December, 1990

Linguistic Applications of LEXIS and WESTLAW

Fred R. Shapiro

Available at: https://works.bepress.com/aallcallforpapers/64/
LINGUISTIC APPLICATIONS OF LEXIS AND WESTLAW

Fred R. Shapiro*

A “full-text online database” is a service which allows a user, working at a terminal connected to a communications network, to search and retrieve the full text of documents such as newspaper articles or judicial decisions. Full-text systems typically have the ability to retrieve occurrences of desired words or combinations of words from the documents within their coverage. Therefore, even though they are created for nonlinguistic purposes, they may be suitable for various kinds of linguistic analysis. Two full-text databases with special potential for both specialized and general linguistic studies are the legal systems, LEXIS and WESTLAW, produced respectively by Mead Data Central and the West Publishing Company.

The documents which the two companies have made available for online full-text searching are closely parallel. Both LEXIS and WESTLAW include vast comprehensive collections of judicial opinions from United States federal and state jurisdictions. At first only opinions from recent decades were loaded, but both systems have steadily pushed back their historical coverage. LEXIS also has files of English, Scottish, Irish, Australian, New Zealand and French case reports. Both systems encompass nonjudicial materials including statutes, administrative decisions and regulations and periodical articles; LEXIS contains United States Supreme Court briefs as well. This enormous coverage

*Assistant Librarian for Public Services, Yale Law School; Consultant, Oxford English Dictionary and Oxford Law Dictionary. S.B., Massachusetts Institute of Technology 1974; J.D., Harvard University 1980; M.S.L.S., Catholic University of America 1982. Some of the ideas in this article were developed when I taught a class on legal semantics and lexicography as part of Morris Cohen’s course, Legal Research: Methods and Sources, at Yale Law School in the spring of 1986. I would like to thank Joyce Saltalamachia and Carl Yirka for their general encouragement and support.
gives LEXIS and WESTLAW great utility for research in legal language and, to some extent, in general speech as well.

This article surveys some of the potential linguistic applications of these extraordinary resources. In particular, I suggest ways in which the legal databases can facilitate the compilation of new law dictionaries which correct the glaring deficiencies of existing legal word-books, and ways in which the databases can be used to improve upon the historical record of legal language contained in the *Oxford English Dictionary (OED)*.

### I. CITATION-COLLECTION

Linguists and lexicographers collect "citations"—quotations including a given word or phrase in the context of actual usage—in order to document the word or phrase's existence, signification, geographical currency and other information such as "collocation, degree of formality, spelling, compounding, etymology, or grammatical data." The most obvious linguistic application of full-text databases is citation-collection, which I will introduce through an example.

In April 1985, Governor Mario Cuomo of New York said that the strongest objections to New York's new seat belt law were coming from "hunters who drink beer, don't vote and lie to their wives about where they were all weekend." After criticism from the National Rifle Association, Cuomo admitted that "My response was inartful." The NRA president, Alonzo Garcelon, responded, "Look up inartful in the dictionary. If you find it, let me know. I couldn't." William Safire of the *New York Times* then attacked Cuomo's usage in his August 4, 1985, "On Language" column, stating that "Inartful is in no dictionary because it is not a word."

It is true that inartful does not appear to be listed in any general or legal dictionary. However, I conducted a LEXIS search that revealed 594 occurrences of this term in American judicial opinions, strongly indicating that it is widely used in legal discourse to describe poorly drafted documents. I sent this information to Safire, who printed it in his September 1, 1985, column, conceding that "I may have been unduly harsh... on the Governor." I was also pleased to receive a letter from Governor Cuomo himself, which read as follows: "Thank you for your support with Mr. Safire. I was inartful; he was inaccurate; you were informative. Long live Lexis!"

