katie Drager, *Factors Influencing Speech Perception in the Context of a Merger-in-Progress*, 34 J. PHONETICS 458, 459–60 (2006).

<sup>98</sup> See generally Isabelle Buchstaller & Karen Corrigan, *How to Make Intuitions Succeed: Testing methods for Analysing Syntactic Microvariation*, in ANALYSING VARIATION IN ENGLISH 30 (Warren Maguire & April McMahon eds. 2011).

<sup>99</sup> Jon Sprouse, *A Validation of Amazon Mechanical Turk for the Collection of Acceptability Judgments in Linguistic Theory*, 43 BEHAV. RES. 155, 155 (2011).

<sup>100</sup> See Klapper et al, *supra* note 8, at 21–22. See generally Kevin P. Tobia, *Testing Ordinary Meaning*, 134 HARV. L. REV. 726 (2020); Ben-Shahar & Strahilevitz, *supra* note 86, at 1766.

<sup>101</sup> See Ben-Shahar & Strahilevitz, *supra* note 86, at 1783.

were presented a variety of prompts in randomized order.<sup>102</sup> In the first survey, respondents were asked to identify whether they definitely or probably agreed with the plaintiff, definitely or probably agreed with the defendant, or were completely uncertain as to the result.<sup>103</sup> In the second survey, respondents were then asked not only to predict how they would resolve the interpretive question, but also how they believed the “average person” would resolve the question.<sup>104</sup>

In their first experiment, the authors addressed a homeowners’ insurance policy that contained the following exclusion: “This policy does not apply . . . to bodily injury or property damage arising out of business pursuits of any insured except activities therein which are ordinarily incident to nonbusiness pursuits.”<sup>105</sup> Standard homeowners’ insurance policies have coverage for tort liability, “but explicitly exclude injuries arising from commercial or business” activity.<sup>106</sup>

Courts called upon to interpret this provision were split. Numerous courts concluded that the provision was ambiguous.<sup>107</sup> And those courts who were convinced that the exclusion was unambiguous were almost “evenly split as to its meaning.”<sup>108</sup> Some cases may be easier than others. A doctor cannot use a homeowners’ policy as malpractice insurance merely because the doctor runs the clinic from home.

The case of *State Farm Fire & Casualty Co. v. Moore*,<sup>109</sup> involved a babysitter. Specifically, the case involved a homeowner “who was watching her own child and was paid by her neighbors to also care for their son. The neighbor’s son was injured when boiling water from the kitchen stove accidentally spilled on him.”<sup>110</sup> The neighbors sued, and the homeowner asked the insurer to defend her.<sup>111</sup>

When presented with these facts, “36% of the respondents said that the policy either definitely or probably

---

<sup>102</sup> *Id.*

<sup>103</sup> *Id.*

<sup>104</sup> *Id.* at 1783–84 (noting that the survey was administered, and respondents selected by Toluna, a survey research firm).

<sup>105</sup> *Id.* at 1785 (alteration in original).

<sup>106</sup> *Id.*

<sup>107</sup> *Id.* at 1785–86 (citing *Stanley v. Am. Fire & Cas. Co.*, 361 So. 2d 1030, 1032 (Ala. 1978); *Crane v. State Farm Fire & Cas. Co.*, 485 P.2d 1129, 1131 (Cal. 1971); *Nationwide Mut. Fire Ins. Co. v. Collins*, 222 S.E.2d 828, 830-31 (Ga. Ct. App. 1975)).

<sup>108</sup> *Id.* at 1785 (citing *Robinson v. Utica Mut. Ins. Co.*, 585 S.W.2d 593, 595–98 (Tenn. 1979) (comparing case outcomes)).

<sup>109</sup> *State Farm Fire & Cas. Co. v. Moore*, 430 N.E.2d 641 (Ill. App. Ct. 1981).

<sup>110</sup> See Ben-Shahar & Strahilevitz, *supra* note 86, at 1785.

