

April 2018

# A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics

Neal Goldfarb

Follow this and additional works at: <https://digitalcommons.law.byu.edu/lawreview>

---

## Recommended Citation

Neal Goldfarb, *A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics*, 2017 BYU L. Rev. 1359 (2018).  
Available at: <https://digitalcommons.law.byu.edu/lawreview/vol2017/iss6/6>

This Article is brought to you for free and open access by the Brigham Young University Law Review at BYU Law Digital Commons. It has been accepted for inclusion in BYU Law Review by an authorized editor of BYU Law Digital Commons. For more information, please contact [hunterlawlibrary@byu.edu](mailto:hunterlawlibrary@byu.edu).

## A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics

Neal Goldfarb\*

*Corpus linguistics is more than just a new tool for legal interpretation. Work in corpus linguistics has generated new ways of thinking about word meaning and about the interpretation of words in context. These insights challenge the assumptions that lawyers and judges generally make about words and their meaning. Although the words that make up a sentence are generally regarded as the basic units of meaning, corpus analysis has shown that in many cases, the meaning of a word as it is used in a given context is a function, not of the word by itself, but of the word's interaction with that context. In the many instances in which that is the case, it will often make sense to regard the basic unit of meaning as a multi-word expression that includes not only the word in question but also the relevant parts of the context. That basic insight, which grew out of work on the world's first dictionary based on an electronic corpus, opens the door to new ways for lawyers and judges to analyze issues of word meaning.*

*This Article begins by contrasting two themes that run through legal interpretation: on the one hand, the view that word meanings are clearly delineated abstract entities that exist independently of the use of the word in context, and on the other hand, the view that word meanings exist only in context. (The Article comes down strongly in support of the latter view.) After setting out these two competing themes, the Article introduces the phenomenon of collocation—the tendency of certain words to co-occur disproportionately with certain other words. The study of collocation served as the starting point for the work that ultimately generated the new insights into the nature of word meaning.*

---

\* Butzel Long, PC; [www.LAWnLinguistics.com](http://www.LAWnLinguistics.com). My thinking on the issues discussed in this paper benefited from discussion at the Law and Corpus Linguistics conference held at BYU in April 2016. I want to acknowledge in particular Stephen Mouritsen's presentation on *Muscarello* and my conversations with Mark Davies and Stefan Th. Gries. The title of this paper is cribbed from John Sinclair, *Meaning in the Framework of Corpus Linguistics*, 20 LEXICOGRAPHICA 20 (2004).

© 2017 Neal Goldfarb

*The Article then describes the development of that work, much of which was done as part of creating the first corpus-based dictionary. The Article summarizes some of the findings that were made by the lexicographers, and the conclusions about word meaning that followed from those findings. Some of those conclusions may strike readers as radical, since they call into question many widely held assumptions about word meaning. Finally, in order to demonstrate how the new approach can be used in legal interpretation, the Article undertakes a corpus analysis of the issue in the well-known case of *Muscarello v. United States*: whether driving to the site of a drug deal with a gun in the glove compartment constitutes “carrying a firearm.”*

## CONTENTS

I. INTRODUCTION .....	1361
II. WORD MEANING IN LEGAL INTERPRETATION .....	1364
A. The Dictionary Paradigm .....	1364
B. Words in Context .....	1366
III. COLLOCATION AND RELATED PHENOMENA: A BRIEF INTRODUCTION.....	1368
IV. FROM COLLOCATION TO CORPUS LEXICOGRAPHY.....	1370
A. John Sinclair and the Cobuild Project.....	1370
B. The Significance of Being Corpus-Based.....	1374
C. Insights from Cobuild.....	1378
1. Findings about word meanings.....	1378
2. Conclusions from the findings.....	1383
V. DO WORD MEANINGS EXIST? .....	1388
A. Doubters: Hanks, Atkins, and Kilgarriff.....	1388
B. Reasons for Doubt .....	1389
1. Sense-division differences among dictionaries .....	1391
2. Some further problems regarding sense division ...	1393
C. “Meaning Is Something You Do.” .....	1395
VI. <i>MUSCARELLO V. UNITED STATES</i> : A CORPUS ANALYSIS OF <i>CARRY A FIREARM</i> .....	1397

A. The Case .....	1397
B. Corpus Pattern Analysis.....	1399
C. Analyzing the Corpus Data .....	1402
1. Framing the inquiry .....	1402
2. The data.....	1404
3. Analysis .....	1405

## VII. CONCLUSION: LEARNING TO THINK LIKE A LINGUIST 1415

### I. INTRODUCTION<sup>1</sup>

As is shown by the other contributions to this symposium, corpus linguistics is a new tool for legal interpretation—one that provides an alternative to dictionaries.<sup>2</sup> But the significance of corpus linguistics for legal interpretation goes beyond the methodological. As I hope to show here, work in corpus linguistics (and in particular corpus-based lexicography) has generated new ways of thinking about word meaning and about the interpretation of words in context. These insights challenge the assumptions that lawyers and judges generally make about words and their meaning.

Those assumptions reflect the way that word meaning is treated in the dictionaries that lawyers and judges typically rely on. For example, it is assumed that the basic units of meaning are individual words. And

---

1. This Article uses the following typographical conventions (all of which are in addition to the ordinary uses of the typographic attributes in question):

*Italics* are used (rather than quotation marks) to indicate that the italicized word or expression is used as an example of a linguistic unit, as in “The plural of *corpus* is *corpora*.”

SMALL CAPITALS are used (1) for technical terms from linguistics when they are used for the first time, and (2) for references to semantic roles and conceptual schemas.

‘Single quotes’ are used to indicate that the material in question represents the meaning of a word or other expression.

Sans serif is used for excerpts from corpus results.

2. See also, e.g., *State v. Rasabout*, 2015 UT 72, ¶¶ 54–134, 356 P.3d 1258, 1275–90 (Utah 2015) (Lee, J., concurring in part and concurring in the judgment); James C. Phillips, Daniel M. Ortner & Thomas R. Lee, *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J. F. 21 (2016), <http://www.yalelawjournal.org/forum/corpus-linguistics-original-public-meaning>; Stephen C. Mouritsen, Note, *The Dictionary Is Not A Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915 [hereinafter *The Dictionary Is Not A Fortress*].

word meanings are seen as discrete entities with (in most cases) clear boundaries.

But things are not that simple. One striking effect of work in corpus lexicography has been to significantly undermine the view of words as the basic units of meaning. Under that view, which embodies the principle of COMPOSITIONALITY, the meaning of a sentence is determined by the meanings of the individual words and how they are grammatically combined.<sup>3</sup> But while that can be a useful idealization, it glosses over a major complication. Many frequently used words have multiple possible meanings, and when such a word is used in a sentence, its meaning depends largely on what the rest of the sentence says. So there is a chicken-and-egg problem: how can individual words be regarded as basic units of meaning when the meaning of a word *in* a particular context is itself affected *by* the context? The answer suggested by corpus linguistics is that individual words often are *not* the basic units of meaning. Rather, it makes more sense in many cases to conceive of the basic unit of meaning as a phrase or other multiword expression. And that makes it necessary to revise our views about word meanings.

Another reason for questioning common views about word meaning is that the boundaries between the meanings of different words, or between the different senses of the same word, are often unclear. Drawing lines between different word senses is unavoidably subjective, as is shown by the fact that the lines are often drawn differently by different dictionaries. Although these are not new discoveries—they have been known to lexicographers at least as far back as Samuel Johnson<sup>4</sup>—work in corpus linguistics has given them new force and prominence.

In discussing these ideas, I will proceed in part by summarizing their intellectual history—because the history matters. The crucial period in the development of corpus lexicography was the mid-1980s,

---

3. See generally THE OXFORD HANDBOOK OF COMPOSITIONALITY (Markus Werning et al. eds., 2012).

4. See Patrick Hanks, *Johnson and Modern Lexicography*, 18 INT'L J. LEXICOGRAPHY 243, 257–59 (2005) [hereinafter *Johnson and Modern Lexicography*]; SAMUEL JOHNSON, A DICTIONARY OF THE ENGLISH LANGUAGE B2r, B2v (1st ed. 1755), reprinted in *Johnson, Preface to the Dictionary* (Jack Lynch, ed.), <http://andromeda.rutgers.edu/~jlynch/Texts/preface.html> (last visited Jan. 30, 2018).

when the first dictionary based on data from a computerized corpus was published. The dictionary's editor-in-chief, John Sinclair, was a leading figure in the development of corpus linguistics and a pioneer in using corpus analysis to investigate the nature of word meaning. The ideas that I will discuss here, such as the notion that words are not the basic unit of meaning, are to a large extent attributable to Sinclair, and the dictionary project he headed served simultaneously as a testing ground for those ideas and as a source of further data and insights. The dictionary that resulted is widely regarded as a major advance in lexicography. So the ideas I will be discussing represent more than mere academic theorizing. The theory has been put to a practical test and has been shown to have real-world applications.

In the final portion of the Article, I will demonstrate how the ideas I describe can contribute to legal interpretation. I will take a fresh look at *Muscarello v. United States*, which presented the question whether driving a car or truck with a firearm in the trunk or glove compartment constituted "carrying" the firearm.<sup>5</sup> Although *Muscarello* has already been the subject of a corpus-based analysis by Steven Mouritsen,<sup>6</sup> his analysis focused on which of two dictionary senses of the word *carry* was more common, and therefore did not challenge the prevailing conception of word meaning. My approach to *Muscarello* will differ from Mouritsen's in two respects. Rather than look at which dictionary sense is more common, I will ask a more open-ended question: when viewed without preconceptions, what does the corpus data tell us about how the word *carry* behaves? And I will look at the data through the lens of Corpus Pattern Analysis, a corpus-driven lexicographic approach that focuses on multiword patterns rather than on individual word meanings. My conclusions, like Mouritsen's, seriously undermine the holding in *Muscarello*. But more importantly, the data reveals details about the use of *carry* that are not recorded in any dictionary.

---

5. *Muscarello v. United States*, 524 U.S. 125, 126 (1998).

6. *The Dictionary is Not a Fortress*, *supra* note 2, at 1958–70; *see also* Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, YALE L.J. (forthcoming 2018), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2937468](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2937468) [hereinafter *Judging Ordinary Meaning*].

## II. WORD MEANING IN LEGAL INTERPRETATION

Two contrasting themes run through legal interpretation, both having to do with the meanings of words. One is a view that sees word meanings as well-delineated abstract entities having an almost Platonic existence. The other is a view that word meaning exists only in context (although that principle may be honored at least as much in the breach as in the observance). To set the stage for discussion of what corpus linguistics can tell us about word meaning, I will briefly describe these themes, the first of which is this Article's principal target.

*A. The Dictionary Paradigm*

It has been said that the dictionary is for many people a “guardian of absolute and eternal truth[.]”<sup>7</sup> For these people “there is, for every word, a ‘true’ meaning, that is stored somewhere, and that the job of the lexicographer is to find it and copy it in the dictionary.”<sup>8</sup> This view is analogous to the discredited conception of the common law as being what Holmes dismissively described as “a brooding omnipresence in the sky.”<sup>9</sup> Courts often seem to share this attitude; that at least is a reasonable inference from cases in which courts have equated ordinary meaning with dictionary definitions, or have otherwise treated such definitions as definitive.<sup>10</sup>

Associated with this view of dictionaries are certain attitudes and assumptions about the nature of word meanings, which I will refer to collectively as the Dictionary Paradigm. These attitudes and assumptions are to a large extent reflected in, and fostered by, the way in which the information in dictionaries is presented.

---

7. HENRI BÉJOINT, TRADITION AND INNOVATION IN MODERN ENGLISH DICTIONARIES 122 (1994), *reprinted in* MODERN LEXICOGRAPHY: AN INTRODUCTION 122 (2000) [hereinafter TRADITION AND INNOVATION IN MODERN ENGLISH DICTIONARIES].

8. HENRI BÉJOINT, THE LEXICOGRAPHY OF ENGLISH: FROM ORIGINS TO PRESENT 235 (2010) [hereinafter THE LEXICOGRAPHY OF ENGLISH]; *see also* TRADITION AND INNOVATION IN MODERN ENGLISH DICTIONARIES, *supra* note 7, at 122.

9. *S. Pac. Co. v. Jensen*, 244 U.S. 205, 222 (1917) (Holmes, J., dissenting).

10. *E.g.*, *Taniguchi v. Kan Pac. Saipan, Ltd.*, 132 S. Ct. 1997, 2002–04 (2012); *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176 (2008); *Muscarello*, 524 U.S. at 127–28; *MCI Telecomms. Corp. v. AT&T Co.*, 512 U.S. 218, 225–28 (1994); *Chapman v. United States*, 500 U.S. 453, 461–62 (1991).

A dictionary's structure, with its basic unit being a separate entry for each word, conveys the impression that individual words are similarly the basic units of meaning. That assumption would probably strike most people as unremarkable, but such a reaction simply confirms that most people have internalized the assumption. And as I will attempt to show, the assumption should be questioned.

Many words are POLYSEMOUS, meaning they can be used in a variety of ways, each with its own meaning.<sup>11</sup> That is reflected in the fact that the dictionary entries for such words are divided into multiple separately numbered senses. That format, according to two leading lexicographers, rests on some unstated assumptions.<sup>12</sup> One assumption, which is congruent with the view of dictionaries as guardians of truth, is "that there is a sort of Platonic inventory of senses 'out there.'"<sup>13</sup> On this view, word meanings are abstract entities that exist independently of the contexts in which they are used.

Note that the context I am referring to includes the grammatical environment in which words occur. Word meaning is seen as residing in the domain of dictionaries, and syntax in the domain of grammar books. There are of course exceptions: words (and parts of words) whose function is primarily grammatical, like *of*, *the*, *-ed*, *-ing*, and *-s*. But the attention that dictionaries devote to these "function" words is insignificant in relation to their overall scope.

In addition to treating word meanings as abstract entities, dictionaries assume that each word has exactly as many senses as are listed, with each sense being "mutually exclusive and [having] clear boundaries."<sup>14</sup> This "offers the comforting prospect of certainty to linguistic inquirers"—an apparently complete menu of each word's possible meanings, from which the user can simply select the one appropriate to the situation at hand.<sup>15</sup> And fittingly, given that we are talking here about word meaning in legal interpretation, the standard definition

11. See, e.g., Ingrid Lossius Falkum & Agustín Vicente, *Polysemy: Current Perspectives and Approaches*, 157 LINGUA 1, 1 (2015).

12. B.T. SUE ATKINS & MICHAEL RUNDALL, THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY 271–72 (2008) [hereinafter THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY].

13. *Id.* at 272.

14. *Id.*

15. PATRICK HANKS, LEXICAL ANALYSIS: NORMS AND EXPLOITATIONS 85 (2013) [hereinafter LEXICAL ANALYSIS].



format reflects the demands of the dictionary's role in resolving disputes over meaning.<sup>16</sup> To perform that function, dictionaries "must draw a line around a meaning, so that a use can be classified as on one side of the line or the other"; if a dictionary "presents [a word's] meaning as context-dependent or variable or flexible, [it] will be of little use for purposes of settling arguments."<sup>17</sup>

The discussion above is not intended as a complete description of the Dictionary Paradigm; there are other aspects that are not relevant for present purposes.<sup>18</sup> And I don't mean to suggest that the assumptions I have described play an active role in every case involving an issue of word meaning. But I do think that judges and lawyers deal with such issues in a way that is consistent with these assumptions, and that the assumptions can influence judges' decisions and the ways in which judges explain those decisions.

### B. Words in Context

In at least partial counterpoint to the Dictionary Paradigm is the principle (described by the Supreme Court as fundamental to both statutory interpretation and language itself) that "the meaning of a

---

16. Adam Kilgarriff, "I Don't Believe in Word Senses," 31 COMPUTERS & HUMAN. 91, 109–10 n.13 (1997) [hereinafter "I Don't Believe in Word Senses"], reprinted in PRACTICAL LEXICOGRAPHY: A READER 143 n.14 (Thierry Fontenelle ed., 2008) [hereinafter PRACTICAL LEXICOGRAPHY: A READER].

