

A Corpus Linguistic Analysis of “*Possessions*” in American English, 1760-1776

*James C. Phillips**

| | |
|--|-----|
| INTRODUCTION | 144 |
| I. BACKGROUND..... | 146 |
| A. Current Litigation | 146 |
| B. The Constitutional Text | 147 |
| C. The Weakness with Traditional Methods and Tools of Constitutional Interpretation | 149 |
| 1. The Limitations of Dictionaries | 149 |
| a. Dictionaries as “museum[s] of words” and linguistic intuition | 149 |
| b. “Lexicographical prescriptivism” | 150 |
| c. Dictionary piracy and idiosyncrasy..... | 151 |
| 2. Non-Systematic Usage Sampling..... | 152 |
| II. A BRIEF INTRODUCTION TO CORPUS LINGUISTICS | 153 |
| III. CORPUS LINGUISTIC ANALYSIS OF “ <i>POSSESSIONS</i> ” | 156 |
| A. Corpus of Founding Era American English (COFEA)..... | 156 |
| B. Collocate Analysis..... | 156 |
| C. Concordance Line Analysis | 157 |
| D. Sample of “ <i>Possessions</i> ” | 162 |
| CONCLUSION | 164 |

* Director and Associate Professor, Constitutional Government Initiative, Wheatley Institute, Brigham Young University. This research was funded in part by the Chapman Law School Summer Research Grant and the Institute for Justice.

INTRODUCTION

The U.S. Constitution's Fourth Amendment protects against unreasonable searches and seizures of persons, houses, papers, and effects.¹ Yet state constitutions often use different language, thus providing a different scope of protection. Specifically, starting with Pennsylvania in 1776, sixteen states have constitutional provisions that include *possessions* as protected from unreasonable searches and seizures.² And currently there is litigation in various state courts, including the Pennsylvania Supreme Court, over the meaning of this constitutional protection.

Possessions potentially implies more than houses, papers, or effects—arguably covering anything one possesses, including private land, which would significantly expand the coverage of such constitutional protection.³ But traditional tools of constitutional interpretation, such as dictionaries or etymology, often fall short in uncovering the original public meaning of constitutional text. Hence, courts (including U.S. Supreme Court Justices) increasingly have looked to corpus linguistics to better answer the linguistic questions that judges face in interpreting the

¹ U.S. CONST. amend. IV.

² ALA. CONST. art. I, § 5 (“possessions” language dates back to CONST. of 1819, art. I, § 9); CONN. CONST. art. I, § 7 (“possessions” language dates back to CONST. of 1818, art. I, § 8); DEL. CONST. art. I, § 6 (“possessions” language dates back to CONST. of 1792, art. I, § 6); ILL. CONST. pt. I art. XVI, § 6 (“possessions” language dates back to CONST. of 1780, pt. I, art. XVI, § 6); KY. CONST. art. I, § 10 (“possessions” language dates back to CONST. of 1792, art. XII, § 9); ME. CONST. art. I, § 5 (“possessions” language dates back to CONST. of 1820, art. I, § 5); MASS. CONST. art. XIV, pt. I (“possessions” language dates back to CONST. of 1780, pt. I, art. XIV); MI. CONST. art. I, § 11 (“possessions” language dates back to CONST. of 1835, art. I, § 11); MS. CONST. § 23 (“possessions” language dates back to CONST. of 1817, art. I, § 9); N.H. CONST. pt. I, art. XIX (“possessions” language dates back to CONST. of 1784, pt. I, art. XIX); OHIO CONST. art. I, § 14 (“possessions” language dates back to CONST. of 1802, art. VIII, § 5); PA. CONST. art. I, § 8 (“possessions” language dates back to 1776 Pa. Decl. of Rights, § 10); R.I. CONST. art. I, § 6 (“possessions” language dates back to CONST. of 1842, art. I, § 6); TENN. CONST. art. I, § 7 (“possessions” language dates back to CONST. of 1796, art. XI, § 7); TEX. CONST. art. I, § 9 (“possessions” language dates back to CONST. of the Repub. of Tex. of 1836, Decl. of Rights, § 5); VT. CONST. ch.1, art. XI (“possessions” language dates back to CONST. of 1777, ch. 1, art. XI).