<sup>111</sup> *Id.*

covered the injuries, whereas almost 46% said that the policy definitely or probably did not cover the injuries.”<sup>112</sup>

The authors then administered a second survey, this time rendering the exclusion in plainer terms: “This policy does not apply to bodily injury arising out of business pursuits of the homeowner.”<sup>113</sup> With the terms of the agreement made more explicit, the prior survey results were amplified. “[R]espondents in the treatment group who were shown the new relatively unambiguous version of the policy overwhelmingly sided with the insurer’s interpretation of no-coverage.”<sup>114</sup> The survey showed that “58% of the sample sided with the insurer (compared to 47% in the control group), and only 28% sided with the policyholder (compared to 34% in the control group).”<sup>115</sup> From this the authors conclude both that respondents “pick up on improvements in clarity” and that it would have been possible for the insurer to draft the policy in a way that would foster more agreement. This is important because, among other things, the court itself had found that the provision was ambiguous and therefore resolved the dispute in favor of the insured.<sup>116</sup> One case to be made for survey methods is that they can assist contract drafters to avoid uncertainty in interpretation.

### B. *People v. Harris*

Survey methods have also been deployed to examine cases of statutory interpretation. For example, in their study, *Ordinary Meaning from Ordinary People*, Shlomo Klapper, Soren Schmidt, and Tor Tarantola address the Michigan Supreme Court’s decision in *People v. Harris*.<sup>117</sup>

In *People v. Harris*, the Michigan Supreme Court addressed what the court characterized as the “ordinary meaning” of the word “information” and whether the ordinary meaning of “information” includes the notion of both true information and false information.<sup>118</sup> The relevant statute, the Michigan Disclosures of Law Enforcement Officers Act (the Act), prevents the use of “any information” in a criminal proceeding if that information is under threat of employment sanction.<sup>119</sup>

In *Harris*, three police officers were charged with obstruction of justice after they denied, falsely, that one of them

---

<sup>112</sup> *Id.* at 1786.

<sup>113</sup> *Id.* at 1787.

<sup>114</sup> *Id.* at 1787–88.

<sup>115</sup> *Id.*

<sup>116</sup> *Statement Farm Fire & Cas. Co. v. Moore*, 430 N.E.2d 641, 247 (Ill. App. Ct. 1981).

<sup>117</sup> See Klapper et al., *supra* note 8, at 29–31.

<sup>118</sup> *People v. Harris*, 885 N.W.2d 832, 837 (Mich. 2016).

<sup>119</sup> *Id.* (citing MICH. COMP. LAWS §§ 15.391(a), .393).

had assaulted a civilian.<sup>120</sup> The obstruction charges were all based on statements obtained under threat of employment sanction. Video evidence confirmed that the statements were false.

Both the majority and dissenting opinions referenced evidence from COCA. Writing for the majority, Justice Brian K. Zahra, found that the word “information” could include both true and false information.<sup>121</sup> The court based this conclusion on the notion that “information” often occurred with modifiers indicating both truth and falsity—including, “truthful,” “accurate,” “false,” and “inaccurate.”<sup>122</sup>

In dissent, Justice Stephen J. Markman reached the opposite conclusion, citing, among other things, evidence from the COCA. Specifically, Justice Markman observed:

The term “information” is found within the COCA 168,187 times, and yet it is only modified by the term “truthful” 28 times, “true” 18 times, “accurate” 508 times, “inaccurate” 112 times, and “false” 271 times. In other words, the term “information” is modified by one of these adjectives 937 times.<sup>123</sup>

Justice Markman further noted: “I do not believe that *any* ordinary or reasonable meaning of the word ‘information’ includes false statements, and the majority identifies none. Would one person of a hundred taken at random from the streets of any community of this state disagree regarding this entirely ordinary meaning?”<sup>124</sup>

In their study, Klapper, Schmidt, and Tarantola tested survey respondents’ judgments regarding the meaning of “information” in the *Harris* case.<sup>125</sup> They presented 541 participants with a short summary of the facts of the case, and then asked the respondents if they agreed or disagreed with the proposition: “The officer’s false statements can be used as criminal evidence against them.”<sup>126</sup> Most respondents agreed with the dissent. 73 percent of respondents agreed that the officers’ statements could be used against them because “information” does not necessarily implicate the notion of false information. 33 percent strongly agreed, 24 percent agreed, and 15 percent somewhat agreed.<sup>127</sup> In this case, the survey responses sharply diverged from the interpretation of the *Harris* majority, though they were consistent with the conclusion of the dissent—which conclusion was itself based on an evaluation of

---

<sup>120</sup> *Id.* at 838–39.

<sup>121</sup> *Id.* at 845.

<sup>122</sup> *Id.* at 839–41.

<sup>123</sup> *Id.* at 850 n.14 (Markman, J., concurring in part and dissenting in part).

<sup>124</sup> *Id.* at n.13.