17. *Id.* As with much else in lexicography, the dispute-resolution function of dictionaries may well trace back to Samuel Johnson. The linguist Geoffrey Nunberg is reported to have suggested that the tradition of relying on dictionaries to resolve disputes can be traced to the cultural environment in which Johnson worked. That connection is described in the following summary of a presentation by Nunberg:

[T]he development of printed discourse, particularly the new periodicals, in England in the early part of the eighteenth century brought about a re-evaluation of the nature of meaning. No longer could it be assumed that a disagreement or confusion about a word's meaning could be settled face-to-face, and it seemed at the time that the new discourse would only be secure if there was some mutually acceptable authority on what words meant. The resolution to the crisis came in the form of Johnson's Dictionary. Thus, from its inception, the modern dictionary has had a crucial symbolic role: it represents a methodology for resolving questions of meaning.

*Id.*; see also Geoff Nunberg, *The Once and Future Dictionary*, (March 17, 1995) (abstract of presentation at Stanford University Linguistics Department colloquium), <http://web.stanford.edu/dept/linguistics/colloq/prev/1995/1995mar17.html>.

18. These issues will be discussed further in the book that I am working on, *Thinking Like a Linguist: Using Linguistics in Legal Interpretation*.

word cannot be determined in isolation, but must be drawn from the context in which it is used.”<sup>19</sup> Taken at face value, this statement could serve to summarize this Article’s thesis. And cases do exist in which the court focuses on the kind of context that is relevant here: the immediate linguistic environment in which the word at issue appears.<sup>20</sup> However, the principle is often cited with regard to aspects of context that are outside this Article’s concerns, such as statutory structure or consistency with other parts of the statute.<sup>21</sup>

An argument frequently made by critics of the courts’ reliance on dictionaries is that dictionaries deal with word meanings out of context.<sup>22</sup> While I agree that the way that courts typically use dictionaries is problematic, it seems to me that the critics’ statement of the problem is not entirely correct.

As I will explain, what lexicographers try to do in writing definitions is to generalize from the ways that the word in question is used out in the world.<sup>23</sup> In all of those uses the word is embedded in a context. So each definition is an attempt to summarize what those instances of meaning-in-context have in common. The problem as I see it is not merely that the dictionaries used by judges and lawyers present word meanings out of context, but that they do little or nothing to identify the kinds of contexts in which each of the various word senses typically appears (a process referred to in linguistics as MAPPING word senses to the appropriate kinds of context). In fact, dictionaries don’t even suggest that such a mapping is possible.

---

19. *Deal v. United States*, 508 U.S. 129, 132 (1993); *accord, e.g., Yates v. United States*, 135 S. Ct. 1074, 1082 (2015); *Tyler v. Cain*, 533 U.S. 656, 662–63 (2001).

20. *See, e.g., Textron Lycoming Reciprocating Engine Div., Avco Corp. v. United Auto., Aerospace & Agric. Implement Workers*, 523 U.S. 653, 657 (1998) (“It is not the meaning of ‘for’ we are seeking here, but the meaning of ‘[s]uits for violation of contracts.’” (alteration in original)).

21. *See, e.g., Reno v. Koray*, 515 U.S. 50, 56–57 (1995).

22. *E.g., James J. Brudney & Lawrence Baum, Oasis or Mirage: The Supreme Court’s Thirst for Dictionaries in the Rehnquist and Roberts Eras*, 55 WM. & MARY L. REV. 483, 502–03 (2013) [hereinafter *Oasis or Mirage*]; Frank H. Easterbrook, *Text, History, and Structure in Statutory Interpretation*, 17 HARV. J.L. & PUB. POL’Y 61, 67 (1994); A. Raymond Randolph, *Dictionaries, Plain Meaning, and Context in Statutory Interpretation*, 17 HARV. J.L. & PUB. POL’Y 71, 74 (1994); *The Dictionary Is Not A Fortress*, *supra* note 2, at 1924–25. For citations to the fairly substantial body of scholarship regarding the courts’ use of dictionaries, see *Oasis or Mirage*, *supra*, at 486–87 nn.3–6.

23. *See infra* notes 56–62 and accompanying text.

Corpus linguistics makes it possible to deal with the mapping problem. The data that corpus linguistics has generated, and the insights that have been drawn from that data, have made it possible to think about the mapping issue systematically and to identify what the mapping process entails. And the methodologies of corpus linguistics provide the means to translate those ideas into practice.

### III. COLLOCATION AND RELATED PHENOMENA: A BRIEF INTRODUCTION

The branch of corpus linguistics that is most important for legal interpretation is corpus lexicography: the use of corpus data in compiling dictionaries, and more generally, the use of such data to investigate word meaning. Although lexicography has long been referred to as a form of applied linguistics,<sup>24</sup> the fact is that until about thirty years ago, lexicography did not pay much attention to linguistics. That situation has changed, however, and current lexicographic practice has been influenced significantly by linguistics (although without much visible effect on the major American dictionaries<sup>25</sup>). This emerging influence has come mainly from corpus linguistics, and in particular from the study of COLLOCATION.

Although the term is used to describe a variety of phenomena, collocation at its core is “a lexical relation between two or more words which have a tendency to co-occur within a few words of each other

---

24. See, e.g., Dirk Geeraerts, *Types of Semantic Information in Dictionaries*, in A SPECTRUM OF LEXICOGRAPHY 1, 1 (Robert Ilson ed., 1987); Hans H. Meier, *Lexicography as Applied Linguistics*, in 3 LEXICOGRAPHY: CRITICAL CONCEPTS: LEXICOGRAPHY, METALEXICOGRAPHY, AND REFERENCE SCIENCE 307, 307 (R.R.K. Hartmann ed., 2003); Alain Rey, *Training Lexicographers: Some Problems*, in LEXICOGRAPHY: AN EMERGING INTERNATIONAL PROFESSION 93, 95 (Robert Ilson ed., 1986).

25. See, e.g., Patrick Hanks, *Lexical Patterns: From Hornby to Huston and Beyond*, in PROCEEDINGS OF THE XIII EURALEX INTERNATIONAL CONGRESS 89, 106 (Elisenda Bernal & Janet DeCesaris eds., 2008) [hereinafter *Lexical Patterns*]; Michael Rundell, *Good Old-Fashioned Lexicography: Human Judgment and the Limits of Automation*, in LEXICOGRAPHY AND NATURAL LANGUAGE PROCESSING: A Festschrift in Honour of B.T.S. Atkins 138, 140 (Marie-Hélène Corréard ed., 2002). I should note that the reference in the text to “major American dictionaries” is not intended to include the *New Oxford American Dictionary*, which is published by the U.K.-based Oxford University Press and which is based on the *Oxford Dictionary of English* (which covers British English). See Patrick Hanks, *Corpus Evidence and Electronic Lexicography*, in ELECTRONIC LEXICOGRAPHY 57, 62 (Sylviane Granger & Magali Paquot eds., 2012) [hereinafter *Corpus Evidence and Electronic Lexicography*].

in running text.”<sup>26</sup> In many cases, collocation is “the lexical realization of the situational context”<sup>27</sup>—a reflection of the topic that is being discussed. Thus, the most frequent collocates of *trial* in the Corpus of Contemporary American English include *murder, fair, lawyers, court, jury, judge, and Simpson (O.J., not Homer)*.<sup>28</sup>

But other collocations are purely linguistic rather than situation-driven. For example, we say that someone *commits a crime*, but typically not that they *do* or *perform* or *make* a crime.<sup>29</sup> A more complex example is provided by the words *strong* and *powerful*. Despite their similarity in meaning, the Google Books Ngram Viewer (American English corpus) reveals that they have different patterns of collocation:<sup>30</sup>

<i>strong</i>	<i>powerful</i>
strong wind	powerful hurricane
strong liquor	powerful engine
strong coffee	powerful car
strong preference	powerful machine
strong opposition	powerful explosive

The tendency of the words in the left-hand column to co-occur with *strong*, and of those in the right-hand column to co-occur with *powerful*, is often referred to as a collocational preference.<sup>31</sup>

26. MICHAEL STUBBS, *WORDS AND PHRASES: CORPUS STUDIES OF LEXICAL SEMANTICS* 24 (2001); see also JOHN SINCLAIR, *CORPUS, CONCORDANCE, COLLOCATION* 170 (1991) [hereinafter *CORPUS, CONCORDANCE, COLLOCATION*].

27. Rosamund Moon, *The Analysis of Meaning, in* LOOKING UP: AN ACCOUNT OF THE COBUILD PROJECT IN LEXICAL COMPUTING AND THE DEVELOPMENT OF THE COLLINS COBUILD ENGLISH LANGUAGE DICTIONARY 86, 92 (J.M. Sinclair ed., 1987) [hereinafter *The Analysis of Meaning*].

28. See Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 560 million words, 1990–present*. Available online at <https://corpus.byu.edu/coca/> [hereinafter *COCA*] (last visited Jan. 22, 2018).

29. This example is taken from THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 303.

30. The difference between how *strong* and *powerful* each behave was first noted in M.A.K. Halliday, *Lexis as a Linguistic Level, in* IN MEMORY OF J. R. FIRTH 148, 150–52 (C. E. Bazell et al. eds., 1966) [hereinafter *Lexis as a Linguistic Level*].

31. In a 2011 Supreme Court case, the author submitted an amicus brief that relied heavily on corpus data regarding collocational preferences. Brief for the Project on Government Oversight, the Brechner Ctr. for Freedom of Info., and Tax Analysts as Amici Curiae in Support of Petitioners, *FCC v. AT&T, Inc.*, 562 U.S. 397 (2011) (No. 09-1279). The brief may have

Words can also have preferences as to the syntactic environments in which they appear, and even more importantly, so can the various possible meanings of a polysemous word. For example, a given meaning may be more strongly associated with one grammatical form of the word than with others.<sup>32</sup> And as we will see, the meaning of a transitive verb as used in a sentence is often sensitive to the word that acts as its direct object.<sup>33</sup> Although phenomena such as these are not typically referred to as a type of collocation, they are closely related to it. In the interest of simplicity, I will use the term to refer not only to purely lexical co-occurrence, but also to this kind of lexical/grammatical co-occurrence.

#### IV. FROM COLLOCATION TO CORPUS LEXICOGRAPHY

##### *A. John Sinclair and the Cobuild Project*

Corpus lexicography is relevant to legal interpretation not only because of its association with dictionary-making, but also because it is a way of studying word meaning. The development of the field was a more-or-less direct outgrowth of the study of collocation by British linguist John Sinclair. While Sinclair was not the first person to focus on collocation,<sup>34</sup> he was the first to seriously study it as an aspect of linguistic theory.<sup>35</sup>

---

influenced the way the opinion was written. See *FCC*, 562 U.S. at 403. For discussion, see Ben Zimmer, *The Corpus in the Court: 'Like Lexis on Steroids,'* ATLANTIC (Mar. 4, 2011), <https://www.theatlantic.com/national/archive/2011/03/the-corpus-in-the-court-like-lexis-on-steroids/72054/>.

32. See, e.g., CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 44–51.

33. See *infra* text accompanying note 85.

34. The study of collocation began (at least with regard to English) in the 1920s and 1930s in Japan in the context of teaching English as a foreign language—and in particular in connection with creating dictionaries intended for that purpose. See, e.g., A. P. COWIE, ENGLISH DICTIONARIES FOR FOREIGN LEARNERS: A HISTORY 2–6, 8–12, 52–58 (1999) [hereinafter ENGLISH DICTIONARIES FOR FOREIGN LEARNERS]. Within linguistics, credit for introducing the concept of collocation is usually given to John Firth, who in 1951 proposed the idea of “meaning by collocation.” J. R. Firth, *Modes of Meaning*, in ESSAYS AND STUDIES OF THE ENGLISH ASSOCIATION 118 (1951), reprinted in J. R. Firth, PAPERS IN LINGUISTICS 190, 194–95 (1957). However, Firth’s writings on the subject were sketchy at best, and are more important for having influenced Sinclair than for what they actually say.

35. Regarding Sinclair’s work and career, see, for example, Patrick Hanks, *John Sinclair (1933–2007)*, EURALEX NEWSL., Summer 2007, in 20 INT’L J. LEXICOGRAPHY 209, 212–13 (2007) [hereinafter *John Sinclair (1933–2007)*]; Michael Stubbs, *A Tribute to John McHardy*

As first framed (in 1966), the object of study had little if anything to do with the meanings of words, but rather was defined as the statistical patterns of co-occurrence between “lexical items” (which might consist of more than a single word).<sup>36</sup> Sinclair recognized that this would raise “problems which are not likely to yield to anything less imposing than a very large computer.”<sup>37</sup> In fact, he was at that point in the preparatory stage of a project devoted to studying collocation using one of the first computerized corpora of spontaneous speech.<sup>38</sup> And despite the purely statistical approach Sinclair had initially outlined, by the time that project was undertaken, its goals included investigating the relationship between collocation and meaning.<sup>39</sup>

That goal proved to be elusive; according to the report that the project generated, the work “tries to define the problem [of the relationship between collocation and meaning] more carefully, without being able to settle the issue.”<sup>40</sup> The absence of progress on that issue is not surprising, considering the small size of the corpus (135,000 words).<sup>41</sup> The primitive state of the technology in the 1960s was also a serious limitation.<sup>42</sup> The project was completed in 1970, and for the next ten years, Sinclair did no further corpus research, focusing instead on technical issues such as improving the necessary software.<sup>43</sup>

But when Sinclair finally ventured back into corpus work, it was to take on an ambitious project that dealt with the issue of meaning

*Sinclair (14 June 1933 – 13 March 2007)*, in *THE PHRASEOLOGICAL VIEW OF LANGUAGE: A TRIBUTE TO JOHN SINCLAIR 1* (Thomas Herbst et al. eds., 2012) [hereinafter *A Tribute to John McHardy Sinclair*].

36. J. McH. Sinclair, *Beginning the Study of Lexis*, in *IN MEMORY OF J.R. FIRTH 410* (C. E. Bazell et al. eds., 1966) [hereinafter *Beginning the Study of Lexis*]; *Lexis as a Linguistic Level*, *supra* note 30, at 148.

37. *Beginning the Study of Lexis*, *supra* note 36, at 410.

38. See J. MCH. SINCLAIR ET AL., *ENGLISH LEXICAL STUDIES: REPORT TO THE OFFICE FOR SCIENTIFIC AND TECHNICAL INFORMATION (1970)* [hereinafter *OSTI REPORT*], reprinted in JOHN SINCLAIR ET AL., *ENGLISH COLLOCATION STUDIES: THE OSTI REPORT 2* (Ramesh Krishnamurthy ed., 2004) [hereinafter *ENGLISH COLLOCATION STUDIES*].

39. *OSTI REPORT*, *supra* note 38, at 3; see generally *A Tribute to John McHardy Sinclair*, *supra* note 35, at 7.

40. *OSTI REPORT*, *supra* note 38, at 3.

41. *Id.* at 5, 18–23.