³ The U.S. Supreme Court has held that the Fourth Amendment does not protect private land. *See Hester v. United States*, 265 U.S. 57 (1924). But in some states where “possessions” is part of their Fourth Amendment equivalent, state supreme courts have held that the term extends to private land, thus protecting such from warrantless searches. *See, e.g., State v. DuPuis*, 197 A.3d 343 (Vt. 2018); *Welch v. State*, 289 S.W. 510 (Tenn. 1926); *Falkner v. State*, 98 So. 691, 692–93 (Miss. 1924). Other states have held that people can expect privacy on their land. *See, e.g., State v. Bullock*, 901 P.2d 61 (Mont. 1995); *People v. Scott*, 593 N.E.2d 1328 (N.Y. 1992); *State v. Dixson*, 766 P.2d 1015 (Or. 1988).

words of the law.⁴ Understandably, judges use economic tools to tackle economic questions and historical tools to answer historical questions. Should they not use linguistic tools for linguistic questions? As Justice Frankfurter observed, “words are . . . the material of which laws are made. Everything depends on our understanding of them.”⁵ We can and should use the right tools for seeking this understanding.

This article will proceed in four parts. Part I introduces the question at issue in the context of the first state constitution to include the term: the Pennsylvania Constitution. It does so, at least in part, because other state constitutions arguably copied the Pennsylvania Constitution, and thus the meaning of that constitution likely sheds light on the state constitutions that followed it. Part I also describes the litigation where the issue of the meaning of *possessions* comes up. Part II highlights shortcomings of the traditional tools usually employed in constitutional interpretation. Part III explains how the tools of corpus linguistics can address these shortcomings. And Part IV presents a corpus linguistic analysis of the term *possessions*. This approach, more rigorous than that usually undertaken, provides data on the linguistic question that undergirds the legal issue—which reading of these state constitutions is more probable than the other. After all, a “problem in [legal interpretation] can seriously bother courts only when there is a contest between probabilities of meaning.”⁶ Corpus linguistics can help with that contest.

⁴ At the Supreme Court level, see *Carpenter v. United States*, 138 S. Ct. 2206, 2238–39 n.4 (2018) (Thomas, J., dissenting) (running a search in the Corpus of Founding-Era American English); *Lucia v. SEC.*, 138 S. Ct. 2044, 2056 (2018) (Thomas, J., concurring, joined by Gorsuch, J.) (citing Jennifer Mascott, Who Are “Officers of the United States?” 70 STAN. L. REV. 443 (2018)); *Facebook, Inc. v. Duguid*, 141 S. Ct. 1163, 1174 (2021) (Alito, J., concurring) (citing Thomas R. Lee & Stephen C. Mouritsen, Judging Ordinary Meaning, 127 YALE L. J. 788 (2018)); *Bostock v. Clayton County*, 140 S. Ct. 1731, 1769 n.22 (2020) (Alito, J., dissenting) (citing James C. Phillips, The Overlooked Evidence in the Title VII Cases: The Linguistic (and Therefore Textualist) Principle of Compositionality 3 (May 11, 2020) (unpublished manuscript), (https://www.supremecourt.gov/opinions/URLs_Cited/OT2019/17-1618/17-1618-3.pdf) [<https://perma.cc/X2HZ-VC2R>])). Numerous lower federal and state courts have cited to, discussed, or even done their own corpus linguistic analysis.

⁵ Garson Kanin, *Conversations with Felix*, READER’S DIGEST, June 1964, at 116–17 (noting that Justice Felix Frankfurter replied to counsel and said a question from the bench was just a matter of semantics).

⁶ Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527, 528 (1947).