---

4. *Id*.
5. *Id*.
6. A LEXIS search for inartfully retrieved an even greater number of opinions employing the adverbial form.
Full-text databases such as LEXIS or WESTLAW, by searching billions of words of text, can retrieve usage examples for less common words that elude even the largest dictionaries. *Inartful* is not such a word, however. It clearly is a common term that should have been included in the more substantial law dictionaries, and is representative of a sizable class of vocabulary items for which LEXIS and WESTLAW can be of tremendous value in redressing the deficiencies of existing legal lexicons.

Some of the shortcomings of existing law dictionaries have been spotlighted by David Mellinkoff in an excellent paper. Mellinkoff finds the two leading contemporary American word-books, *Black’s Law Dictionary* and *Ballentine’s Law Dictionary*, to be stuffed with “a legal miscellany . . . scraps of what is said to be law—Anglo-Saxon law, Hindu, Japanese, Jewish, Greek, Spanish, French, Roman, canon, ecclesiastical, civil, and something called ‘Old European’ law,” “disembodied snatches of law French and Latin,” “claptrap from the feudal system . . . the armor, the weapons, the ancient customs. Page after page of trivia.” William Anderson made the same point in 1894:

> The antiquated expressions in question . . . for one to ten and more centuries have been dead beyond resuscitation . . . It is believed that, in all the enacted and adjudicated law of the States . . . there can scarcely be found a single reference to the dead things I have in mind. Yet the pretense, in some quarters, still is, that an immense mass of these odds and ends of expressions, originating, as intimated, back six hundred years and more, should be minutely explained in one’s “law” dictionary, because he may “encounter” them in his reading or at law lectures.

Mellinkoff also identifies “another chunk” of *Black’s* and *Ballentine’s* as consisting of general terms of well-known meaning. He cites proper, improper, garden, horsepower, martini and sex as examples. I would add allergy, family Bible, north, revolution, happiness, usefulness and love making, which is defined in *Ballentine’s* as “one of the arts of seduction” and duly supported by a citation to the *American Jurisprudence 2d* article on Seduction. These words, like the antiquarian terms, have been selected without rhyme or reason, with countless comparable items being left out. In fact, both classes owe their inclusion largely to uncritical copying of older dictionaries which were compiled in a highly idiosyncratic and unscientific manner, and to publishers’ desires for pretentiousness and bulk.

---

10*BLACK’S LAW DICTIONARY* (5th ed. 1979).
12Mellinkoff, *supra* note 9, at 435.
13*Id.*
14*Id.*
16Mellinkoff, *supra* note 9, at 436.
Mellinkoff's critique neglects the flip side of Black's and Ballentine's hospitality towards valueless filler, namely, their failure to collect important newer vocabulary, with "newer" denoting terms entering currency within the last half-century. The reader will search in vain in Black's and Ballentine's for the word genocide or the phrase human rights. Theoretical terms such as legal realism or folk-law are nowhere to be found, nor are law-related social-scientific terms.

Some day a real dictionary of law will be compiled, in which the overgrown thicket of extraneous entries will be pruned and the space saved will be devoted to the addition of modern vocabulary selected on sound lexicographical principles.18 The compilers of the first real legal dictionary will find their task considerably eased by LEXIS and WESTLAW, which can be used to verify the existence of previously undictionaryed terms and to collect citational evidence illuminating their meaning and usage. By the same token, the databases' citation-gathering capabilities can be employed to assemble documentation supporting improved treatment of terms already listed in dictionaries.

With the help of evidence generated by LEXIS and WESTLAW, the first real dictionary will supply the kinds of information for each entry that are routine in general dictionaries but virtually unheard-of in Black's and Ballentine's: variant spellings, printed form, part of speech, inflected forms, etymology, field labels, usage labels and notes, derivatives, synonym notes, illustrative citations. Most importantly, such a dictionary will feature clear, concise definitions written by professional lexicographers on the basis of available evidence of overall usage, as opposed to the traditional law dictionary practice of relying heavily on undiscriminating collection of judicial authority in lieu of good definitions. The definitions will indicate the range of meanings of a term more fully than in past lexicons.