42. Wolfgang Teubert, *Interview with John Sinclair*, in *ENGLISH COLLOCATION STUDIES*, *supra* note 38, xvii, xix–xx.

43. *Id.*

head on: leading the creation of the world's first dictionary based on a computerized corpus.<sup>44</sup> This project literally created the field of corpus lexicography, and its impact on lexicography more generally has been described as revolutionary.<sup>45</sup>

The project was a collaboration between Birmingham University (where Sinclair was based) and the British publisher Collins.<sup>46</sup> Both the project and the dictionary it produced are referred to by the name of the corpus: "Cobuild," which stood for "Collins Birmingham University International Language Database." The dictionary itself was called the *Collins Cobuild English Language Dictionary*, and it was published in 1987.<sup>47</sup> It is now in its eighth edition, under the name *Collins Cobuild Advanced Learner's Dictionary*, although Sinclair and his team have not been involved since the second edition.<sup>48</sup>

---

44. COLLINS COBUILD ENGLISH LANGUAGE DICTIONARY (John Sinclair et al. eds., 1st ed. 1987) [hereinafter COBUILD1]. On Cobuild's status as the first dictionary to be based on corpus data, see, for example, Rosamund Moon, *Explaining Meaning in Learners' Dictionaries* [hereinafter *Explaining Meaning in Learners' Dictionaries*], in THE OXFORD HANDBOOK OF LEXICOGRAPHY 123, 131 (Philip Durkin ed., 2016) [hereinafter THE OXFORD HANDBOOK OF LEXICOGRAPHY]; John Sinclair, *Introduction to COBUILD1*, *supra*, at xv [hereinafter *Introduction to COBUILD1*]; *A Tribute to John McHardy Sinclair*, *supra* note 35, at 8. A few dictionaries had previously consulted corpus data for limited purposes, but without any significant change in lexicographic practices. See, e.g., SIDNEY I. LANDAU, DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY 279 (2d ed. 2001) [hereinafter DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY]; *Corpus Evidence and Electronic Lexicography*, *supra* note 25, at 61; William Morris, *The Making of a Dictionary—1969*, 20 C. COMPOSITION & COMM. 198, 201 (1969).

45. E.g., ENGLISH DICTIONARIES FOR FOREIGN LEARNERS, *supra* note 34, at 118; HOWARD JACKSON, LEXICOGRAPHY: AN INTRODUCTION 131 (2002).

46. John Sinclair, *Introduction* to LOOKING UP: AN ACCOUNT OF THE COBUILD PROJECT IN LEXICAL COMPUTING AND THE DEVELOPMENT OF THE COLLINS COBUILD ENGLISH LANGUAGE DICTIONARY at vii–viii (J M Sinclair, ed., 1987) [hereinafter LOOKING UP]. The book that I've just cited provides detailed description of the project, with chapters written by Sinclair and several of his colleagues.

47. COBUILD1, *supra* note 44; see LOOKING UP, *supra* note 46 (providing a detailed description of the project, with chapters written by Sinclair and his colleagues). For additional discussion, see, for example, THE LEXICOGRAPHY OF ENGLISH, *supra* note 8, at 177–84; ENGLISH DICTIONARIES FOR FOREIGN LEARNERS, *supra* note 34, at 123–25, 147–48, 152–54, 158–60, 163–65, 170–71; Rosamund Moon, *Sinclair, Lexicography, and the Cobuild Project: The Application of Theory*, in WORDS, GRAMMAR, TEXT: REVISITING THE WORK OF JOHN SINCLAIR 1 (Rosamund Moon ed., 2009).

48. COLLINS COBUILD ADVANCED LEARNER'S DICTIONARY (8th ed. 2014).

Cobuild was from the beginning a “learner’s dictionary,” meaning that it was intended for people learning English as a foreign language.<sup>49</sup> Now, some readers might be dismissive of the idea that a dictionary for learners could represent a serious work of lexicography, thinking of it as a dumbed-down version of a “real” dictionary. But such a view would be mistaken. To begin with, Cobuild is intended for *advanced* learners, and that is reflected in the selection of words it includes; on a single page of the first edition, for example, we find *consortium*, *conspiratorial*, and *consternation*.<sup>50</sup> More importantly, Cobuild and other advanced-learners’ dictionaries provide *more* information in some respects than do dictionaries intended for native speakers. Dictionaries for advanced learners are typically more informative about the grammatical patterns in which the defined words appear and about the words’ most frequent collocates.<sup>51</sup> The reason is obvious: learners need to be given such information explicitly, while native speakers do not. There is, of course, a tradeoff. Learners’ dictionaries focus more on words and word meanings that are relatively frequent: “[A] native-speaker dictionary will tell you less about more whereas a learners’ dictionary will tell you more about less.”<sup>52</sup>

The difference between the two kinds of dictionaries is potentially significant in connection with the use of dictionaries in legal interpretation. Because learners’ dictionaries provide more information about the kinds of contexts in which particular words typically appear, they are less susceptible than native-speaker dictionaries to the criticism that their definitions are acontextual. As I have said, the real problem with lack of context is that the dictionaries on which judges and lawyers most often rely do little or nothing to map individual word senses to their preferred contexts. In contrast, learners’ dictionaries attempt

---

49. See, e.g., ENGLISH DICTIONARIES FOR FOREIGN LEARNERS, *supra* note 34, at 118–21; John Sinclair (1933–2007), *supra* note 35, at 212–13; *Introduction to COBUILD1*, *supra* note 44, at xvi.

50. COBUILD1, *supra* note 44, at 301.

51. See, e.g., THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 300–02; THE LEXICOGRAPHY OF ENGLISH, *supra* note 8, at 197–98; Reinhard Heuberger, *Learners’ Dictionaries: History and Development; Current Issues*, in THE OXFORD HANDBOOK OF LEXICOGRAPHY, *supra* note 44, at 35–38; *Explaining Meaning in Learners’ Dictionaries*, *supra* note 44, at 133–35.

52. THE LEXICOGRAPHY OF ENGLISH, *supra* note 8, at 197 (attributing quote to C. McGregor).



to provide such information. This suggests that lawyers and judges who use dictionaries as interpretive tools should start consulting learners' dictionaries.

*B. The Significance of Being Corpus-Based*

For its first edition, Cobuild used two corpora that together included about twenty million words.<sup>53</sup> Although that is small by today's standards,<sup>54</sup> it was almost 150 times the size of the main corpus that had been used in Sinclair's project in the late 1960s.<sup>55</sup> As we will see, that increase in size made a big difference. But first, it is worth looking at the ways in which the use of corpora improved on prior lexicographic practice. The difference was not a matter of simply automating what had previously been performed by hand. Rather, the use of corpora made possible major improvements in the data that lexicographers relied on and in the methods for analyzing that data. To understand the nature of those improvements, it is necessary to first understand what that data consists of.

Simply stated, lexicographic data consists of information about how words are actually used. Although dictionaries for native speakers typically portray word meanings as abstractions that exist independently of the actual use of the word, that is precisely the opposite of how usage relates to meaning. The meanings of the words used in a given linguistic community are matters of tacit convention, and lexicographers regard it as their job to accurately reflect those conventions. As two leading lexicographers say in their recent textbook, "A reliable dictionary is one whose generalizations about word behavior approximate closely to the ways in which people normally use (and understand) language when engaging in real communicative acts (such as writing novels or business reports, reading newspapers, or having conversations)."<sup>56</sup>

---

53. See Antoinette Renouf, *Corpus Development*, in LOOKING UP, *supra* note 46, at 7, 11–12.

54. See, e.g., THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 58.

55. See OSTI REPORT, *supra* note 38, at 18 (noting that the corpus used for the collocation study contained 135,000 words).

56. THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 46–47. Controversy can arise (mainly outside the lexicographic community) as to how dictionaries treat

To achieve reliability by this standard, lexicographers demand objective evidence: namely, data that comes from “observing language in use.”<sup>57</sup> And that requires “looking at what speakers and writers actually do when they communicate with listeners and readers.”<sup>58</sup> Then, generalizing from “the mass of available language data,” lexicographers attempt “to make explicit the meaning distinctions which—in normal communication—humans deal with unconsciously and effortlessly.”<sup>59</sup>

But what exactly is the nature of the evidence that lexicographers work with, and how do they assemble it? Before there were electronic corpora, and to some extent even now, lexicographers relied on examples of actual usage that had been compiled in “citation files”—collections of sentences excerpted from books, newspapers, magazines, and so on, each of which was selected to illustrate the use of a particular word.<sup>60</sup> The quotations were gathered by hand, one at a time, with each one being recorded on its own slip or card.<sup>61</sup>

Computerized corpora offer two advantages compared to relying on individually gathered citations. The first results from how the data was collected. The selection of individual citations is unavoidably ad hoc and subjective; citation readers are attracted to unusual words and

certain usages that are widespread but nevertheless regarded by some as being wrong, such as the use of *infer* to mean *imply*. See, e.g., *MCI Telecomms. Corp. v. AT&T Co.*, 512 U.S. 218, 228 n.3 (1994); HERBERT C. MORTON, *THE STORY OF WEBSTER'S THIRD: PHILIP GOVE'S CONTROVERSIAL DICTIONARY AND ITS CRITICS* 171–214 (1994) [hereinafter *THE STORY OF WEBSTER'S THIRD*]. However, when views about how language should be used are at odds with how language actually is used, they should play no role in legal interpretation. See Neal Goldfarb, *Prescriptivist Statutory Interpretation? (Part 2 of Scalia and Garner on Statutory Interpretation)*, *LAWNLINGUISTICS* (July 6, 2012), <https://lawlinguistics.com/2012/07/06/prescriptivist-statutory-interpretation-part-2-of-scalia-and-garner-on-statutory-interpretation/>. But cf. ANTONIN SCALIA & BRYAN A. GARNER, *READING LAW: THE INTERPRETATION OF LEGAL TEXTS* 422 (2012) (urging caution in using *Webster's Third New International Dictionary* “because of its frequent inclusion of doubtful, slipshod meanings without adequate usage notes”).

57. *THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY*, *supra* note 12, at 47.

58. *Id.*

59. *Id.* at 311.

60. E.g., *id.* at 48–53; *DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY*, *supra* note 44, at 189–207; *THE STORY OF WEBSTER'S THIRD*, *supra* note 56, at 94–98. The practice goes back at least as far as Samuel Johnson. See HENRY HITCHINGS, *DEFINING THE WORLD: THE EXTRAORDINARY STORY OF DR. JOHNSON'S DICTIONARY* 78–84 (2005).

61. E.g., *THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY*, *supra* note 12, at 50; *DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY*, *supra* note 44, at 190; *THE STORY OF WEBSTER'S THIRD*, *supra* note 56, at 94–95.

unusual meanings, so the ordinary uses of ordinary words are underrepresented.<sup>62</sup> Citation files are therefore less likely to provide reliable evidence of the full range of normal usage.<sup>63</sup> And they cannot provide evidence about the relative frequencies of the various senses of a word, because there is no way to quantify the universe of texts from which the citations were selected.<sup>64</sup> A corpus, on the other hand, is made up of a variety of different types of texts (often including transcripts of spontaneous speech) that are chosen in an effort to approximate the range of conventional use.<sup>65</sup> And because the total size of the corpus is known, computing frequencies is easy.

The second advantage of corpora over citation files has to do with how the data is presented. In a traditional citation file, each citation is an independent mini-text, recorded on its own card or slip of paper.<sup>66</sup> That is not a format designed to facilitate the discovery of recurrent patterns of usage. In contrast, corpus data can be reviewed using a key-word-in-context (KWIC) display, which makes such patterns easier to spot. Such a display (often referred to as a concordance) is in effect a spreadsheet, each line of which presents one use of the key word, with the context that immediately precedes and follows it; the key word appears by itself in a column in the middle, flanked on either side by the columns that present the context, as shown in this example:

---

62. *E.g.*, DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY, *supra* note 44, at 104, 192–93; THE STORY OF WEBSTER'S THIRD, *supra* note 56, at 95.

63. *See* BO SVENSÉN, A HANDBOOK OF LEXICOGRAPHY: THE THEORY AND PRACTICE OF DICTIONARY-MAKING 44–45 (2009) [hereinafter A HANDBOOK OF LEXICOGRAPHY]; *The Dictionary Is Not A Fortress*, *supra* note 2, at 1915–16 (discussing the evidence on this point).

64. *See, e.g.*, DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY, *supra* note 44, at 104; A HANDBOOK OF LEXICOGRAPHY, *supra* note 63, at 44.

65. *See, e.g.*, THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 53–78; Marc Kupietz, *Constructing a Corpus*, in THE OXFORD HANDBOOK OF LEXICOGRAPHY, *supra* note 44, at 63–68.

66. *See, e.g.*, sources cited *supra* note 61.

...Mrs. Beck's third grade students watch her	carry	a cardboard box into the classroom	False
's defenses have been weakened, the fighter will	carry	a heavier load on wing pylons	False
OK. All right, ready? Mr-MOONEN: Poblano peppers	carry	a little bit of heat, so you do n't have to	
...David Souter tonight in New Hampshire. We will	carry	a live-stream of the event on our Web site	False
I have met him. Sometimes the museum asks me to	carry	a message to him. He is muy pacifico. A gentleman	
...Frewer than we did in the trailer; Gordon-Levitt can	carry	a movie in a comic leading role. # Oct. 7 #	
... can mean substantial health benefits but can also	carry	a serious increased risk. (Voiceover) So how should	
in a single green olive without a toothpick. We each	carry	a serving plate and head into the large front room,	
eager to try the new technique there. # Most rocks	carry	a small amount of the element strontium (Sr). All	
every level of the organization. Planning has helped	carry	a young company over its first formidable hump	
boot with a standard liner, and shops typically don't	carry	an extensive offering of custom-fit boots. So...	
this, of course, did nothing to help the administration	carry	out a coherent Middle Eastern policy. 39 # Kissinger	

Reviewing a concordance is obviously different than reading an independent text. While “[a]n individual text is designed to be read as a whole, linearly, from left to right,” concordances are “designed to be read as a series of fragments, vertically, from top to bottom.”<sup>67</sup> It is easy to scan down the page (or monitor screen) and see patterns: recurring meanings, frequent collocates of the key word, grammatical constructions in which the key word appears, and so on. For example, in the short concordance excerpt above, the pattern that should be readily apparent is the frequent occurrence of nonliteral uses of *carry*. The format makes it easy to identify the different types of such uses. As Sinclair said, “The language looks rather different when you look at a lot of it at once.”<sup>68</sup>

Due to the combined effect of better data and a better way to review it, the Cobuild lexicographers found the corpus data to be a revelation. In an account of the project, one of them wrote, “English had been thoroughly described many times, yet the team felt that they were discovering it like a new-found territory, and mapping its features and composition as if for the first time from a scientific perspective.”<sup>69</sup> The project’s managing editor is similarly enthusiastic about what he

67. See, e.g., Michael Stubbs, *The Search for Units of Meaning: Sinclair on Empirical Semantics*, 30 APPLIED LINGUISTICS 115, 117 (2009).

68. CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 100.

69. Rosamund Moon, *The Cobuild Project*, in 2 THE OXFORD HISTORY OF ENGLISH LEXICOGRAPHY 436, 442 (A. P. Cowie ed., 2009) [hereinafter *The Cobuild Project*].

describes as the “overwhelming” impact of corpora: “At last lexicographers have sufficient evidence to make the generalizations that they need to make with reasonable confidence. We can now see that pre-corpus lexicography was little more than a series of stabs in the dark.”<sup>70</sup>

The view that corpora provide better data than citation files is widely shared.<sup>71</sup> In *Dictionaries: The Art and Craft of Lexicography*, Sidney Landau said that “the electronic corpus has replaced the citation file as the essential research tool in general lexicography,”<sup>72</sup> and that “other things being equal, any new dictionary not based on a linguistic corpus is bound to be inferior to one that is.”<sup>73</sup> Similarly, Howard Jackson says in *Lexicography: An Introduction*, “It is no exaggeration to say that computer corpora have revolutionized the lexicographic process, in terms both of the quality of lexical data that can be obtained and of the reliability of the conclusions that can be drawn from that data.”<sup>74</sup>

### *C. Insights from Cobuild*

Having seen what corpora can offer by way of improvements in lexicographic tools, let’s look at the insights regarding word meaning that those tools helped to deliver. I will divide those results into two broad categories, which I will label *Findings* and *Conclusions*. The first category will deal with the evidence of usage patterns that the corpora revealed, and will draw some generalizations from that evidence. The second will shift the emphasis more toward the theoretical, by considering what the evidence and generalizations can tell us about how meaning arises from the words that make up a text.