<sup>125</sup> See Klapper et al., *supra* note 8, at 29–31.

<sup>126</sup> *Id.* at 30 (emphasis omitted).

<sup>127</sup> *Id.*



corpus evidence. As discussed below, there are a number of instances in which survey evidence will diverge from corpus evidence, but this is not always the case. The difficulty (and question for future research) will come in understanding why corpus evidence and survey evidence diverge.

### C. *Some Advantages and Limitations of Surveys*

Surveys share some advantages of corpora in legal interpretation. Surveys, like the corpus, may provide evidence of language usage or perception that is not available via intuition and may not be recorded in any dictionary. And surveys may be constructed to show language in context. While there is potential for bias in the construction of a pool of respondents, a survey may be administered to members of any speech community.

Surveys also have some advantages over the use of linguistic corpora when examining questions of usage and meaning. As noted above, surveys allow their creators to take into account much more pragmatic information than may be available for a given linguistic phenomenon in a corpus. And, surveys have no limitations with respect to whether or not a given linguistic phenomenon is rare, as the survey architect can put anything in the survey prompt.

In addition, surveys have a number of obvious disadvantages to corpora. Administering a survey may prove more costly than examining a usage question in a corpus. And while the results of a survey are replicable and falsifiable, replicating a survey may prove more costly than replicating the results of a corpus-based analysis. Moreover, unlike corpora, surveys may be limited in time because “there are no members of the relevant societies left for us to gather opinions from and there were no public surveys of social attitudes undertaken then.”<sup>128</sup>

## IV. WHEN CORPUS EVIDENCE AND SURVEY EVIDENCE DISAGREE

Corpus evidence will sometimes agree with survey evidence. And corpus evidence may, in this respect be thought of as a way to test and validate survey evidence. But what happens if they disagree? Two recent examples from participants in this symposium illustrate this problem.

---

<sup>128</sup> Tony McEnery, Helen Baker & Carmen Dayrell, *Working at the Interface of Hydrology and Corpus Linguistics: Using Corpora to Identify Droughts in Nineteenth-Century Britain*, in *USING CORPUS METHODS TO TRIANGULATE LINGUISTIC ANALYSIS* 52, 53 (Jesse Egbert & Paul Baker eds., 2020).



A. *Is an Airplane a Vehicle?*

In *Judging Ordinary Meaning*, the authors examined the “no vehicles in the park” problem introduced by H.L.A. Hart<sup>129</sup> in his famous debate over positivism and natural law with Lon Fuller.<sup>130</sup> Hart insists that “[p]lainly” the rule “forbids an automobile,” but asks “about bicycles, roller skates, toy automobiles” and “airplanes.”<sup>131</sup> Here, *airplane* invokes an actual case—*McBoyle v. United States*—in which the U.S. Supreme Court held that an airplane was not a vehicle under the National Motor Vehicle Theft Act, which prohibited transporting stolen “vehicles” across state or national borders.<sup>132</sup> In that case, writing for the Court, Justice Oliver Wendell Holmes, Jr., concluded: When a rule of conduct is laid down in words that evoke in the common mind only the picture of vehicles moving on land . . . .<sup>133</sup> Corpus evidence appears to support the *McBoyle* Court’s conclusion. In *Judging Ordinary Meaning*, my coauthor and I examined contemporary and historical usage evidence for the words “vehicle” and “airplane” in the NOW Corpus and COHA.<sup>134</sup> In both corpora, for the relevant periods of the 1910s, 1920s, and the present, there was not a single collocation or concordance line that reflected the use of “vehicle” to mean “airplane.”<sup>135</sup> We concluded: “[B]ased on its absence from any of our corpus data, we might ask if *airplane* is even a possible sense of *vehicle*.” “Airplane” may fit the “necessary and sufficient conditions for inclusion in the class of *vehicles*,” but it is unattested in the corpus.<sup>136</sup>

In his article, *Testing Ordinary Meaning*, Kevin Tobia examines this same question using evidence from survey respondents. He observes that his survey respondents—including ordinary people, law students, and judges—“evaluated airplanes to be vehicles.”<sup>137</sup> From this and other results, Tobia concludes that “ordinary meaning diverges from ordinary use.”<sup>138</sup> But this is only true if we define “ordinary meaning,” as Tobia does, as synonymous with the self-reported linguistic judgments of survey respondents.

---

<sup>129</sup> See Lee & Mouritsen, *supra* note 4, at 836; H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 606–15 (1958).