I. BACKGROUND

A. Current Litigation

The Pennsylvania Constitution provides that “[t]he people shall be secure in their persons, houses, papers and possessions from unreasonable searches and seizures.”⁷ This constitutional clause is currently being litigated before the Pennsylvania Supreme Court.⁸ The case involves two private hunting clubs that own thousands of acres of land.⁹ Only members of the clubs may use the land, and they do so “to hunt, vacation, and enjoy nature.”¹⁰ For privacy, the properties are gated, and “no trespassing” signs are posted.¹¹

However, state game wardens have entered and traversed the properties without permission, probable cause, or a warrant.¹² And they do so because state statutes authorize game wardens “unfettered discretion” to come and go on this private property as they see fit.¹³ Specifically, these statutes give game wardens “the right and authority to go upon or enter any property, posted or otherwise, outside of buildings” when “exercis[ing] . . . their powers and duties.”¹⁴ So these officers of the state may “[g]o upon any land or water outside of buildings, except curtilage, posted or otherwise, in the performance of [their] duty.”¹⁵

The hunting clubs argued that these statutes are in conflict with the Pennsylvania Constitution because the term *possessions* includes land.¹⁶ Therefore, when property owners “signal that their land is not open to the public,” they “have a reasonable expectation of privacy and must be entitled to protection under [the Pennsylvania Constitution].”¹⁷ And “game wardens who want to search it must obtain consent or a warrant, or show a warrant

⁷ PA. CONST. art. I, § 8.

⁸ See *Punxsutawney Hunting Club, Inc. v. Pa. Game Comm’n*, Dkt. 23, No. 456 M.D. 2021, 2023 Pa. Commw. Unpub. LEXIS 541.

⁹ *Punxsutawney Hunting Club, Inc. v. Pa. Game Comm’n*, No. 456 M.D. 2021, 2023 Pa. Commw. Unpub. LEXIS 541, at *2 (Pa. Commw. Ct. Sept. 29, 2023).

¹⁰ *Id.*

¹¹ *Id.*

¹² See *id.*

¹³ *Id.*

¹⁴ 34 PA. CONS. STAT. § 303(c) (2023).

¹⁵ *Id.* § 901(a)(2) (2023).

¹⁶ *Punxsutawney Hunting Club, Inc.*, 2023 Pa. Commw. Unpub. LEXIS 541, at *3–4.

¹⁷ *Id.* at *4.

exception.”¹⁸ Yet the challenged statutes “authorize warrantless searches of land that is used and marked as private,” and so, the argument goes, are unconstitutional.¹⁹

B. The Constitutional Text

The Pennsylvania Constitution prohibits “unreasonable searches” of “persons, houses, papers and possessions.”²⁰ *Possessions* could have at least three meanings in the Pennsylvania Constitution. First, it could mean anything one possesses.²¹ Such a meaning would make *houses* and *papers* surplusage, as a house and papers are also things one possesses—and by that broad meaning, also possessions. This would violate the canon against surplusage,²² but it may be that the common meaning of *possessions* in this context is the broad meaning and does violate the canon, as the canon is not absolute.²³

Second, in the Pennsylvania Constitution, *possessions* could mean a subset of anything one owns, referring just to certain things one owns. We see this sense used currently in phrases like, “all of your worldly possessions,” “all of her possessions,” etc., wherein the word *possessions* appears to be used to refer to movable things one owns, other than land or a house: one’s belongings—personal property as opposed to real property. If that were the sense being used in the constitution, then it would not turn *houses* into surplusage, but arguably would still turn *papers*

¹⁸ *Id.*

¹⁹ *Id.*

²⁰ PA. CONST. art. I, § 8.

²¹ See, e.g., SAMUEL JOHNSON, 2 A DICTIONARY OF THE ENGLISH LANGUAGE (4th ed. 1773), <https://tinyurl.com/y23jff6jn> [<https://perma.cc/9X7F-7FVB>] (“The state of owning or having in one’s own hands or power; property The thing possessed.”); THOMAS SHERIDAN, 2 A COMPLETE DICTIONARY OF THE ENGLISH LANGUAGE 39 (3d ed. 1790), <https://tinyurl.com/578kbd53> [<https://perma.cc/8D4E-CUA6>] (“To have as an owner . . . to enjoy, or occupy actually”); NOAH WEBSTER, 2 AN AMERICAN DICTIONARY OF THE ENGLISH LANGUAGE 328 (1828), <https://tinyurl.com/ypb826b4> [<https://perma.cc/F6PC-UW8N>] (“[P]roperty in one’s power or command; actual seizin or occupancy . . . The thing possessed; land, estate or goods owned. . . .”).