#### *1. Findings about word meanings*

I will begin this discussion by approaching the specifics of word meaning from the side, as it were, by talking about what the Cobuild

---

70. Patrick Hanks, *The Impact of Corpora on Dictionaries*, in CONTEMPORARY CORPUS LINGUISTICS 214, 230 (Paul Baker ed., 2009) [hereinafter *The Impact of Corpora on Dictionaries*].

71. See, e.g., THE LEXICOGRAPHY OF ENGLISH, *supra* note 8, at 368–69.

72. DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY, *supra* note 44, at 193.

73. *Id.* at 77.

74. HOWARD JACKSON, LEXICOGRAPHY: AN INTRODUCTION 169 (2002) (citation omitted).

data revealed about the relative frequencies of the various senses of polysemous words. I do that partly because talking about frequency will provide a good entry point into the issue of word meaning, and partly because frequency is independently relevant to legal interpretation

Courts sometimes seem to equate “ordinary meaning” with “most common meaning.”<sup>75</sup> But such an appeal to overall frequency will often be misguided; as we will see, the meaning of a particular usage of a word is more likely to be determined by the immediate linguistic context in which it appears than by which sense of the word is the most frequent in general.<sup>76</sup> So the relative frequency of different senses will typically be relevant only if the inquiry focuses on the specific usage that is at issue.<sup>77</sup>

But even putting all of that aside, if courts are going to consider the relative frequencies of different senses, they will need to find out which sense is the most common one. And the corpus data suggests that our intuitions about frequency are often unreliable—as are frequency judgments in dictionaries that are not corpus-based.<sup>78</sup>

75. *See, e.g.*, *Yarbro v. Comm’r*, 737 F.2d 479, 483 (5th Cir. 1984); *Mu-Hun Kim v. Ariz. State Bd. of Dental Exam’rs*, No. 1 CA-CV 10-0374, 2011 WL 797466, at \*2 (Ariz. Ct. App. Mar. 8, 2011); *Wallbeoff v. Wallbeoff*, No. FA064004613S, 2009 WL 4282286, at \*3 (Conn. Super. Ct. Nov. 3, 2009); *People v. Ellison*, No. 313422, 2014 WL 806115, at \*2 (Mich. Ct. App. Feb. 27, 2014).

76. *See infra* notes 84–113 and accompanying text.

77. For an example of such an inquiry, see my examination of *Muscarello v. United States* in Part VI, below, and the prior discussion of the same case by Stephen Mouritsen, *The Dictionary is Not a Fortress*, *supra* note 2, at 1958–70. *See also, e.g.*, *In re Adoption of Baby E.Z.*, 2011 UT 38, ¶ 96, 266 P.3d 702, 726 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment) (“I cannot imagine how we can have a meaningful conversation about the ‘ordinary’ meaning of a statutory term without asking how a given term is most commonly used in a given context.”); *Judging Ordinary Meaning*, *supra* note 6, at 9 (“[E]ven a common sense of a term might not be the most frequent use of it in a certain context.”).

78. John Sinclair, *Introduction to LOOKING UP*, *supra* note 46, at vii; Patrick Hanks, *Evidence and Intuition in Lexicography*, in *MEANING AND LEXICOGRAPHY* 31, 35 (Jerzy Tomaszczyk & Barbara Lewandowska-Tomaszczyk eds., 1990); *The Dictionary Is Not A Fortress*, *supra* note 2, at 1935–36. Of the pre-corpus dictionaries in which the ordering of senses was based at least in part on the editors’ views about the senses’ relative frequencies, the most important was probably the Random House Dictionary. *See A Guide to the Dictionary*, in *THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE* xv, xix (unabridged ed. 1967).

Contrary to what one might expect, a word's most frequently appearing sense is in many cases not its most literal or concrete sense.<sup>79</sup> As one of the Cobuild lexicographers wrote later, "metaphorical meanings might be much more frequent than literal ones, as with *reflect/reflection* and *torrent*."<sup>80</sup> For many other words, "the most frequent uses were phrasal or delexicalized."<sup>81</sup> Phrasal uses include fixed expressions such as *of course*, instances of which "vastly outnumbered instances of *course* used in reference to education, courses of action, or routes."<sup>82</sup> Delexicalized uses include verb phrases in which the verb itself has little meaning, and most of the meaning is provided by the direct object, as in *give a speech*, *have a conversation*, and *take a guess*.

Looking at the delexicalized uses of verbs provides a gateway to new ways of thinking about word meaning—especially with respect to words that are polysemous. When one thinks about a polysemous word in isolation, what most often comes to mind first is not a delexicalized sense like one of those discussed above, but a sense that most people would describe as the word's "literal," "basic," or "primary" meaning.<sup>83</sup> For *give*, that would be 'transfer'; for *have*, it would be 'possess'; and for *take* it would probably be something like 'acquire possession by grasping.' Perhaps the most important difference between the two kinds of meanings is that the literal/basic/primary sense has a stable and relatively definable meaning independent of a particular context, while the delexicalized sense does not. Consider the following expressions and try to describe the meaning that is contributed to each one by the verb:

give meaning [to something], give reason to believe [something], give effect to [something], give priority to [something], give thought to [something], give [somebody] a break, give advice  
have a conversation, have an effect/influence, have a party, have a meeting, have a seat, have sex, have an experience, have a chance

---

79. CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 112–13; LOOKING UP, *supra* note 46, at vii; John Sinclair, *The Dictionary of the Future*, 36 LIBR. REV. 268, 272–73 (1987) [hereinafter *The Dictionary of the Future*]; see *The Cobuild Project*, *supra* note 69, at 448–50.

80. *The Cobuild Project*, *supra* note 69, at 443.

81. *Id.*

82. *Id.*

83. See CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 113.

take a walk, take a nap, take a swim, take a shot, take a swing, take the trouble to [do something], take care of [something], take a chance, take a vacation, take charge, take the blame

I suspect that you will find this task to be challenging. In some cases, the verb seems to serve no purpose other than enabling the event denoted by the phrase to be designated by a noun rather than a verb (*give thought*, *have a conversation*, *take a walk*). In other cases, the verb seems to have more semantic content, but only in an abstract and generalized sense. For example, one can view the events denoted by *give advice* and *take the blame* as involving metaphorical acts of giving and taking. But the same can't be said for some of the other examples. In what way does the phrase *have a meeting* evoke the element of possession that is integral to the literal meaning of *have*? In what way does *give thought to [something]* or *take a walk* evoke even metaphorical acts of giving or taking?

The fact is that these phrases are not fully compositional—their meanings are not fully determined by what the words mean and how they are syntactically combined. Phrases like the ones above therefore pose a problem for the Dictionary Paradigm. That is a topic that I will return to, but not until after discussing a number of issues revealed by the corpus data that pose similar problems.

One of those issues relates to “high-frequency general nouns” such as *fact*, *matter*, *time*, and *way*. The corpus data showed that these were most often used in multiword expressions whose meanings were, like those of the verb phrases discussed above, not fully compositional.<sup>84</sup> (And I would add that some of the expressions exhibit a degree of delexicalization.) As reflected in the entries in Cobuild (and as can be seen in COCA), these include the following:

the fact that, in fact, in point of fact, as a matter of fact, the fact is, is that a fact, the fact remains  
 no matter, no matter who/what/when/where/why, what's the matter, a matter of course/time/opinion, for that matter  
 at times [=sometimes], at all times, at any time, at the same time [=simultaneously], at the same time [≈various discourse functions], (just) in time, out of time [=no time left], out of time [=late (in litigation context)], behind/ahead of time, for the time being, make time, take time, pass the time, waste time

---

84. *The Cobuild Project*, *supra* note 69, at 444.



by the way, by way of, all/part of/some of/most of the way, come a long way [literally and/or figuratively], in the way, out of the way [=not obstructing], out of the way [=away from crowds and hard to get to], go one's own way, way out there, no way! [often followed by *dude*], way [as response to preceding item] (Alright, the last two didn't really come from Cobuild.)

Of these words, it seems to me that *time* displays the smallest degree of delexicalization, so that the various expressions that it appears in preserve a good deal of the meaning inherent in the noun. But what is interesting is that different aspects of that meaning—different ways of thinking about time—are highlighted in different expressions:

*time as a point in the flow of time*: at the same time (as), (at) a different time, at this time, at that time

*time as an occasion or instance of an event*: the first/second/. . . time, that time, this time, last time, next time, every time, many times, another time [=again], time after time,

*time as duration*: a long time, a short time, a day's time, a period of time, the allotted time,

*time in relation to need, obligation, opportunity, etc.*: in time, on time, ahead of time, out of time, enough time, insufficient time, now is the time, the wrong time, it's time to leave

*time as a resource*: my/your/her. . . time, enough time, insufficient time, need more time, [action, process, etc.] takes time, take your time, out of time, keep track of your time

*time as an experience*: have a good/bad time, a difficult time, give someone a hard time, having the time of my life

*time as (pre)history*: the time of the dinosaurs, in colonial times, in Isaac Newton's time, before my time, the best/worst of times

What we're seeing here is that *time* has a range of possible meanings and that in a given use of the word, the intended meaning is activated by the grammatical and collocational context in which the word is embedded.

The last phenomenon that I will discuss here is that of transitive verbs. We have already looked at delexicalized uses of transitive verbs, but I will focus now on uses in which the verb is not delexicalized (or at least not very much). And what we will see is that the meaning of

the verb in context is in many cases determined by the word that functions as the verb's direct object:<sup>85</sup>

throw a football, throw a football game, throw a party, throw a fit, throw a switch  
 drop a dish, drop a course, drop a hint, drop a claim (from a lawsuit)  
 observe a religious holiday, observe somebody's activities  
 break a window, break a date, break a law  
 sell a product, sell an idea  
 smoke a cigarette, smoke a salmon  
 file a piece of metal, file a lawsuit  
 grill a steak, grill a witness  
 examine a witness, examine a patent, examine a patient  
 exhibit symptoms, exhibit paintings  
 run a race, run a machine, run a risk, run an advertisement, run an errand  
 pull a rope, pull rank, pull a prank, pull a gun, pull a switch

In those examples, the differences in meaning for each verb are fairly substantial. But in other instances, the difference is subtler:

cut the bread, cut the grass, cut somebody's hair [different types of cutting]  
 write a novel, write your name [different levels of creativity & composition]  
 bake a potato [=cook a potato by baking], bake a cake [=create a cake by baking cake batter]

In still other cases, a particular meaning is activated by a more complex construction, such as V+NP+PP (Verb plus Noun Phrase plus Prepositional Phrase):

smoke somebody out of hiding  
 sell somebody out, sell a stock short  
 throw something out, throw somebody out of somewhere  
 run somebody out of town, run something into the ground

## 2. *Conclusions from the findings*

Based on the kind of data I have been discussing, the Cobuild team concluded that distinctions among the various possible meanings of a given word depended on context in some very specific ways. Specific meanings were associated with textual patterns that could be ob-

---

85. This phenomenon was known before Cobuild. See ALEC MARANTZ, ON THE NATURE OF GRAMMATICAL RELATIONS 25 (1984); Uriel Weinreich, *Webster's Third: A Critique of Its Semantics*, 30 INT'L J. AM. LINGUISTICS 405, 405-06 (1964).

jectively identified and described, such as collocations between different words and collocations between specific words on the one hand and particular grammatical categories and patterns on the other.<sup>86</sup> On a practical level, this enabled the lexicographers to use that patterning as a guide in deciding how to draw the lines dividing the different senses of a word.<sup>87</sup> It also heavily influenced the way in which the entries in the dictionary were written.

The emphasis in writing the entries was very much on showing the kinds of context in which each meaning typically occurred, with *context* referring to both lexical collocations and grammatical structures.<sup>88</sup> One way this was done was through the examples that were given of the word in use, which were much more extensive than in dictionaries for native speakers, and which were almost entirely taken from the corpus.<sup>89</sup> Sinclair insisted that it was vital to draw examples from real text: “Those dictionaries that stand as milestones in our cultural history use real citations: Dr. Johnson’s Dictionary of 1755 and the Oxford English Dictionary begun by Murray in 1878. They understood that the dictionary is really just a commentary on the examples; the examples have a justification of their own.”<sup>90</sup> Context was also emphasized through the novel style in which Cobuild’s explanations were written.<sup>91</sup> (“Explanations” was the word used by the Cobuild team in place of “definitions,” based on their view that it better described what

---

86. See, e.g., *The Cobuild Project*, *supra* note 69, at 443–44; *The Analysis of Meaning*, *supra* note 27, at 90–94; COBUILD1, *supra* note 44, at xvii.

87. *The Analysis of Meaning*, *supra* note 27, at 89–94.

88. See, e.g., JOHN SINCLAIR, *Words About Words*, in CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 123–37; Patrick Hanks, *Definitions and Explanations*, in LOOKING UP, *supra* note 46, at 116–36 [hereinafter *Definitions and Explanations*].

89. E.g., Gwyneth Fox, *The Case for Examples*, in LOOKING UP, *supra* note 46, at 137–49; COBUILD1, *supra* note 44, at xv–xvi.

90. *The Dictionary of the Future*, *supra* note 79, at 269.

91. The style of the explanations and its rationale are described in CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 123–37; *Definitions and Explanations*, *supra* note 88, at 116–36; and *The Dictionary of the Future*, *supra* note 79, at 270–71. For an evaluation of the style that offers both praise and criticism, see Michael Rundell, *More than One Way to Skin a Cat: Why Full-Sentence Definitions Have Not Been Universally Adopted*, in PROCEEDINGS XII EURALEX INTERNATIONAL CONGRESS 323 (Elisa Corino et al. eds., 2006), reprinted in PRACTICAL LEXICOGRAPHY: A READER, *supra* note 16, at 197–210.

the entries actually did.<sup>92</sup>) As Sinclair explained in his introduction to the first edition, “The word being explained is normally mentioned in the explanation in such a way that you can see how it is typically used in English. In many cases, the explanation provides an illustration of the word in its typical grammatical context.”<sup>93</sup> Thus, the first sense for *conceal* was given as follows: “If you **conceal** something, you cover it or hide it carefully, so that it cannot be seen.”<sup>94</sup> The theory behind this style of wording is explained in Sinclair’s introduction, which simultaneously indicates the kind of linguistic data that was deemed relevant in crafting the explanation: “[The wording] suggests that the verb in this sense is typically used with a human subject and a wide range of direct objects, which are typically inanimate or abstract rather than human.”<sup>95</sup>

Underlying these practical aspects of the Cobuild dictionary was an explicit theory of word meaning. And that theory, which is set forth in several of Sinclair’s writings, largely repudiates some unstated assumptions that underlie the Dictionary Paradigm.<sup>96</sup>

To begin with, Sinclair almost entirely rejected the idea that individual words constitute the primary units of meaning. In a paper that appeared shortly before the first edition of Cobuild was published, he said, “Most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up text.”<sup>97</sup> In the introduction to Cobuild’s first edition, he said that the meanings of particular words are “[o]verwhelmingly . . . bound up with a particular usage—a syntactic pattern,

92. *Definitions and Explanations*, *supra* note 91, at 116–36; *The Cobuild Project*, *supra* note 69, at 448–50; *The Dictionary of the Future*, *supra* note 79, at 270–71.

93. *Introduction to COBUILD1*, *supra* note 44, at xvi.

94. COBUILD1, *supra* note 44, at 287 (boldface in original).

95. *Introduction to COBUILD1*, *supra* note 44, at xvi.