<sup>130</sup> Lon L. Fuller, *Positivism and Fidelity to Law—A Reply to Professor Hart*, 71 HARV. L. REV. 630, 662–69 (1958).

<sup>131</sup> See Hart, *supra* note 129, at 607–08.

<sup>132</sup> *McBoyle v. United States*, 283 U.S. 25, 26–27 (1931).

<sup>133</sup> *Id.* at 27.

<sup>134</sup> See Lee & Mouritsen, *supra* note 4, at 836–40.

<sup>135</sup> *Id.* at 844.

<sup>136</sup> *Id.* at 844–45.

<sup>137</sup> See Tobia, *supra* note 100, at 735.

<sup>138</sup> *Id.* (emphasis omitted).

Tobia defines “ordinary meaning” as “how readers of the text would actually understand it.”<sup>139</sup> Tobia sometimes uses the phrases “ordinary understanding” or “ordinary judgments” as synonymous with “ordinary meaning.”<sup>140</sup> He states, for example, that the “judges’ use of legal corpus linguistics and dictionary methods did not consistently reflect their ordinary judgments about category membership.”<sup>141</sup> He further states that “dictionaries and legal corpus linguistics often provide dramatically different verdicts from each other,” and “they often provide different recommendations about meaning from what is reflected in ordinary judgments,”<sup>142</sup> and that “the dictionary use did not reflect ordinary judgment[s].”<sup>143</sup> Thus, “ordinary meaning,” “ordinary judgments,” and “ordinary understanding,” are all used by Tobia to refer to some conception of “ordinary people’s understanding of legal texts,” and this understanding is, ostensibly, synonymous with the self-reported judgments of survey respondents.<sup>144</sup>

But, as outlined below, there are reasons to be very skeptical of a claim that “ordinary people’s understanding of legal texts” is reflected in their self-reported judgments.<sup>145</sup> Surveys do not give us the ground truth about language perception. They do not present a direct window into the linguistic perceptions of survey respondents.

#### B. *Taniguchi v. Kan Pacific Saipan, Ltd.*

A second illustration of a divergence between corpus evidence and surveys is highlighted by comparative analyses of the United States Supreme Court’s decision in *Taniguchi v. Kan Pacific Saipan*.<sup>146</sup> *Taniguchi* addressed the question of whether a person hired to translate written documents constitutes an “interpreter” for the purposes of a statute authorizing costs to prevailing parties for interpreters.<sup>147</sup> In *Taniguchi*, the court held that “an interpreter is normally understood as one who translates orally from one language to another.”<sup>148</sup> The Court conceded that the text-to-text translation sense of “interpreter” is possible, but concluded that it is “hardly a

---

<sup>139</sup> *Id.* at 739 (emphasis omitted).

<sup>140</sup> *Id.* at 754, 764.

<sup>141</sup> *Id.* at 764.

<sup>142</sup> *Id.* at 766.

<sup>143</sup> *Id.* at 764.

<sup>144</sup> *Id.* at 736 (emphasis omitted).

<sup>145</sup> *Id.* (emphasis omitted).

<sup>146</sup> See *Taniguchi v. Kan Pac. Saipan, Ltd.*, 566 U.S. 560 (2012).

<sup>147</sup> *Id.* at 562, 566.

<sup>148</sup> *Id.* at 569.

common or ordinary meaning.”<sup>149</sup> The Court even went so far as to characterize this potential sense of “interpreter” as “obsolete.”<sup>150</sup>

In examining the usage evidence for “interpreter” in the NOW Corpus, the authors of *Judging Ordinary Meaning* did not find a “single instance of anyone referred to as an interpreter performing a text-to-text translation” of one language to another.<sup>151</sup> This, the authors note, raises the question not merely of whether “interpreter” is commonly used to refer to text-to-text translation, but whether such a reading is even possible—that is whether the word “interpreter” is *ever* used in reference to text-to-text translation.<sup>152</sup>

When presented with a summary of the facts of *Taniguchi* by the authors of *Ordinary Meaning from Ordinary People*, however, the overwhelming majority of survey participants disagreed with the Supreme Court. Some “76 percent of participants agreed that compensation for a written translator is eligible for recovery; 32 percent agreed strongly, 28 percent agreed, and 16 percent agreed somewhat.”<sup>153</sup>

Here again, we are left with the question of what to make of this difference. If usage evidence disagrees with survey responses, can we meaningfully say that one represents “ordinary meaning” and the other does not, or should we simply conclude that the results of either approach are in error?