In a recent paper, Robert Burchfield maintained that "[t]he problem of monitoring the obsolescence of words (as opposed to their emergence) remains intractable."19 For legal terminology, however, LEXIS and WESTLAW can serve as barometers of obsolescence. A new law dictionary, seeking to exclude obsolete entries from its word-list, could adopt a policy that terms which do not appear in any nineteenth or twentieth century case on LEXIS or WESTLAW will be left out. Similarly, a law dictionary could be guided by data from the full-text systems in applying temporal labels like "obsolete" or "archaic" to certain entries.

Another class of linguistic applications of full-text databases involves frequency studies in which the number of search "hits" is the desired datum. Full-text databases usually will not give actual word-frequency figures, but will

---

18Legal slang will also be included. The student or beginning lawyer needs a guide to the everyday oral language of the profession far more than a compendium of archaisms which will never in fact be encountered.

enumerate the number of documents containing the search term. Such totals can be used as the bases for relative or approximate frequency statistics, bearing always in mind that commercial databases are not balanced representations of the printed English language as a whole, nor even of printed legal language as a whole.

LEXIS or WESTLAW frequency figures can be utilized for a host of linguistic purposes. A dictionary staff can use LEXIS frequency or WESTLAW frequency as a test of whether a term is common enough to merit inclusion in a word-list. The relative frequencies of variant spelling can be investigated.20

A law dictionary edited on these principles could exert a significant influence on the quality of legal writing. It could also improve the quality of LEXIS and WESTLAW research, since questions of synonymy and ambiguity are central to the challenge of formulating precise, effective full-text search strategies. An online version of the new dictionary’s synonym notes could provide LEXIS or WESTLAW with a badly needed automatic thesaurus function.

I believe that law students (and, not so incidentally, the general public) get a subliminal message from Black’s and Ballentine’s. The abundance of antiquarian esoterica crowding each page and the relatively unhelpful definitions convey the message that the law is arcane, that considerable training is necessary to comprehend even its language. The publication of a real law dictionary, explicating legal terminology in a less forbidding manner, could have a tangible impact in demystifying the law.

II. HISTORICAL LEXICOGRAPHY

Because searches on LEXIS, WESTLAW and other full-text systems can be limited by date, and retrieved documents are usually ranked chronologically, full-text databases are suitable for historical-lexicographical research. Historical lexicography is the study of the etymology, chronology and meaning of words and phrases by means of a method first proposed in Germany in the early nineteenth century and later exemplified by the Oxford English Dictionary. This historical method requires that each meaning of each word be traced, to the extent practicable, to its earliest appearance in print, and that all developments in the word’s usage be illustrated by dated and documented quo-

20To the extent that British materials are included, relative currency in the United States and Britain can be tested. For example, the Second Barnhart Dictionary of New English lists the word adversarial with the regional label British and the note, “In the U.S., adversary is the standard adjective used in legal and judicial contexts . . . adversarial occurs only in British citations in our file.” C. BARNHART, S. STEINMETZ & R. BARNHART, infra note 29, at 12. In fact, a LEXIS or WESTLAW search shows that adversarial has been common in American legal usage for some time. On LEXIS, for example, there are 3,706 American cases using this adjective.

tations using the word. 22 The project of tracing words and phrases to their earliest appearances in print is an enormous and difficult one, involving research of a highly sophisticated and ingenious nature. Now, however, human ingenuity can be supplemented by automated searches retrieving the earliest usage of a term in the documents covered by the database.