96. *E.g.*, John Sinclair, *The Phrase, the Whole Phrase, and Nothing but the Phrase*, in PHRASEOLOGY: AN INTERDISCIPLINARY PERSPECTIVE 407 (Sylviane Granger & Fanny Meunier eds., 2008); John Sinclair, *The Lexical Item*, in CONTRASTIVE LEXICAL SEMANTICS 1 (Edda Weigand ed., 1998), *reprinted in* TRUST THE TEXT: LANGUAGE, CORPUS AND DISCOURSE 131–48 (John Sinclair & Ronald Carter eds., 2004); John Sinclair, *The Search for Units of Meaning*, 9 TEXTUS 75 (1996) [hereinafter *The Search for Units of Meaning*], *reprinted in* TRUST THE TEXT, *supra*, at 24–48.

97. John McH. Sinclair, *First Throw Away Your Evidence*, in THE ENGLISH REFERENCE GRAMMAR: LANGUAGE AND LINGUISTICS, WRITERS AND READERS 56, 64 (Gerhard Leitner ed., 1986), *reprinted in* CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 108.

perhaps, or a close association of words or a grouping of words into a set phrase.” As a result, Sinclair continued, “It is not really possible to talk about the meaning of [a] word in isolation—it only has a particular meaning when it is in a particular environment.”<sup>98</sup>

In another paper from the same year, Sinclair set out what was perhaps the most complete statement of his views about word meaning.<sup>99</sup> He argued that “in order to explain the way in which meaning arises from language text, we have to advance two different principles of interpretation.”<sup>100</sup> The first one he called “the open-choice principle,” under which words are seen as having the freedom to appear anywhere in a text, subject only to the constraints of syntax.<sup>101</sup> What is important for present purposes is that this principle is generally consistent with the Dictionary Paradigm in treating individual words as basic units of meaning. The second principle was what Sinclair called “the idiom principle.”<sup>102</sup> Under the idiom principle, “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments.”<sup>103</sup> In Sinclair’s view, the idiom principle provided the default interpretive principle for “normal texts,” but interpretation will switch into the open-choice mode as needed.<sup>104</sup> However, “[s]ome texts may be composed in a tradition which makes greater than normal use of the open-choice principle.”<sup>105</sup> (One genre he described as possibly falling into this category was “legal statements”;<sup>106</sup> he was apparently unaware of how much formulaic language is found in legal documents.)

---

98. COBUILD1, *supra* note 44, at xvii.

99. John Sinclair, *Collocation: A Progress Report, in 2 LANGUAGE TOPICS: ESSAYS IN HONOUR OF MICHAEL HALLIDAY* 319 (Ross Steele & Terry Threadgold eds., 1987), *reprinted in* CORPUS, CONCORDANCE, COLLOCATION, *supra* note 26, at 109–21 [hereinafter *Collocation*].

100. *Collocation*, *supra* note 99, at 109.

101. *Id.* at 109–10.

102. *Id.* at 110–15.

103. *Id.* at 110 (British spelling changed to American spelling).

104. *Id.* at 114.

105. *Id.*

106. *Id.* The other was poetry. *Id.*

Although Sinclair's use of the terms *open-choice principle* and *idiom principle* might seem to suggest a binary distinction, they are better viewed as identifying the opposite ends of a continuum. Sinclair himself said that "the preponderance of usage lies between the two."<sup>107</sup> And the fact that he was talking about a spectrum rather than a dichotomy is reflected in another pair of terms that he coined: *terminological tendency* and *phraseological tendency* (corresponding to the open-choice principle and the idiom principle, respectively).<sup>108</sup> In any event, the idea that the basic unit of meaning is not necessarily the individual word is broadly applicable, especially with the common words that make up the core vocabulary of English.

Along with substantially rejecting the view that words are the primary units of meaning, Sinclair rejected the idea that vocabulary and syntax are independent of one another. He argued instead that "[t]here [is] in practice no clear distinction between grammar and lexis" (*lexis* being the word corpus linguists use for *vocabulary*).<sup>109</sup> Sinclair described the corpus data as showing that grammar and lexis were intertwined in multiple ways. "In nearly every case," he said, "a structural pattern seemed to be associated with a sense. Despite the broad range of material in the corpus, when the instances were sorted into 'senses', a recurrent pattern emerged."<sup>110</sup> And in many cases, "including most of the common meanings of the common words," there existed an even closer relationship between "the sense and the phraseology."<sup>111</sup> The relationship that he describes here is the same relationship that supports the view that meaning often resides in units larger than the individual word, but with a focus on grammatical as well as lexical collocation: "[I]t was clear that in these central patterns of English the meaning was only created by choosing two or more words simultaneously *and disposing them according to fairly precise rules of position*."<sup>112</sup> Thus, the view that lexis and grammar are inseparable and the view that words are the basic units of meaning are merely opposite sides of the same coin.

---

107. *The Search for Units of Meaning*, *supra* note 96, at 29.

108. *Id.*

109. JOHN SINCLAIR, *Grammar in the Dictionary*, in LOOKING UP, *supra* note 46, at 110.

110. *Id.* at 109.

111. *Id.* at 110.

112. *Id.* (emphasis added).

## V. DO WORD MEANINGS EXIST?

*A. Doubters: Hanks, Atkins, and Kilgarriff*

The heading “Do Word Meanings Exist” is taken from the title of an article (and later a book chapter) by Patrick Hanks,<sup>113</sup> who was the managing editor of Cobuild’s first edition and was later Chief Editor of Current English Dictionaries at Oxford University Press, where he supervised the first edition of what was then titled the *New Oxford Dictionary of English*.<sup>114</sup> While it may seem odd for a professional lexicographer to ask whether word meanings exist, Hanks is not the only member of the profession to question their existence. Sue Atkins, who played “a formative role” in the Cobuild project’s design<sup>115</sup> and who more recently co-authored the “Oxford Guide to Practical Lexicography,” is quoted by Hanks as having said, “I don’t believe in word meanings.”<sup>116</sup> And to the list of nonbelievers one can add Adam Kilgarriff, a computational linguist whose doctoral dissertation investigated the nature of dictionary word senses, and who (before his death in 2015 at age fifty-five) did important work at the intersection of lexicography and computer science.<sup>117</sup> Hanks and Atkins are among

---

113. LEXICAL ANALYSIS, *supra* note 15, at 65–84; Patrick Hanks, *Do Word Meanings Exist?*, 34 COMPUTERS & HUMAN. 205 (2000), *reprinted in* PRACTICAL LEXICOGRAPHY: A READER, *supra* note 16, at 125–34 [hereinafter *Do Word Meanings Exist?*].

114. For Hanks’s background and publications, see his website, PATRICK HANKS, LEXICOGRAPHER, [www.patrickhanks.com](http://www.patrickhanks.com) (last visited Jan. 22, 2018). The New Oxford Dictionary of English (and its later editions, which dropped the “New” from the title) deals with modern-day British English, and should not be confused with the Oxford English Dictionary, which is a historical dictionary that traces the meanings of words back to their first recorded use in English, whenever that was. An “Americanized” version of the Oxford Dictionary of English has been published as the New Oxford American Dictionary. See *Corpus Evidence and Electronic Lexicography*, *supra* note 25, at 62.

115. John Sinclair, *Foreword* to COBUILD1, *supra* note 44, at v. For Atkins’s background and publications, see *Sue Atkins*, LEXICOM, [https://sites.google.com/a/lexmasterclass.com/www2/people\\_sa](https://sites.google.com/a/lexmasterclass.com/www2/people_sa) (last visited Jan. 29, 2018); *Sue Atkins: Publications*, LEXICOM, [https://sites.google.com/a/lexmasterclass.com/www2/people\\_sa\\_pub](https://sites.google.com/a/lexmasterclass.com/www2/people_sa_pub) (last visited Jan. 29, 2018).

116. *Do Word Meanings Exist?*, *supra* note 113, at 125. Or maybe what she said was “I don’t believe in word senses” See LEXICAL ANALYSIS, *supra* note 15, at 65; “*I Don’t Believe in Word Senses*,” *supra* note 16. Or maybe both, at different times.

117. For information on Kilgarriff’s background, see *Curriculum Vitae: Adam Kilgarriff*, ADAM KILGARRIFF, <https://www.kilgarriff.co.uk/cv.htm> (last visited Jan. 29, 2018). For a bibliographic essay describing his work, see *Adam Kilgarriff: Structured Bibliography*, SKETCH

the leading figures of modern lexicography, as was Kilgarriff before his untimely death.

Kilgarriff's inclusion among the word-meaning skeptics highlights the fact that the skepticism results not only from the use of corpora in lexicography but also from work in the field of natural language processing (NLP). NLP is an interdisciplinary field focusing on the processing of human language by computers.<sup>118</sup> One of the topics it deals with is word-sense disambiguation (WSD)—“the task of examining word tokens in context and determining which sense of each word is being used.”<sup>119</sup> This has required attention to be devoted to the nature of word meaning and lexical ambiguity. Whereas people can understand language effortlessly, without any conscious knowledge of semantics, the same is not true of computers. If we want a computer to “understand” natural language, the computer needs to be given an explicit model of meaning and ambiguity.<sup>120</sup> And developing such models requires an understanding of those phenomena as they operate in real life. Thus, lexicography has obvious relevance for work in WSD, and vice versa.<sup>121</sup>

### *B. Reasons for Doubt*

Having set the stage, let's return to the question that is asked in this section's title: do word meanings exist? If by “word meanings” you are talking about the kinds of word senses that are found in traditional dictionaries, Hanks, Atkins, and Kilgarriff, would say no.

ENGINE, <https://www.sketchengine.co.uk/adam-kilgarriff-structured-bibliography/> (last visited Jan. 29, 2018).

118. See generally, e.g., DANIEL JURAFSKY & JAMES H. MARTIN, *SPEECH AND LANGUAGE PROCESSING: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING, COMPUTATIONAL LINGUISTICS, AND SPEECH RECOGNITION* (2d ed. 2009).

119. *Id.* at 637.

120. YORICK A. WILKS ET AL., *ELECTRIC WORDS: DICTIONARIES, COMPUTERS, AND MEANINGS* 69 (1996).

121. See, e.g., Adam Kilgarriff, *Dictionary Word Sense Distinctions: An Enquiry Into Their Nature*, 26 *COMPUTERS & HUMAN* 365 (1992) [hereinafter *Dictionary Word Sense Distinctions*]; “*I Don't Believe in Word Senses*,” *supra* note 16, at 142–51; Adam Kilgarriff, *Word Senses*, in *WORD SENSE DISAMBIGUATION: ALGORITHMS AND APPLICATIONS* 29 (Eneko Agirre & Philip Edmonds eds., 2007) [hereinafter *Word Senses*]; Adam Kilgarriff, *Word Senses Are Not Bona Fide Objects: Implications for Cognitive Science, Formal Semantics, NLP*, in *PROCEEDINGS OF THE FIFTH CONFERENCE ON THE COGNITIVE SCIENCE OF NATURAL LANGUAGE PROCESSING* 193 (1996).



Hanks, Atkins, and Kilgarriff see that conception of word meaning as being at odds with the evidence.

For example, Atkins has described dictionary definitions of the kind we are familiar with as “trying to do what the language simply will not allow.”<sup>122</sup> She argues that word meaning “cannot be sliced up into distinct bundles, labelled (however carefully) and packaged into a dictionary entry which will tell the truth, the whole truth and nothing but the truth about that word.”<sup>123</sup> And similarly: “Faced now with the overwhelming richness and subtlety of the language in a computerized corpus, I no longer believe that it is possible to give a faithful, far less a true, account of the ‘meaning’ of a word within the constraints of the traditional entry structure.”<sup>124</sup>

At this point, a caveat is in order. I take Atkins to be referring mainly to high-frequency words, which tend to be the most polysemous, and not necessarily to *all* words. Standard dictionary definitions may be adequate for relatively unusual words (particularly technical words), which are less likely to have meanings that are dependent on context. Such words fall on the open-choice end of Sinclair’s distinction between the open-choice principle and the idiom principle.

However, as to words whose meanings *are* sensitive to context, it seems obvious in hindsight that the slice-and-dice approach to meaning won’t work. As to such words, there is a mismatch between, on the one hand, the presentation in dictionaries of word meaning as abstractions attributed entirely to the word being defined and, on the other hand, the way that meaning is actually made in texts or discourse. As we have seen, meaning often arises from multiword constructions, with the relevant sense of the word at issue being activated by its linguistic surroundings. To the extent that meaning is spread across multiple words, one should not expect good results from an approach to meaning that focuses on individual words.

---

122. B.T.S. Atkins, *Building a Lexicon: The Contribution of Lexicography*, 4 INT’L J. LEXICOGRAPHY 167, 180 (1991) [hereinafter *Building a Lexicon*].

123. *Id.*

124. B.T.S. ATKINS, *Theoretical Lexicography and Its Relation to Dictionary-Making*, 14 DICTIONARIES: J. DICTIONARY SOC’Y NORTH AM., 4, 20 (1993).

*1. Sense-division differences among dictionaries*

One symptom of the problem with word senses is that dictionaries differ in how they identify and carve up the different senses of a polysemous word. Atkins (working in some cases with prominent linguists) has compared how different dictionaries have treated various words, and has found that mismatches are not uncommon.<sup>125</sup> One dictionary may include a sense that another one does not; where one dictionary gives two or three senses, another may lump them together into a single sense that is broader but less specific; and the dividing lines between senses may be drawn in different places.

I will illustrate this phenomenon with one example, taken from a paper that Atkins wrote with Charles Fillmore (one of the most important linguists of the past fifty years). The paper constituted an exhaustive study of the semantics of *risk*, as both a verb and a noun, and it found extensive disagreement as to sense distinctions among the dictionaries they studied.<sup>126</sup>

The study began by examining how the use of *risk* as a verb was treated in three dictionaries. From that examination, Fillmore and Atkins concluded that the dictionaries “diverge[d] fundamentally in the selection of facts to present.”<sup>127</sup> The three dictionaries’ definitions, when combined, provided three senses:<sup>128</sup>

- [1] to expose to danger or loss; hazard: to *risk one's life*.
- [2] to do something in spite of the possibility of (unfortunate consequences): to *risk falling*; to *risk a fall*.
- [3] to incur the chance of unfortunate consequences by (doing something): to *risk climbing the cliff*.

The differences between these senses concerned the semantic role played by the entity that is denoted by the verb’s direct object. In

125. See B.T.S. Atkins, *Analyzing the Verbs of Seeing: A Frame Semantics Approach to Corpus Lexicography*, in PROCEEDINGS OF THE TWENTIETH ANNUAL MEETING OF THE BERKELEY LINGUISTICS SOCIETY: GENERAL SESSION DEDICATED TO THE CONTRIBUTIONS OF CHARLES J. FILLMORE 42 (1994); *Building a Lexicon*, *supra* note 122; Charles J. Fillmore & B.T.S. Atkins, *Starting Where the Dictionaries Stop: The Challenge of Corpus Lexicography*, in COMPUTATIONAL APPROACHES TO THE LEXICON 349 (B.T.S. Atkins & A. Zampolli eds., 1994) [hereinafter *Starting Where the Dictionaries Stop*]; Beth Levin, Grace Song & B.T.S. Atkins, *Making Sense of Corpus Data: A Case Study of Verbs of Sound*, 2 INT’L J. CORPUS LINGUISTICS 23, 29–30 (1997).

126. *Starting Where the Dictionaries Stop*, *supra* note 125, at 351–55.

127. *Id.* at 352.

128. *Id.* The three senses set out in the text are adapted from this source but are not direct quotations.

sense 1, that direct object denotes the thing that is put in danger; in sense 2, it denotes the bad thing that might happen; and in sense 3, it denotes it denotes the conduct that creates the risk.

All three of the dictionaries included sense 1. However, two included sense 2 but not sense 3, while the third included sense 3 but not sense 2. When Fillmore and Atkins looked at more dictionaries, they found several that followed the 2-but-not-3 pattern, plus some that followed still different patterns. Two dictionaries included all three senses, and three included sense 1 plus a single sense that blurred together senses 2 and 3.