### C. *Interpretation and Natural Language*

Linguists have relied on surveys of linguistic judgments (sometimes called acceptability or grammaticality judgments) for decades to gather evidence of language perception. But some skepticism has been raised because responding to a survey question about grammaticality may involve entirely different cognitive processes from that of ordinary communication.<sup>154</sup>

Put simply, when survey respondents are asked for their judgments about language, they are not performing a natural language task—that is, they are not producing language in a natural communicative setting. Instead, they are responding to an experiment in an experimental setting. The introspection and inferential reasoning involved in responding to a linguistic prompt in an experimental setting, and the offering of a

---

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

<sup>150</sup> *Id.*

<sup>151</sup> See Lee & Mouritsen, *supra* note 4, at 849–50 (emphasis omitted).

<sup>152</sup> *Id.*

<sup>153</sup> See Klapper et al., *supra* note 8, at 27.

<sup>154</sup> Thomas R. Lee & Stephen C. Mouritsen, *The Corpus and the Critics*, 88 U. CHI. L. REV. 275, 318–19 (2021) (setting forth similar responses to survey methods).

linguistic judgment, are not the same cognitive task as simply engaging in communication. Thus, “[p]erformance in an experiment, including performance on the standard linguistic task of making grammaticality judgments, cannot be equated with grammatical knowledge. To determine properties of the underlying knowledge system [of language] requires *inferential reasoning, sometimes of a highly abstract nature*.”<sup>155</sup>

Self-reported judgments about language are sometimes said to be unreliable because of their “unavoidable meta-cognitive overtones.”<sup>156</sup> Thus, “there have been continuing doubts about the empirical reliability and theoretical interpretation of judgment data.”<sup>157</sup>

Indeed, because of this skepticism about the reliability of self-reported linguistic judgments, linguists may check their survey evidence against usage evidence from a corpus and, if differences are discovered, conclude that the *survey evidence* is flawed because it does not conform to the corpus evidence, and not vice versa.<sup>158</sup> Linguistic corpora are designed to gather evidence of *natural language* used in natural linguistic settings—writing and reading, speaking and hearing. Most corpora are *not* composed of language that was elicited specifically for the purpose of study or where the speaker, hearer, reader, or writer would have been aware that they were being observed specifically as an object of study.

In the most extreme version of this objection, the comparison of corpus evidence to survey evidence is tantamount to attempting to validate language claims based on evidence of *natural language* with evidence from *unnatural language* (i.e., language produced in the artificial linguistic environment of a survey response). This is because “[e]ven the best designed elicitation tasks are removed from how people use (and think about) language in everyday life, and people’s reports of their linguistic usage may or may not match up with what they actually do.”<sup>159</sup> Surveys may be phrased in a way that will “induce respondents to claim knowledge and use of features they have never heard prior to the research situation.”<sup>160</sup>

<sup>155</sup> Jane Grimshaw & Sara Thomas Rosen, *Knowledge and Obedience: The Developmental Status of the Binding Theory*, 21 LINGUISTIC INQUIRY 187, 188 (1990) (emphasis added).

<sup>156</sup> Shimon Edelman & Morten H. Christiansen, *How Seriously Should We Take Minimalist Syntax?*, 7 TRENDS COGNITIVE SCIS. 60, 60 (2003).

<sup>157</sup> WAYNE COWART, EXPERIMENTAL SYNTAX: APPLYING OBJECTIVE METHODS TO SENTENCE JUDGMENTS 2 (1997).

<sup>158</sup> See generally Joan Bresnan, *Is Syntactic Knowledge Probabilistic? Experiments with the English Dative Alternation*, in ROOTS: LINGUISTICS IN SEARCH OF ITS EVIDENTIAL BASE 75 (Sam Featherston & Wolfgang Sternefeld eds., 2007).

<sup>159</sup> See Schilling, *supra* note 96, at 102.

<sup>160</sup> *Id.*

Survey respondents may be influenced by ordering or framing effects.<sup>161</sup> Consider the case of “airplane” and “vehicle.” Both words are fairly common English words. Both occur very frequently. But in the NOW Corpus, a corpus of billions upon billions of words, there does not appear to be any evidence that the word “vehicle” is ever used to refer to an “airplane.”<sup>162</sup> This does not necessarily mean that this usage is impossible, only that it is unattested in a very large corpus that contains many uses of both words.