The coverage of LEXIS and WESTLAW extends far enough back in time to be useful for historical lexicography, with some of the databases’ files beginning in the eighteenth or nineteenth centuries. 23 In my own researches I have explored the utility of LEXIS and WESTLAW for tracing early uses of several important legal terms, such as executive privilege. Arthur Schlesinger, Jr. concluded in 1973 that “Executive privilege” seems itself to be of fairly recent American usage.” 24 Raoul Berger asserted on page 1 of his celebrated 1974 book, Executive Privilege: A Constitutional Myth, that “the very words ‘executive privilege’ were conjoined only yesterday, in 1958,” identifying Assistant Attorney General George C. Doub as the coiner. 25

Subsequently, two discussions in the journal American Speech pushed the date of first use back to 1950, when both parties in the case of Touhy v. Ragen 26 used the term in their briefs. 27 Safire’s Political Dictionary and the Second Barnhart Dictionary of New English cite the American Speech articles as the last word on the subject. 28 The LEXIS search EXECUTIVE PRIVILEGE AND DATE BEF 1950 or the WESTLAW search “EXECUTIVE PRIVILEGE” AND DATE (BEFORE 1950), however, readily yields two earlier occurrences of the expression, in a 1940 federal case, Glass v. Ickes, 29 and a 1941 Ohio case, Bigelow v. Brumley. 30

A central term in contemporary jurisprudence is human rights. Jordan J. Paust has written a number of articles on the application of this concept and

22 For an introduction to historical lexicography in the context of a remarkable biography, see K. M. E. MURRAY, CAUGHT IN THE WEB OF WORDS: JAMES A. H. MURRAY AND THE OXFORD ENGLISH DICTIONARY (1977).
23 As of mid-1989, LEXIS covers United States federal case law beginning in 1789, Ohio cases from 1821, Colorado from 1864, California from 1883, New York from 1884, Florida and Texas each from 1886, Arizona from 1898 and Illinois and New Jersey each from 1899. More minor collections of pre-1900 materials on LEXIS include: United States Attorneys General Opinions (1791), Delaware cases construing Delaware Corporation Law (1895) and United Kingdom Tax Cases (1875). WESTLAW’s pre-1900 files, with beginning dates, are: federal courts (1789), Ohio (1821), Colorado (1864), California (1883), Illinois (1885), New York (1885), Pennsylvania (1885), Florida (1886), Texas (1886), Louisiana (1887), Arizona (1898), Minnesota (1898), Massachusetts (1899), Michigan (1899), New Jersey (1899), Washington (1899), United States Attorneys General Opinions (1791) and Delaware corporation cases (1888).
27 Russell & Porter, Among the New Words, 48 AM. SPEECH 250 (1973); RUSSELL & PORTER, Among the New Words, 55 AM. SPEECH 112 (1980).
29 117 F.2d 273, 277 (D.C. Cir. 1940).
30 138 Ohio St. 574, 37 N.E.2d 584, 589 (1941).
term by United States courts\textsuperscript{31} and has identified \textit{Ex parte Milligan} (1866)\textsuperscript{32} as the earliest case he was able to find (by noncomputerized methods) in which the phrase was used.\textsuperscript{33} At the time Paust wrote, LEXIS and WESTLAW had not yet pushed their case coverage beyond the twentieth century, but now research on either system brings to light usages of human rights in two highly important earlier cases: \textit{Chisholm v. Georgia} (1793)\textsuperscript{34} and \textit{Fletcher v. Peck} (1810).\textsuperscript{35} They also retrieve occurrences of the expression in a number of prewar cases relating to slavery. Furthermore, WESTLAW has a Bicentennial database including documents pertaining to the framing of the Constitution. A search here reveals that human rights appeared in the notes of James McHenry at the Constitutional Convention in 1787,\textsuperscript{36} four years before the previous earliest known use of the words in any context, in Thomas Paine's \textit{The Rights of Man} (1791).\textsuperscript{37}

In many other instances, it is possible to improve upon the historical record of legal vocabulary as set forth in the \textit{Oxford English Dictionary} through full-text database searches. For example, the \textit{OED}'s earliest citation for plea bargaining is dated 1964 and for plea bargain is dated 1969, but LEXIS or WESTLAW searches retrieve a 1959 opinion using plea bargaining\textsuperscript{38} and a 1961 opinion using plea bargain.\textsuperscript{39} Reverse discrimination is documented by \textit{OED} with a 'first use' of 1969, but is found in the databases as early as 1964.\textsuperscript{40} These are but two of hundreds of 'antedatings' which I have accomplished on LEXIS and WESTLAW.