For uses of *risk* as a noun, the situation was even worse. Considering all the dictionaries together, Fillmore and Atkins concluded that “lexicographers who had all the facts at their disposal might have distinguished at the most five dictionary senses.”<sup>129</sup> And each of the ten dictionaries handled the word differently, with each one including and omitting different combinations of the five possible senses, and in some cases lumping two or three of the senses together.

Disparities such as these are problematic for the Dictionary Paradigm. Such differences should not exist if, as many lawyers and judges seem to believe, the word senses given in dictionaries are abstract entities that have some kind of independent existence. That the differences between dictionaries exists is evidence that distinguishing between word senses often depends at least in part on the exercise of judgment and discretion by the lexicographer. That process is unavoidably subjective; there are no objective standards for distinguishing between senses.<sup>130</sup> In the *Oxford Guide to Practical Lexicography*, Atkins and her coauthor say that there is “nothing ‘definitive’” about how lexicographers divide words into separate senses.<sup>131</sup> They note that James Murray, editor-in-chief of the *Oxford English Dictionary*, said “that the best any lexicographer could hope for would be that

---

129. *Id.* at 354.

130. *E.g.*, THE STORY OF WEBSTER’S THIRD, *supra* note 56, at 81; *The Impact of Corpora on Dictionaries*, *supra* note 70, at 224; *Johnson and Modern Lexicography*, *supra* note 4, at 258; “I Don’t Believe in Word Senses,” *supra* note 16, at 143.

131. THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 275.

readers would feel, on scanning a multisense dictionary entry, that ‘this is not an unreasonable way of exhibiting the facts.’”<sup>132</sup>

## 2. *Some further problems regarding sense division*

Moving beyond the mere fact that dictionaries differ in how they divide up word meaning, the divisions themselves can raise issues. Let’s consider the word *bank* (in its financial sense), which most dictionaries divide into two subsenses: one for a bank as an institution, establishment, business, or organization (the INSTITUTION sense) and another for a bank as a building or other location where the institution carries on its banking business (the BUILDING sense). Of the dictionaries I looked at, only the Merriam-Webster dictionaries (the *Unabridged* and the *Collegiate*) did not include the BUILDING sense.

The fact that all the other dictionaries provided two senses creates the impression that each sense is separate from the other. But are they really? Merriam-Webster didn’t think both senses were necessary. So maybe the conception of a bank as a building can be inferred from the conception of a bank as an institution, with the inference being activated when the word is used in an appropriate context. But if so, that suggests that what are presented as separate senses are not necessarily independent of each other. And that in turn casts doubt on the idea that an ambiguous statutory term can be interpreted simply by choosing between different dictionary senses (as the Supreme Court did in *Muscarello*<sup>133</sup>). On the other hand, if one assumes the BUILDING sense is independent of the ESTABLISHMENT sense, that would seemingly open the door to dividing up the latter sense further: BANK-AS-EMPLOYER (*the bank laid off 100 employees*), BANK-AS-CORPORATION (*the bank has a new board of directors*), BANK-AS-INSURED (*the bank is covered by a \$50 million insurance policy*), BANK-AS-SPORTS-TEAM (*the bank made it to the finals of the summer softball tournament*).

If lexicographers had to choose one of these two approaches in defining *bank*, I suspect that most would lump everything into a single broad sense. But either way, the decision would be based on *practical* considerations, not on which approach was somehow considered a more accurate representation of reality.

---

132. *Id.* (crediting the Murray quotation to *The Analysis of Meaning*, *supra* note 27, at 86).

133. *Muscarello v. United States*, 524 U.S. 125, 127–32 (1998).

Staying with the example of *bank*, we can see that a use of the word may simultaneously activate more than one sense. Consider the sentence, *I have a meeting at the bank*. This pretty clearly activates the BUILDING sense, but not necessarily *only* that sense. An utterance of the sentence would probably be understood to communicate not only that the speaker had a meeting that was to be held on the bank's premises, but also that the meeting would be with representatives of the bank and would concern the bank's business, thereby activating the INSTITUTION sense. So, the different dictionary senses of a word are not always mutually exclusive.<sup>134</sup>

A related issue concerns the fact that different senses of a word are not always divided by clear lines; sometimes they blend into one another. Penelope Stock, who was one of the Cobuild lexicographers, discussed this problem in a paper that was published while the dictionary was being compiled; she spoke of a semantic "blurring" that occurs "where a word seems either to operate on a cline between two or more meanings, or to bring in its train various extra nuances so that any individual utterance might suggest one strong aspect of a meaning but is, as it were, strengthened or supported by various other possible close meanings."<sup>135</sup> To illustrate this problem, Stock pointed to the word *culture* and provided a number of corpus examples, from which the following list is drawn.<sup>136</sup>

There does seem evidence [sic] that Eastern cultures have more right  
brain emphasis.  
the big colloquium on African culture and African civilization that's to be held  
the great cultures of Japan and China  
Infanticide was practised by many early cultures  
desire to live as a nation that has its own culture and individuality  
a multicultural society where cultures can live side by side  
to give value and literate dress to an oral culture we have forgotten how  
to appreciate.  
the Ministry of Culture  
by removing all traces of black ethics and culture.  
the culture of machismo  
Man dresses the part his culture tells him he is called upon to play.

134. See *Dictionary Word Sense Distinctions*, *supra* note 121, at 376–77.

135. Penelope F. Stock, *Polysemy*, in LEXETER '83 PROCEEDINGS: PAPERS FROM THE INTERNATIONAL CONFERENCE ON LEXICOGRAPHY AT EXETER 131, 137 (R.R.K. Hartmann ed., 1983), reprinted in PRACTICAL LEXICOGRAPHY: A READER, *supra* note 16, at 158.

136. *Id.* at 158–59.

nevertheless absorbed enough of Spanish political culture to build  
 authoritarian principles  
 Newspaper-reading, word-and-trade-conscious urban culture.  
 I have shown how Caro's work belongs to the culture of the early 1960's  
 culture shock

has led to the development of a specific 'pop' culture  
 the extension of the throw-away culture

I have tried to arrange the order of these examples so that as you read down the list, *culture* displays a changing continuum of meanings. In the first example, *culture* means  $X_1$ , then in the second example it means  $X_2$ , which is similar to but slightly different from  $X_1$ , and so on down the list, with each item differing slightly from the one before it. I have tried to order the examples so that in each line *culture* is used more narrowly than in the one before it (in an intuitive sense of "narrow" that I would be hard put to explain), so that the difference from one line to the next is small. And while there seems to be a clear difference in meaning between the first example and the last, it is difficult if not impossible to find the boundary between the two meanings, or to say whether there are intermediate meanings that might count as a third sense. There are quite possibly several ways of grouping the examples, none of which would be *the* right answer.

### C. "Meaning Is Something You Do."

After much discussion of what word meanings are *not*, it is time to talk about describing what they *are*. Or rather, how to describe word meanings within the framework of (Sinclairian) corpus linguistics.

I'm going to start with the statement by Kilgarriff that "meaning is something you do"<sup>137</sup>—a view endorsed by Hanks and Atkins. Hanks says that "[i]n the everyday use of language, meanings are events, not entities."<sup>138</sup> Similarly, in the *Oxford Guide to Practical Lexicography*, Atkins and her coauthor say, "[M]eanings and dictionary senses aren't the same thing at all. Meanings exist in infinite numbers of discrete communicative events, while the senses in a dictionary rep-

---

137. *Word Senses*, *supra* note 121, at 32.

138. *Do Word Meanings Exist?*, *supra* note 113, at 130.

resent lexicographers' attempts to impose some order on this babel."<sup>139</sup> To be sure, it is "convenient shorthand," as Hanks puts it, to talk about the meanings of words in isolation, as they are set out in dictionaries, "but strictly speaking these are not meanings."<sup>140</sup> Rather, they are what Hanks calls "meaning potentials"—potential contributions to the meanings of texts and conversations in which the words are used, and activated by the speaker who uses them."<sup>141</sup>

Kilgarriff describes dictionary senses, as distinguished from meanings, as representations of a "corpus cluster"—a grouping of corpus lines in which the word in question is used in more or less the same way.<sup>142</sup> Such a cluster is prepared using a KWIC concordance for the word in question. The concordance lines are grouped "so that, as far as possible, all members of each cluster have much in common with each other, and little in common with members of other clusters."<sup>143</sup> The lexicographer then "works out what it is that makes [each cluster's] members belong together, reorganising clusters as necessary," and writes up those conclusions "in the highly constrained language of a dictionary definition."<sup>144</sup>

The clustering process is an intuitive one. "The lexicographer may or may not be explicitly aware of the criteria according to which he or she is clustering"; the process of working out what the members of each cluster have in common "is just a fallible *post hoc* attempt to make the criteria explicit."<sup>145</sup> And although Kilgarriff describes this process in the context of using corpus data, it seems to me that his description applies equally to what lexicographers do when they work from old-fashioned citations.<sup>146</sup>

---

139. THE OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 311.

140. *Do Word Meanings Exist?*, *supra* note 113, at 130.

141. *Id.* The concept of meaning potentials is also invoked, though in terms that might be slightly different, in Jens Allwood, *Meaning Potentials and Context: Some Consequences for the Analysis of Variation in Meaning*, in COGNITIVE APPROACHES TO LEXICAL SEMANTICS 29 (Hubert Cuyckens et al. eds., 2003).

142. "I Don't Believe in Word Senses," *supra* note 16, at 144–45.

143. *Id.* at 145.

144. *Id.*

145. *Id.*

146. Indeed, at least one lexicographer noted during the pre-corpus era that "citations tend to fall into what may be called 'contextual clusters.'" Allen Walker Read, *The Relation of Definitions to their Contextual Basis*, 39 ETC: REV. GEN. SEMANTICS 318, 319 (1982).

All of this points toward the conclusion that rather than think of words as *having* meanings, we should think about words as things that are used in order to *make* meanings.<sup>147</sup> And it follows that we should make a corresponding adjustment in how we think about dictionary definitions and in how we deal with disputes about word meaning. The latter topic is one that I will take up in the next, and final, part of this Article.

VI. *MUSCARELLO V. UNITED STATES*:  
A CORPUS ANALYSIS OF *CARRY A FIREARM*

A. *The Case*

*Muscarello v. United States*<sup>148</sup> is famous (or should I say infamous?) for two things. One is the Supreme Court's exercise in kinda-sorta corpus linguistics. The issue was whether the phrase *carry a firearm* applies to driving somewhere with a gun locked in your glove compartment or trunk; in holding that it does, the Court relied on evidence of actual usage showing that *carry* is sometimes used that way.<sup>149</sup> Stephen Mouritsen has persuasively criticized the Court's analysis on the ground that the Court's choice of search terms effectively predetermined what the data would show.<sup>150</sup> He also used corpus data to show that the Court got the ordinary meaning of *carry* wrong, at least if "ordinary meaning" is taken to mean "most-frequently-expressed meaning."<sup>151</sup>

The other point for which *Muscarello* is known is the Court's display of poor dictionary skills: the majority pointed to the first senses listed for *carry* in the *Oxford English Dictionary* and *Webster's Third New International Dictionary* as the word's primary sense, rather than

---

147. Patrick Hanks, *How People Use Words to Make Meanings: Semantic Types Meet Valencies*, in *NLPCS 2010: PROCEEDINGS OF THE 7TH INTERNATIONAL WORKSHOP ON NATURAL LANGUAGE PROCESSING AND COGNITIVE SCIENCE 3* (Bernadette Sharp & Michael Zock eds., 2010).

148. *Muscarello v. United States*, 524 U.S. 125 (1998).

149. *Id.* at 129–30.

150. *The Dictionary is Not a Fortress*, *supra* note 2, at 1947–48.

151. *Id.* at 1957–66.



(as is actually the case) its *oldest* sense.<sup>152</sup> Mouritsen skewered the Court on this point, too.<sup>153</sup>

My purpose here in dealing with *Muscarello* is to use *carry* as a vehicle for a purely corpus-based analysis of word meaning, drawing on a methodology developed by Hanks called Corpus Pattern Analysis. This analysis will differ from Mouritsen's in that his purpose was to decide which of the two dictionary senses that were at issue in *Muscarello* was the more common one. In doing so, he did not, as far as I can tell, use the approach that I will follow here. On the other hand, my purpose here is more open ended. Rather than adjudicating between two different word senses, I want to provide an example of corpus analysis in action, in order to show how it works and what it can do.

First, some necessary background. The defendants in *Muscarello* had been convicted under a statute that mandates the imposition of at least five years' imprisonment on anyone who "during and in relation to any crime of . . . drug trafficking . . . uses or carries a firearm."<sup>154</sup> Each of the defendants had driven to the site of a drug deal with a gun in his vehicle; in one case it was in the glove compartment and in the other it was in the trunk.<sup>155</sup> They each challenged their conviction on the theory that their actions had not constituted "carrying" a firearm. By a 5–4 vote, the Court rejected that argument.<sup>156</sup>

In discussing what the ordinary meaning of *carry* is, the Court saw its job as being to select between two dictionary senses of the word: (1) what the Court believed was "the first, or primary, meaning," under which "one can, as a matter of ordinary English, 'carry firearms' in a wagon, car, truck, or other vehicle that one accompanies," and (2) what the Court described as "a different, rather special" use of the word, under which it means "for example, 'bearing' or (in slang) 'packing' (as in 'packing a gun')." <sup>157</sup>

---

152. *Muscarello*, 524 U.S. at 128; see 1 THE OXFORD ENGLISH DICTIONARY xxix (2d ed. 1989); WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 17a (Philip Babcock Gove ed., 1971).

153. *The Dictionary is Not a Fortress*, *supra* note 2, at 1930–35.

154. 18 U.S.C. § 924(c) (2012).

155. *Muscarello*, 524 U.S. at 127.

156. *Id.*

157. *Id.*

As a check on its reliance on the dictionaries, the Court investigated the usage of *carry* in news reports “to make certain that there is no special ordinary English restriction (unmentioned in dictionaries) upon the use of ‘carry’ in respect to guns.” In particular, the Court searched the *New York Times* database in Nexis and the “US News” database in Westlaw, looking for “sentences in which the words ‘carry,’ ‘vehicle,’ and ‘weapon’ (or variations thereof) all appear.” Thousands of such sentences were found, many of which were “used to convey the meaning at issue here, *i.e.*, the carrying of guns in a car.”<sup>158</sup>

The Court also acknowledged that one of the “primary” definitions it relied on “define[d] ‘carry’ as ‘to move while supporting,’ not just in a vehicle, but also ‘in one’s hands or arms.” But it said that although “one who bears arms on his person ‘carries a weapon, . . . one may *also* ‘carry a weapon’ tied to the saddle of a horse or placed in a bag in a car.”<sup>159</sup>

### *B. Corpus Pattern Analysis*

My analysis of the corpus data regarding *carry* will draw on a method developed by Hanks, which he calls Corpus Pattern Analysis (CPA).

Hanks describes CPA as “a technique for mapping meaning onto words in text.”<sup>160</sup> Its purpose is to identify and systematically describe the different patterns in which words appear, and the meanings associated with those patterns, as shown by corpus data.<sup>161</sup> CPA does not attempt to identify word meanings in isolation; rather, “meanings are associated with prototypical contexts.”<sup>162</sup>

The process has two steps. The first is essentially a variant on the corpus-clustering process that is discussed above. For each word being analyzed (referred to here as “the key word”), “[c]oncurrence lines

---

158. *Id.* at 129.

159. *Id.* at 130.

160. Patrick Hanks, *Corpus Pattern Analysis*, in 1 PROCEEDINGS OF THE ELEVENTH EURALEX INTERNATIONAL CONGRESS 87, 87 (Geoffrey Williams & Sandra Vessier eds., 2004) [hereinafter *Corpus Pattern Analysis*]; see also, e.g., LEXICAL ANALYSIS, *supra* note 15, at 113–43.