²² See ANTONIN SCALIA & BRYAN A. GARNER, *READING LAW: THE INTERPRETATION OF LEGAL TEXTS* 174 (2012) (“If possible, every word and every provision is to be given effect None should be ignored. None should needlessly be given an interpretation that causes it to duplicate another provision or to have no consequence.”)

²³ *Id.* at 176–77 (“Put to a choice, however, a court may well prefer ordinary meaning to an unusual meaning that will avoid surplusage. So like all other canons, this one must be applied with judgment and discretion, and with careful regard to context. It cannot always be dispositive because (as with most canons) the underlying proposition is not *invariably* true.”).

into surplusage, unless somehow one's papers were just not seen as a possession in this narrower sense.

Third, there may be an implied "other" before *possessions*. This taps into the concept of *implicature*, wherein "what is said implicitly includes something else that is closely related."²⁴ For example, the brackets in the examples below show what was implied:

"Jack and Jill are married [to each other].

Bill insulted his boss and [as a result] got fired.

Nina has had enough [to eat]."²⁵

An example from the U.S. Constitution is Article I, Section 9: "No Bill of Attainder or ex post facto Law shall be passed [by Congress]."²⁶

If such implicature is contained in the Pennsylvania Constitution, so that it should read "persons, houses, papers, and [other] possessions,"²⁷ it would provide a reading between the broad first one and the narrow second one just noted above, both of which have surplusage problems. The "other" would mean that the word *possessions* includes all other possessions except those already listed, meaning that all types of possessions are covered without a surplusage problem. There is also some historical support for this implied "other" in lists that ended in the term *possessions*, as shown in search results below.

Additionally, it may not be technically accurate that this implied "other" triggers the *eiusdem generis* canon, wherein general, catch-all words that follow a list of specific words are limited by that specific list.²⁸ That is because the list includes an item that may not be property—persons—depending on natural rights theories prevalent at the Founding.²⁹ Still, if one applied the canon to the last two items on the list, which are types of possessions, we see that one is real property and one is personal

²⁴ Lawrence B. Solum, *Legal Theory Lexicon 086: Context and Meaning*, LEGAL THEORY LEXICON (last modified Sept. 18, 2022), https://lsolum.typepad.com/legal_theory_lexicon/2019/01/legal-theory-lexicon-086-context-and-meaning.html [<https://perma.cc/7Z32-UA3H>].

²⁵ *Id.*

²⁶ *Id.*

²⁷ PA. CONST., art. I, § 8.

²⁸ SCALIA & GARNER, *supra* note 22, at 199 ("Where general words follow an enumeration of two or more things, they apply only to persons or things of the same general kind or class specifically mentioned (*eiusdem generis*).").

²⁹ Would a right to protect one's body be found in the right to property? Right to liberty? Right to life?

property. This arguably means the broad catch-all term *possessions* should include both types—and thus includes property in the form of land.³⁰

C. The Weakness with Traditional Methods and Tools of Constitutional Interpretation³¹

1. The Limitations of Dictionaries

a. Dictionaries as “museum[s] of words” and linguistic intuition

Dictionaries do not declare which sense of a word is the ordinary one as dictionaries often struggle to deal with context. Perhaps that should not be surprising since dictionaries are just “museum[s] of words”³²—“historical records (as reliable as the judgment and industry of the editors) of the meanings with which words have in fact been used by writers of good repute.”³³ Thus, dictionaries “are often useful in answering hard questions of whether, in an appropriate context, a particular meaning is linguistically *permissible*,” not what is linguistically *probable* in a particular context.³⁴