But what happens if a survey respondent who has never used the word “vehicle” to refer to an “airplane,” who has never heard the word “vehicle” used to refer to an “airplane,” who has never thought about whether the word “vehicle” can be used to refer to an “airplane,” is asked the question: Is an airplane a vehicle?<sup>163</sup> This question may be the first time the respondent has so much as considered this usage, or seen these two words in the same sentence. And this framing may create the very association that the survey is trying to test. If that is the case, then the answer about whether the usage is acceptable may tell us something about the respondent’s metalinguistic attitudes, but it may not tell us how the respondent ordinarily uses or perceives the word “vehicle.”

This is just one way in which “[t]he traditional method of eliciting language attitudes is plagued by the same problems as elicitations of speech production. The tasks are unnatural and there is no guarantee that the results are reflective of listeners’ genuine attitudes . . . .”<sup>164</sup> This is because “listeners do not have free access to their attitudes or the ability to accurately convey them, or because they do not wish to express negative attitudes they might really hold.”<sup>165</sup>

Framing effects are just one way in which survey methods may suffer from what sociolinguist William Labov referred to as an observer’s paradox. Labov stated that “the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation.”<sup>166</sup> If we are aiming to understand natural language use or natural language perception (i.e., “ordinary understanding”), then unnatural language produced in an artificial setting may not be satisfactory.

---

<sup>161</sup> See *id.* at 103.

<sup>162</sup> See Lee & Mouritsen, *supra* note 4, at 844.

<sup>163</sup> See Tobia, *supra* note 100, at 754 (surveying participants with no legal corpus linguistics or dictionary data about the term “vehicle”).

<sup>164</sup> See Schilling, *supra* note 96, at 106.

<sup>165</sup> *Id.*

<sup>166</sup> WILLIAM LABOV, SOCIOLINGUISTIC PATTERNS 209 (1972).

Labov proposed some workarounds for the observer paradox. He said that “[o]ne way of overcoming the paradox is to break through the constraints situation by various devices which divert attention away from speech.”<sup>167</sup> But even in the face of such mitigation techniques, respondents are still generally aware that they are speaking in an unnatural linguistic setting and subject to evaluation.<sup>168</sup> Moreover, such techniques would impose significant costs in time and resources, and adversely effect one of the principal benefits of new survey technologies like Mechanical Turk and Lucid—that they are comparatively fast and inexpensive methods for administering surveys to large quantities of people.<sup>169</sup>

These challenges to the use of survey methods to evaluate claims about ordinary meaning paint a bleaker picture than is likely warranted. As noted above, survey methods have a wide variety uses, and a wide variety of advantages over other methods of gathering information about language use and perception. But we should be cautious when faced with a disagreement between corpus results and survey results before concluding that one or the other is in error. The better response is to carefully interrogate why the differences exist in the first place and to attempt to account for these differences in research design.

## V. THE PATH FORWARD FOR EVIDENCE-BASED INTERPRETIVE METHODS

If we accept that the object (or one of the objects) of interpretation is to accurately account what is taking place in the mind of the interpreter—the recovery of the Holmsian picture<sup>170</sup>—we are forced to acknowledge that no existing empirical method can give us a complete picture of what is going on in the mind.

We have no way to directly observe the cognitive processes that take place during ordinary communication. We cannot see the picture that words evoke in the mind of a test subject. If we want to know what is going on inside someone’s head during communication, we are left with a collection of imperfect empirical tools to gather evidence about language and meaning. One way or another, we are observing linguistic behavior in order to try to understanding meaning.

Both corpora and surveys allow us to make evidence-based claims about language usage and meaning. Both allow us to check our intuition against evidence. Both corpus methods and survey

---

<sup>167</sup> *Id.* at 209.

<sup>168</sup> See Schilling, *supra* note 96, at 109–10.

<sup>169</sup> See sources cited *supra* note 3.

<sup>170</sup> See Holmes, *supra* note 10, at 418.

methods allow us to make claims about language that are replicable and falsifiable. Both provide access to information that cannot be intuited and may not be available through both ordinary means of perception and through examination of traditional interpretative sources like dictionaries.

My modest proposal for this symposium is that where corpus evidence and survey evidence disagree, we should carefully interrogate why they disagree, and work to take the basis for that disagreement into account when designing further research and evaluating interpretative claims.