161. See *Corpus Pattern Analysis*, *supra* note 160, at 87–88. The symbol  $\approx$  is used to indicate that “grammatical” is a rough paraphrase of “syntagmatic,” not an exact synonym.

162. *Id.* at 88.

are grouped into semantically motivated syntagmatic [ $\approx$ grammatical] patterns.”<sup>163</sup> (I take the description of the patterns as being “semantically motivated” to mean that each cluster is limited to lines in which the pattern is used to express essentially the same meaning. Any such semantic sorting presumably takes account of similarities and differences in the key word’s collocates.)

In the second step, “a ‘meaning’ [is associated] with each pattern . . . in close coordination with the assignment of concordance lines to patterns.”<sup>164</sup> In dealing with the key word’s collocates, the process requires more than just making a list of words. Rather, it involves grouping the collocates into categories, with the categories being organized into a taxonomy, of which the following is a toy example:

```

ANYTHING
- ENTITY
- - ANIMATE
- - - HUMAN
- - - - LITIGANT
- - - - LAWYER
- - - NONHUMAN
- - - - ...
- - INANIMATE
- - INSTITUTION
- - - - ...
- - OTHER
- - - - ...
- EVENTUALITY
- - EVENT
- - STATE

```

Under the last level of each branch of the taxonomy, when the most specific category is reached, would be the set of words that falls into that category. And for each pattern, there would typically be preferences for the categories of collocates that would fill a particular grammatical slot. For example, for a given pattern, there might be a preference for the subject slot to be filled with a noun from the category HUMAN.

---

163. *Id.*

164. *Id.*

To illustrate what the end result of this process looks like, set out below are five patterns associated with the verb *need*.<sup>165</sup> The example is taken from the *Pattern Dictionary of English Verbs* (PDEV), a project headed by Hanks, the purpose of which is to list the patterns associated with each verb in a large corpus (currently, the 100-million-word British National Corpus).<sup>166</sup> These are all the patterns that were identified for *need*. Semantic categories are shown in SMALL CAPITAL LETTERS; types of grammatical constructions are shown in brackets. Other than formatting changes, the only difference from how the patterns appear on the PDEV website is that in the column containing the line labels, I have substituted *meaning* for *implicature*, out of consideration for readers unfamiliar with the latter word.

Pattern:	HUMAN OR INSTITUTION needs EVENTUALITY OR ENTITY
Meaning:	HUMAN OR INSTITUTION requires that EVENTUALITY OR ENTITY must be realized or available, in order to accomplish some goal
Example:	<i>Even the largest schools may need outside help.</i>
Pattern:	ENTITY 1 OR EVENTUALITY 1 needs ENTITY 2 OR EVENTUALITY 2
Meaning:	ENTITY 1 OR EVENTUALITY 1 is an essential precondition for or attribute of ENTITY 2 OR EVENTUALITY 2
Example:	<i>Punishing people certainly needs a justification, since it is almost always something which is harmful, painful or unpleasant to the recipient.</i>
Pattern:	ENTITY OR EVENTUALITY needs [to-infinitive]
Meaning:	An essential precondition for the realization of EVENTUALITY is that [verb] must be realized
Example:	<i>Moreover, labels need not be permanent and irreversible.</i>
Pattern:	HUMAN needs [to-infinitive]
Meaning:	HUMAN must do [verb]
Example:	<i>I need to explain why this was happening in the first place.</i>
Pattern:	PLANT OR ANIMATE needs EVENTUALITY OR STUFF
Meaning:	PLANT OR ANIMATE must have EVENTUALITY OR STUFF in order to survive and flourish

165. Patrick Hanks et al., *Need*, PATTERN DICTIONARY ENG. VERBS, <http://pdev.org.uk/#browse?q=need;f=A;v=need> (last visited Jan. 29, 2018).

166. Patrick Hanks et al., PATTERN DICTIONARY ENG. VERBS, <http://pdev.org.uk/#browse> (last visited Jan. 22, 2018). Out of the 5396 verbs in the corpus, *need* is one of the 1393 that had been completed as of mid-October 2017; analysis of *carry* had not yet begun. *Id.*

Example: *I had to have a tracheostomy operation, and from then on I needed twenty-four-hour nursing care.*

CPA's value for both lexicography and legal interpretation lies in bringing discipline and structure to the process of investigating meaning. It forces one to focus on the two aspects of context that are most important: the specific grammatical structure that is at issue, and the semantic characteristics of the key collocates of the word in question. In discussing *Muscarello*, the grammatical structure is the transitive-verb construction: verb plus noun phrase. The important collocates are the words that serve as the verb's subject (denoting the entity doing the carrying) and as its direct object (denoting what is being carried). But a different case might involve some other grammatical construction, with its own set of semantic characteristics.<sup>167</sup> The important point is that CPA provides a framework for organizing the corpus analysis.

### *C. Analyzing the Corpus Data*

#### *1. Framing the inquiry*

The verb *carry* is often used in ways that are irrelevant to the statute at issue in *Muscarello*. People can carry diseases, as can birds and other animals. Products carry warnings, television networks carry programs and ads, crimes carry penalties, fiber optic cables carry data. So it was necessary to identify the characteristics that distinguish the kind of carrying prohibited by § 924(c) from those other kinds.

This is where my analysis was informed by CPA. In particular, I used the idea of semantically categorizing the collocates of the verb. Deciding on the appropriate categories was easy. The statute prohibits actions by humans, and firearms are tangible objects, so the relevant corpus lines would be those that denoted the carrying of tangible objects by humans. Thus, we have the initial categories HUMAN and OBJECT, and the basic conceptual schema HUMAN CARRY OBJECT.

Within the set of corpus lines instantiating that schema, I was interested in seeing what was indicated by each line with respect to the

---

167. See, e.g., Neal Goldfarb, *The Semantics of Sleeping in Railway Stations*, *LAWNLINGUISTICS* (June 5, 2017), <https://lawlinguistics.com/2017/06/05/the-semantics-of-sleeping-in-railway-stations/>.

manner in which the object was carried: in the person's hands or arms, strapped to their back, in the trunk of their car, and so on. Therefore, I attempted to categorize the manner of carrying that was reflected in each concordance line that denoted the carrying of a tangible object by a human. I set up two categories, each one corresponding to a separate subschema: HUMAN CARRY OBJECT IN VEHICLE, which encompassed events analogous to those in *Muscarello* and HUMAN CARRY OBJECT, which covered everything else.

There were two ways in which the concordance line might provide the information needed for this categorization. The first is that the information might be linguistically encoded:

carried a hacksaw blade in his toolbox  
 he was carrying it behind the seat of his pick-up.

Alternatively, the necessary information may be inferred, with varying degrees of specificity, from the context:

Lindsay was going up the stairs, carrying a ceramic teacup  
 She drives her car and carries her dogs places and goes to openings of clubs.  
 . . . seem underfoot, a noise clicked. Wood on rock, like a rifle butt carried  
 too low.

There was also a separate issue that I wanted to explore. *Carry* is sometimes used in sentences in which the subject noun denotes a vehicle:

There is a distinct possibility that the truck was carrying a back-up tactical rocket  
 still trying to determine if a second man helped rent a truck believed to have carried the bomb.  
 They're walking perhaps a hundred yards or so to a car waiting to carry them into Cape Town.

One might argue that uses like these support the Court's holding in *Muscarello*, since the schema VEHICLE CARRY is similar to CARRY IN A VEHICLE. But on the other hand, such uses are impersonal—the notion of a person doing the carrying is not expressly encoded in the text and is not otherwise brought within the reader's focus of attention. Given that I didn't know in advance whether this kind of use would be relevant, I kept track of it separately, using a category corresponding to the schema VEHICLE CARRY OBJECT.

## 2. *The data*

The data that I analyzed came from the Corpus of Contemporary American English (COCA).<sup>168</sup> The corpus contains just over 100,000 separate instances of the lemma *carry* being used as a verb. *Lemma* is the word used in lexicography and corpus linguistics to refer collectively to all the forms of a given word. Therefore, the 100,000+ instances of the lemma *carry* (also referred to as “tokens”) include tokens of the word forms *carry*, *carries*, *carried*, and *carrying*. I should also add that some of the tokens of *carry* and *carries* are in fact uses of the word as a noun that are incorrectly tagged in the corpus as verbs.

I obviously looked at only a small fraction of all the tokens of *carry*. I looked individually at 901 concordance lines, which I believe is enough data to provide a reliable picture of how *carry* behaves.<sup>169</sup>

These concordance lines were gathered by means of several different searches:

- 400 lines resulted from two lemmatized KWIC searches on *carry*, each returning 200 lines. (There was no overlap between the two sets of results.) In order to provide a look at the behavior of *carry* in general, these searches did not require the presence of any particular collocates.
- An additional 299 lines resulted from a series of collocate searches calling for corpus lines in which various vehicle-related search terms occurred within four words after the lemma *carry*: *car* (100 lines); *truck* (99); *seat*, *backseat*, or *frontseat* (48); *trunk* (41); and *glovebox*, *glove*, or *compartment* (11). Given that the collocates in these searches either denote a type of vehicle (*car*, *truck*) or denote things that are to varying degrees associated with vehicles, it might be expected that corpus lines that contained any of these collocations would have a

---

168. COCA, *supra* note 28, <http://corpus.byu.edu/coca/> (last visited Jan. 22, 2018).

169. In the compilation of the *Pattern Dictionary of English Verbs*, 250 corpus lines are typically analyzed for each verb. See *DVC: Disambiguation of Verbs by Collocation*, RES. GROUP COMPUTATIONAL LINGUISTICS, <http://rgcl.wlv.ac.uk/research/dvc-disambiguation-of-verbs-by-collocation/> (last visited Jan. 29, 2018). In addition, the *Oxford Guide to Practical Lexicography* states that lexical-profiling software, which partially automates the corpus-analysis process, “only works well for lemmas with at least 500 hits[.]” OXFORD GUIDE TO PRACTICAL LEXICOGRAPHY, *supra* note 12, at 61. Therefore, 902 corpus lines seems adequate.

higher-than-random probability of falling within the category HUMAN CARRY OBJECT IN VEHICLE. As we will see, that is exactly what the data showed.

- I ran several collocate searches calling for lines in which *carry* collocated (within a four-word span in either direction) with *gun, pistol, firearm, rifle, or shotgun*. I did two searches limited to 100 lines each, one in which the collocate terms were singular and the other in which they were plural.

The data from all these searches can be found in the appendix to this Article, which is available online.<sup>170</sup>

### 3. Analysis

*a. Summary.* The most important conclusion is that corpus lines categorized as HUMAN CARRY OBJECT IN VEHICLE—i.e., those corresponding to the defendants' conduct in *Muscarello*—were greatly outnumbered by those categorized as simply HUMAN CARRY OBJECT. Moreover, my reading of the corpus lines in the latter category is that all of them involve people carrying things in their arms, in their hands, in bags and suitcases, slung over their shoulders, and so on. And that was clear even though in the great majority of cases, the manner of carrying was not expressly encoded. In those cases, the manner of carrying had to be inferred from the nature of the event and surrounding circumstances as described in the text. All of which gives reason to believe that phrases following the pattern [*human*] *carry* [*object*] are not used to express the meaning HUMAN CARRY OBJECT IN VEHICLE unless the IN VEHICLE part is explicitly encoded or otherwise supported by something in the context.

That said, when you look at a concordance limited to uses in which *carry* is collocated with words denoting firearms, some uncertainty creeps in. It's not that there are more uses falling into the category of HUMAN CARRY OBJECT IN VEHICLE; the proportion of such uses is comparable to the proportion of such uses in general. However, within the subcategory HUMAN CARRY FIREARM, there are a significant number of uses that feel to me like they are more amenable to an interpretation that would support the holding in *Muscarello*.

---

170. Neal Goldfarb, *A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics (Appendix)* (2017), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3039015](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039015).



*b. The use of carry in general.* Because of the special issues that arise with respect to uses in which *carry* is collocated with words like *gun*, *pistol*, and *rifle*, I will focus first only on the corpus lines that resulted from the two KWIC searches and from the searches involving vehicle-related collocates. Those searches yielded a total of 699 corpus lines, which broke down as follows:

	Human carry object	Human carry object in vehicle	Vehicle carry object	Other	Total
KWIC (note 1)	124	2	24	250	400
car	53	5	19	23	100
truck (note 2)	26	8	52	13	99
seat, backseat, frontseat	27	7	3	11	48
trunk	20	11	3	7	41
glovebox, glove, compartment (note 2)	2	5		4	11
Total	252	38	101	308	699

## Notes:

1. The results from the KWIC search included four lines in which *carry* was used, without a direct object, to denote carrying a gun (e.g., “cabdrivers who are tested and trained to carry and protect themselves”). These are included under “other,” as are six lines that were unclear or ambiguous.

2. The results on these rows for HUMAN CARRY OBJECT IN VEHICLE include a few corpus lines that were ambiguous between HUMAN CARRY OBJECT IN VEHICLE and VEHICLE CARRY OBJECT. I counted them toward the former reading so as to give the *Muscarello* majority the benefit of the doubt.

3. There were a few corpus lines in which I interpreted *carry* as being used as being generally synonymous with *accompany* or *escort*, a usage peculiar to American Southern dialect. These are included under “Other.”

Overall, the HUMAN CARRY OBJECT IN VEHICLE pattern accounted for 15.4% of the corpus lines that involved humans carrying tangible objects, either in a vehicle or otherwise, and 11.4% of the all of corpus lines that referred to the carrying of a tangible object (whether by a human in any manner or in a vehicle). Thus, the HUMAN CARRY OBJECT IN VEHICLE pattern represented only a small minority of the relevant corpus lines. However, these figures almost certainly overstate the actual frequency of the HUMAN CARRY OBJECT IN VEHICLE pattern, since they include the results of the searches that included vehicle-related collocates. As I said above, that category is likely to include a higher-than-average proportion of lines fitting the pattern. And that prediction is borne out by the data. Looking only at the vehicle-related collocate searches, 17.7% of all lines referring to the carrying of a tangible object fit the HUMAN CARRY OBJECT IN VEHICLE pattern, compared to only 1.4% for the KWIC searches by themselves. The KWIC results are presumably more representative of *carry*'s overall behavior. In contrast to that low percentage, 79.5% of the KWIC results that involved the carrying of a tangible object fit the HUMAN CARRY OBJECT pattern. So if the KWIC data is taken as being

representative, the frequency of that pattern is roughly fifty-seven times that of the HUMAN CARRY OBJECT IN VEHICLE pattern.

With respect to the both the HUMAN CARRY OBJECT and the HUMAN CARRY OBJECT IN VEHICLE categories, I looked to see how often the manner of carrying was explicitly encoded. Each category's results were virtually the mirror-image of the other's. Within the HUMAN CARRY OBJECT category, the great majority of lines did not encode the manner in which the object was carried, leaving that fact to be inferred. The opposite was true of the category HUMAN CARRY OBJECT IN VEHICLE; most of the uses in that category did explicitly encode the fact that the object was being carried in a vehicle, as in the following examples:

He dug out the flashlight he carried in the glove box and clicked it on.  
... no reason to keep carrying stuff in our trunk  
Two were charged with stripping parts from a parked car and three with carrying a sawed-off shotgun and drugs in a car, police said.  
In the front seat, I'm carrying a gadget slightly larger than an electric shaver.

In all but one of the remaining lines, the manner-of-carrying information was partially encoded, by which I mean that there were explicit references to vehicles (underlined in the examples below), which prompted inferences that the object was carried in that vehicle:

Shagren said what sparked the proclamation was concern over truck drivers carrying dairy products not being able to drive more than 12 hours a day  
...  
He was thrown into the back of a deuce-and-a-half truck and carried south in a pile of wounded and dying men . . .  
We returned with the truck to carry loot for the soldiers, mainly furniture.