The ultimate beneficiary of historical research in legal terminology is the historiography of legal thought, because the history of words sheds light on the history of ideas. In 1929, upon the completion of the original \textit{Oxford English Dictionary}, the legal historian Theodore F.T. Plucknett called for the compilation of an historical dictionary of law, capacious enough to accommodate discussion of 'the questions which interest the lawyer most—namely, the relation


\textsuperscript{35}71 U.S. (4 Wall.) 2, 119 (1866).

\textsuperscript{33}Paust, \textit{Application of International Human Rights Standards by United States Courts}, supra note 31, at 501.

\textsuperscript{34}2 U.S. (2 Dall.) 419, 423 (1793) (Randolph, argument for plaintiff).

\textsuperscript{35}10 U.S. (6 Cranch) 87, 133 (1810) (Marshall, J.).


\textsuperscript{37}T. Paine, \textit{The Rights of Man} 110 (1791).

\textsuperscript{38}United States v. McCue, 178 F. Supp. 426, 436 (D. Conn. 1959). I have, however, discovered a prior citation by manual research: 'It is only in the privacy of his [the prosecutor's] office that he can conduct the plea bargaining which, although viewed as unethical in many quarters, is a highly useful as well as usual procedure in criminal administration.' Comment, \textit{Reform in Criminal Procedure}, 50 Yale L.J. 107, 110 (1940).


\textsuperscript{40}Strippoli v. Bickal, 42 Misc. 2d 475, 485, 248 N.Y.S.2d 588, 599 (Sup. Ct. 1964).
between changes in a particular word and changes in the law.\textsuperscript{41} More recently, David Mellinkoff described the goal of a complete historical law dictionary as "tantalizing," but argued that the "endless" historical and regional variation in the language of the law would pose an insuperable stumbling block to the creation of such a lexicon.\textsuperscript{42}

Mellinkoff is doubtless correct that an historical dictionary would be an ambitious undertaking, but is too pessimistic as to its prospects. He is probably unfamiliar with professional lexicographical methods which have succeeded in tackling areas of vocabulary much larger and more varied than the legal word-stock. Given a sufficient commitment of resources, an historical glossary of law can be published. I assert this in part because LEXIS and WESTLAW are available to facilitate citation-gathering. The databases would aid in tracing early uses for more modern vocabulary items, and would greatly ease the collection of illustrative examples for the modern history of all items. Oxford University Press has in fact announced plans for an Oxford Law Dictionary on historical principles, to be compiled with extensive aid from LEXIS and WESTLAW searching.

The vocabulary of the judicial opinions which comprise the bulk of searchable materials on LEXIS and WESTLAW is, of course, limited to legal jargon and such other words as are admitted into the conservative discourse of appellate judges. The corpus is sufficiently large, however, that it may sometimes yield valuable historical information on nontechnical items of general speech.

The most controversial word-usage of our time is undoubtedly the use of the word \textit{hopefully} as a sentence adverb meaning "it is hoped," as in "Hopefully, the project will take no more than a week." One panelist for the Harper Dictionary of Contemporary Usage said of this construction, "Its adherents should be lynched."\textsuperscript{43} The sentence adverb \textit{hopefully} is usually said to have sprung fully blown from the brows of "hack translators"\textsuperscript{44} (as an equivalent for the German \textit{hoffentlich}) around 1960. The OED does print a 1932 citation, but Robert Burchfield, the editor of the OED Supplement, tells me that this is an isolated instance; the second earliest example in their files is dated 1965.\textsuperscript{45} According to Sol Steinmetz of Barnhart Dictionaries, the Barnhart citation files have only one pre-1962 citation, dated 1956.\textsuperscript{46}

Yet LEXIS and WESTLAW research, executed by searching for any occurrences of the word \textit{hopefully} and examining the results to extract sentence-

\textsuperscript{41} Plucknett, \textit{Words}, 14 Cornell L. Q. 263, 265 (1929).
\textsuperscript{42} Mellinkoff, supra note 9, at 438-40.
\textsuperscript{43} W. Morris & M. Morris, Harper Dictionary of Contemporary Usage 311–12 (1975). The outrage inspired by \textit{hopefully} is rooted more in social factors than in grammar or logic; other sentence adverbs with similar grammatical and semantic functions, such as \textit{happily}, excite no objections. See Whitley, Hopefully: A Shibboleth in the English Adverb System, 58 Am. Speech 126 (1983).
\textsuperscript{44} W. Follett, Modern American Usage 169 (1966).
\textsuperscript{45} Letter from Robert W. Burchfield to Fred R. Shapiro (May 9, 1984).
\textsuperscript{46} Letter from Sol Steinmetz to Fred R. Shapiro (Sept. 7, 1984).
adverbial usages, turns up a number of usages from the 1940s and 1950s,\(^4\) as well as evidence of early transitional uses suggesting the process by which the sentence adverb became established. If judges were using *hopefully* in writing by the 1940s, we can assume that it was current in informal speech long before that.\(^4\) *Hopefully* provides a good illustration of the potential value of full-text legal databases for general linguistic studies, and also of online lexicography’s ability to retrospectively track items which are now of interest but which probably would have been missed by even the most alert human readers at the time of the earliest usages.\(^9\)

**CONCLUSION**

LEXIS and WESTLAW are now well established as tools for legal research. Their utility for scholarly investigations has, however, only begun to be explored. In this article I have suggested various applications of these systems to research in legal language. Other linguistic applications, such as frequency studies of phrases, studies of changes over time in word frequency, and grammatical studies, are also possible. Words, it is frequently said, are a lawyer’s only tools, and legal scholars now have extraordinary opportunities to scrutinize the words of the law through full-text database searching.


\(^{48}\) I also used the databases to find early examples of the sentence adverbs *thankfully* and *regretfully*. *Thankfully* in this sense means “let us be thankful.” *Regretfully* as a sentence modifier signifies “it is to be regretted” and is used synonymously with *regrettably*. OED dates *thankfully* 1966, but LEXIS and WESTLAW have it as early as 1961. Smyly v. United States, 287 F.2d 760, 764 (5th Cir. 1961). OED has *regretfully* from 1976, but the databases supply citations beginning in 1956, Rank v. United States, 142 F. Supp. 1, 62 (S.D. Cal. 1956), and again furnish evidence of earlier transitional usage.

\(^{49}\) Lexical usage—choices between options that are thought important by language users—is a subject of increasing interest to linguists and lay commentators, particularly in the United States. In this area I see full-text databases as having great potential utility as tools for marshalling evidence on the currency of controversial (or, for that matter, less controversial) locutions. Another usage controversy I have researched involves the expression *verbal agreement* or *verbal contract*. This is often criticized as an illogical use of the adjective *verbal*, since the most common meaning of *verbal* is “of, relating to, or consisting of words,” and all agreements are verbal in this sense. However, a LEXIS search retrieves 3,057 American cases employing *verbal agreement* and 1,420 employing *verbal contract*, providing overwhelming proof that this meaning of *verbal* is well established in legal language. (The *Oxford English Dictionary* documents usage of *verbal* in the sense “expressed or conveyed by speech instead of writing” as far back as 1591, and quotes writers including Pepys, Swift, Hume, Scott and Lincoln as illustrations of this meaning of *verbal* or *verbally*.)