Only one corpus line instantiating the HUMAN CARRY OBJECT IN VEHICLE schema did not explicitly refer to a vehicle, but even that line can be understood as referring to vehicles by implication:

and the other delivery men at Yasgur's were planning to carry milk to the hordes, while my mother would help the Ladies

What all of this shows, I think, is not only that the pattern [*human*] *carry* [*object*] is seldom used to refer to events in which the object is carried in a vehicle, but also that the pattern is not used to express that meaning without the text or utterance containing some kind of overt indication to that effect.

I haven't yet discussed the results for corpus lines categorized as VEHICLE CARRY OBJECT—i.e., those in which the subject noun denoted the vehicle in which the object was carried. This pattern was less frequent than the HUMAN CARRY OBJECT pattern. Overall, the number of VEHICLE CARRY uses was 26.4% of the combined HUMAN CARRY uses. But that figure includes the results from the collocate searches, which would be expected to overrepresent VEHICLE CARRY uses in comparison to the entire corpus. Looking only at the KWIC results, for which no collocates were specified, the number of VEHICLE CARRY uses was roughly 19.2% of the number of HUMAN CARRY uses.

The VEHICLE CARRY category also included more corpus lines than did HUMAN CARRY OBJECT IN VEHICLE. The size of the former was 237% of the size of the latter overall (including the results from the collocate searches), and was 1500% the size of the former when only the KWIC searches are considered.

Even when both vehicle-related categories are combined, the total number of corpus lines was smaller than the number for the HUMAN CARRY OBJECT category. The combined size of the vehicle-related categories was 37.8% of the size of the HUMAN CARRY CATEGORY overall, and 20.5% when considering only the KWIC searches.

Thus, the data supports the following generalizations. First, *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. That may or may not suggest that the former use is more "basic" or "primary" than the latter, but it certainly does not suggest the opposite.

Second, when people do talk about objects being carried in vehicles, they are significantly more likely to do so using the impersonal pattern [*vehicle*] *carry* [*object*] than to use the pattern [*human*] *carry* [*object*] *in* [*vehicle*]. This difference in frequency might suggest that the two patterns differ in their meanings to at least some extent, because if their meanings were the same, one might expect their relative frequencies to be more similar. Before reaching a more definite conclusion, I would want to do some research into the relationship between frequency and similarity of meaning. But my initial inclination

would be to separate these two categories into separate corpus clusters and therefore to treat them as separate senses.<sup>171</sup>

In short, the data that has been examined so far weighs against the majority's interpretation in *Muscarello*.

*c. Carry collocated with words denoting firearms.* I turn now to the results of my searches for corpus lines in which *carry* is collocated with what I will refer to as "firearm collocates."<sup>172</sup> As with the results that are discussed above, this category included only a handful of lines referred to firearms being carried in vehicles. These represented about 2.4% of the lines that were relevant to the analysis (meaning the lines instantiating the pattern [*human*] *carry* [*firearm*]). However, there were also some notable differences, some of which might provide a nonfrivolous basis for arguing in support of the result in *Muscarello*.

The first difference is one that affected how I structured my analysis, but turned out not to affect my ultimate conclusions. In roughly 15% of the relevant firearm-collocate corpus lines, *carry* was used as part of a statement about the permissibility or legality of carrying firearms. These uses, which I labeled as DEONTIC,<sup>173</sup> included the following:

Representative Charles Collins, who sponsored legislation in Little Rock that allows gun owners to carry their weapons to church . . . .  
Wisconsin has become one of the most progressive states now for owning a gun and carrying concealed weapons. . . .

---

171. That is how these patterns are treated in Cobuild, but with the HUMAN CARRY OBJECT pattern being split into two senses. Here are the relevant entries from the first edition of the Cobuild dictionary:

1 If you **carry** something, you hold it or support it so that it does not touch the ground, and take it with you as you go somewhere.

. . .

3 When a vehicle **carries** people, they travel inside it from one place to another.

4 If you **carry** something with you, you have it with you wherever you go, for example by keeping it in your pocket or in your handbag.

COBUILD1, *supra* note 44, at 208 (examples omitted; boldface in original).

172. I should note that as a result of additional analysis that I undertook in September and October 2017, the discussion here and in that appendix goes beyond what was in the draft of this Article that was posted on SSRN in February 2017 and the revised version that was posted in September of the same year. In addition, some of the coding has been changed. However, the underlying data is the same.

173. *See, e.g.*, Monica Bucciarellia & P.N. Johnson-Laird, *Naïve Deontics: A Theory of Meaning, Representation, and Reasoning*, 50 COGNITIVE PSYCH. 159 (2005).

killed a bill that would have allowed Tennesseans to openly carry guns without a permit and several other gun-related proposals . . . .

Uses such as these arguably should not be considered as part of an inquiry into ordinary meaning, since they occur in a context having legal overtones. Ordinary meaning is generally regarded as (among other things) nontechnical meaning,<sup>174</sup> but common words are sometimes used in a technical legal sense.<sup>175</sup> When the pattern [*human*] *carry* [*firearm*] appears in a deontic expression, one has to consider the possibility that in any given case, its meaning is influenced to some extent by how the pattern is used in legal texts. I tend to think that any corpus lines reflecting such an influence should be excluded from the ordinary-meaning analysis. However, it will not be easy to reliably identify such uses. So rather than wrestling with that problem, I decided to tally the results for such uses separately from the other relevant uses.

The second difference displayed by the firearm-collocate results was that unlike the overwhelming majority of the other relevant corpus lines that were reviewed, a substantial number of the lines involving firearm collocates could be read as being not inconsistent with the schema HUMAN CARRY OBJECT IN VEHICLE. And that was the case regardless of whether the lines classified as deontic were counted: roughly 13.1% of the non-deontic lines were coded as being not inconsistent with the HUMAN CARRY OBJECT IN VEHICLE schema, and about 10.7% of all the relevant lines.<sup>176</sup>

Thus, a significant number of corpus lines could be understood as being not inconsistent with the interpretation in *Muscarello*. That is the strongest piece of corpus evidence supporting that interpretation. But I do not want to overstate the strength of that evidence. My description of these corpus lines as being “not inconsistent” with the

---

174. See, e.g., *District of Columbia v. Heller*, 554 U.S. 570, 576–77 (2008); *Nix v. Hedden*, 149 U.S. 304, 306–07 (1893); WILLIAM ESKRIDGE, *INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION* 60 (2016); Frederick Schauer, *Is Law a Technical Language?*, 52 *SAN DIEGO L. REV.* 501 (2015) (hereinafter *Is Law a Technical Language?*).

175. See, e.g., *Is Law a Technical Language?*, *supra* note 174, at 502.

176. The remaining of the lines were divided into two groups. With respect to about 41% of the relevant non-deontic lines, I was reasonably confident that the carrying of firearms in vehicles was not involved. (If the deontic lines were factored in, that percentage decreased to about 34%.) As to a smaller but still significant number of lines (24% of the non-deontic lines, and 20% overall), I leaned in that direction but with more uncertainty.

*Muscarello* is wishy-washy on purpose. In saying that such an interpretation is not ruled out, I do not mean to suggest that it is the most natural reading.

Third, unlike most if not all of the corpus lines that are discussed in the previous section, many of the corpus lines involving firearm collocates, referred to the carrying of firearms generally, as a *type* of action, rather than to a particular instance of a firearm being carried. Within that group of lines, there were two subgroups. In one of them, the AGENT (the person referred to as doing the carrying) was described as a specific person, so that the reference was to the general actions of a specific person. In others, the reference was to the carrying of firearms by people in general or by members of a particular category of people. Thus, the lines could be categorized as (1) general as to both the agent and as to the carrying, (2) specific as to the agent but general as to the carrying, or (3) specific as to both the agent and the carrying. There were no examples in which specific events of carrying did not involve the actions of a specific agent. That is not surprising; it is not possible for a specific carrying event to occur without the action of carrying being the act of a specific person.

The lines that were general as to both the agent and the action of carrying included the following:

the current law, which allows gun owners to openly carry rifles but generally prohibits openly displayed sidearms.

In response to the increasing number of people carrying guns inside their stores, Chipotle and Starbucks recently asked customers to refrain from  
Hawken is a replica of the rifles the mountain men carried. It comes in .50-caliber percussion or flintlock versions. \$637, percussion;

Lines that involved a specific agent but were general as to the carrying included these:

Lisa's turnaround began two years ago when she inadvertently brought the gun she carried to work for protection onto school property in her mother's car, she said

That's easy. If they want to carry a rifle, they can carry mine.

was in the thick of the action. He says he got the rifle he carries now in hand-to-hand combat. . . .

Finally, the following lines involved specific acts of carrying by specific agents:

border when the Marines said he fired at them with the .22-caliber rifle he was carrying. . . .

Early in grouse season last year, Hank had slipped. The shotgun he was carrying had the safety off. He tumbled forward and the gun went off, spraying

Out of the 164 lines that I identified as involving relevant uses, about 68% were general as to the agent, the carrying, or both.

This distinction between specific and general references strikes me as being significant primarily because it illustrates the kind of unexpected insight that corpus analysis can provide into how words are used. Considered on its own, I do not think it does much to support the holding in *Muscarello*. However, some of the lines that were coded as general in one respect or another could arguably provide the holding with some qualitative support.

While most of the HUMAN carry OBJECT lines in the KWIC and vehicle-related searches involved carrying the object for the purpose of moving it from one place to another, many of those in the firearms-related searches probably refer to carrying firearms for the purpose of keeping them with you at all times, like carrying a wallet or carrying keys (which the Cobuild dictionaries treat as a separate sense<sup>177</sup>):

- . . .In response to the increasing number of people carrying guns inside their stores, Chipotle and Starbucks recently asked customers to refrain from
- . . .For many of the women who are carrying guns, the appeal is not a symbolic demonstration of their Second Amendment rights
- . . .macho world, a city ruled by guns, whether carried by Somalis or United Nations soldiers.

These uses could arguably be clustered together with carrying a wallet and carrying keys, and therefore be covered by the separate sense that Cobuild provides for those uses. While that does not directly suggest that driving with a gun in the glove compartment counts as carrying the gun, it felt to me that in these lines, there was a diminution of the element of physical contact and manipulation that is inherent in uses such as these:

They moved stiffly, because the guns they carried under their sport coats or in the waistbands of their pants were uncomfortable and Most are armed with rifles, mainly carried on their shoulders

Moreover, some of the lines that were coded as general conveyed an overtone similar to what is conveyed by this definition of *carry arms or weapons* from the 6th edition of *Black's Law Dictionary*: "To wear,

---

177. See COBUILD1, *supra* note 44, at 208.



bear or carry them upon the person or in the clothing or in a pocket, for the purpose of use, or for the purpose of being armed and ready for offensive or defensive action in case of a conflict with another person.”<sup>178</sup> And while this definition focuses on carrying something “upon the person or in the clothing or in a pocket,” it is common for an expression to be used in a way that is related to an earlier use, but with the earlier meaning being broadened (and possibly weakened).<sup>179</sup> Therefore, the kind of use described by the *Black’s* definition might be extended to contexts in which the purpose for carrying the gun is the same but the gun is not kept on one’s person. Indeed, the dissent in *Muscarello* seemed to say that it would constitute carrying a firearm for the driver of a car to have a loaded gun accessible on the seat next to them.<sup>180</sup> And a similar interpretation was endorsed by Judge Kozinski, who justified it by appealing to something similar to the kind of broadening that I have described:

The key aspect of the narrow definition is not that the weapon actually be borne on the person. Rather, it is that the weapon remain within easy reach while the individual is in motion. Where an individual is walking, a gun in hand certainly amounts to carrying, but so does a gun in a holster or a shopping bag. The essence is that the weapon moves with the person and can be swiftly put to use. Where the individual is in a car, he need not actually be touching the weapon to make it move with him. [Footnote: The car might be thought of as a large, self-propelled shopping bag containing both the gun and the individual.] Because the car and its contents move in unison, any weapon that is within hand’s reach while the car is in motion can be said to be carried. The same would be true, of course, if the individual had the weapon concealed in a train compartment, a bus or, heaven forbid, an airplane.<sup>181</sup>

Overall, however, these attempts to tease out overtones from corpus data turn on distinctions that are impressionistic and probably too delicate to make a difference to a judge. Nevertheless, discussing them

---

178. *Carry Arms or Weapons*, BLACK’S LAW DICTIONARY 214 (6th ed. 1990).

179. *See, e.g.*, PHILIP DURKIN, THE OXFORD GUIDE TO ETYMOLOGY 235–37 (2009); DONKA MINKOVA & ROBERT STOCKWELL, ENGLISH WORDS: HISTORY & STRUCTURE 172–77 (2d ed. 2009).

180. *Muscarello v. United States*, 524 U.S. 125, 147 (1998) (Ginsburg, J., dissenting).

181. *United States v. Foster*, 133 F.3d 704, 706 (9th Cir.) (footnote omitted), *vacated*, 525 U.S. 801 (1998).

has hopefully given some additional sense of what corpus analysis can do.

*d. A new insight.* Putting aside whatever bearing the corpus data may have on the issue in *Muscarello*, it nevertheless is interesting for what it tells us about word meaning and the shortcomings of dictionary definitions. The data reveals shades of meaning that are not reflected in any dictionary I'm aware of and that, as far as I know, have never before been remarked on. One question this raises is whether the uses I've focused on should be regarded as instantiating a separate sense of *carry*.

But that question has no single right answer. Rather, the answer depends on what level of granularity is thought to be appropriate, and that in turn depends on what aspect of the meaning of *carry a [fire-arm]* is in question.

This resonates with a point that was made by Adam Kilgarriff—a point that I didn't fully appreciate until after completing my corpus analysis of *carry*: “[A] task-independent set of word senses for a language is not a coherent concept. Word senses are simply undefined unless there is some underlying rationale for clustering, some context which classifies some distinctions as worth making and others as not worth making.”<sup>182</sup>

This statement was made with respect to word-sense disambiguation in the field of Natural Language Processing, but I think it applies to lexicography and legal interpretation as well. In writing definitions that will appear in dictionaries, lexicographers know that the dictionary will be consulted by all kinds of different people who are looking for all kinds of different information about all kinds of different words. This compels a one-size-fits-all approach to definition-writing. But there is no reason to think that the level of granularity at which a dictionary is written is one that will best serve the needs of a particular legal issue. This is yet another way in which dictionaries are inadequate to the demands of legal interpretation.

## VII. CONCLUSION: LEARNING TO THINK LIKE A LINGUIST

I said at the beginning of this Article that the importance of corpus linguistics for legal interpretation goes beyond mere methodology. I

---

182. “I Don't Believe in Word Senses,” *supra* note 16, at 150.

have tried to show that corpus linguistics has generated insights about language and meaning that call into question some of the assumptions underlying the way that judges and lawyers deal with issues of word meaning. Indeed, one of the things that is called into question is the phrase *word meaning* itself. The relevant unit of meaning in many cases will be more than a single word: a phrase or some other multi-word expression. But at this point, what needs to be addressed is the nature of the appropriate analysis; questions of labeling can wait until later.

Although the significance of corpus linguistics is not limited to methodology, neither is it purely theoretical. It opens up new ways of thinking about issues of (what I will continue to call) word meaning: new kinds of questions to ask, new modes of analysis, and new lines of argument for litigators to pursue. And these new ways of thinking turn out to have methodological implications. We've seen that in the use of Corpus Pattern Analysis in reexamining the issue that divided the Supreme Court in *Muscarello*.

What this means (or rather, one of the things that it means) is that if we want corpus linguistics to fulfill its potential for enriching legal interpretation, simply teaching lawyers and judges how to run corpus searches will not be enough. Intelligently doing corpus analysis will in some cases require a little grounding in relevant ideas from linguistics, such as those that I have discussed here. It will be incumbent on those of us who advocate the use of corpus linguistics to develop materials that can provide that grounding. Because thinking like a lawyer will sometimes require the ability to think like a linguist.