So choosing one dictionary definition over another as the ordinary meaning of a word or phrase reveals more about one’s own linguistic intuition than objective ordinary meaning because it is that intuition that analytically closes the gap from dictionary evidence to the interpretive conclusion. Individual intuition lacks transparency, which is particularly problematic when that individual is a judge. Additionally, individual linguistic intuition suffers from at least two problems given it is informed by exposure

³⁰ The Oxford English Dictionary provides the following senses of *possession*:

2a. That which is possessed or held as property; something belonging to one, a piece of property; (in *plural*) belongings, property, wealth. . . . 2b. *Scottish*. A tenancy; a small farm, etc., held under lease, a smallholding. *Obsolete*. . . . 2c. A territory subject to a sovereign ruler or state; (now chiefly) any of a country’s foreign dominions.

Possession, OXFORD ENGLISH DICTIONARY, https://www.oed.com/dictionary/possession_n# [<https://perma.cc/HP2Y-ETGJ>] (last visited Feb. 20, 2024).

³¹ Part I.C draws somewhat upon James C. Phillips & Jesse Egbert, *A Corpus Linguistic Analysis of “Foreign Tribunal”*, 108 VA. L. REV. ONLINE 207, 213–19 (2022).

³² Frank H. Easterbrook, *Text, History, and Structure in Statutory Interpretation*, 17 HARV. J.L. & PUB. POL’Y 61, 67 (1994).

³³ HENRY M. HART, JR. & ALBERT M. SACKS, *THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 1375* (William N. Eskridge, Jr. & Philip P. Frickey eds., 1994).

³⁴ *Id.* at 1375–76 (emphasis added).

to language over one's lifetime. First, most lawyers, scholars, and judges are not representative of ordinary members of society, as they usually come from higher socioeconomic levels and have experienced significantly more education. These demographic factors affect the language to which they are exposed.

Second, people are still products of their time. This limits their linguistic intuition regarding times during which they did not live, evidenced by the phenomenon of linguistic drift.³⁵ If English never changed, then laws written prior to one's lifetime could be interpreted by relying on a later person's linguistic intuition. But English does change, sometimes significantly and speedily. For example, the constitutional term *domestic violence*, from the 1770s through the 1970s, meant insurrection, rebellion, or rioting within a state.³⁶ Yet starting in the 1980s, that rapidly began to change. And within about a decade, *domestic violence* was almost always used to mean "violent or aggressive behaviour within the home, esp[ecially] violent abuse of a partner."³⁷ The older sense, used almost exclusively for two centuries, has now almost completely disappeared. And that change occurred very quickly, in about a decade. Thus, relying on one's own linguistic intuition formed in a time after a constitution was adopted may cause someone to miss that linguistic drift has occurred and so inaccurately understand a constitutional word or phrase.

b. "Lexicographical prescriptivism"

Dictionaries can either define words according to proper usage or actual usage. Normative, or prescriptive, dictionaries "establish[] what is right in meaning and pronunciation," providing readers with what the dictionary editor considers the "proper" usage of each entry.³⁸ Thus, "the prescriptive school of thought relie[d] heavily on the editors of dictionaries to define and publish the proper meaning and usage of the terms."³⁹ By contrast,

³⁵ See generally Anne McCrary Sullivan, *Basic Students, Linguistic Drift, and the Language of the Future*, ENG. J., 1991, at 43–47.

³⁶ Thomas R. Lee & James C. Phillips, *Data-Driven Originalism*, 167 U. PA. L. REV. 261, 298–300 (2019).

³⁷ *Domestic Violence*, OXFORD ENGLISH DICTIONARY (Mar. 2006) (emphasis omitted), <https://www.oed.com/view/Entry/56663?redirectedFrom=domestic+violence#eid41827739> [<https://perma.cc/A5ZN-RQRV>]; Lee & Phillips, *supra* note 36, at 300.

³⁸ *Webster's Way Out Dictionary*, BUS. WK., Sept. 16, 1961, at 89, reprinted in JAMES SLEDD & WILMA R. EBBITT, *DICTIONARIES AND THAT DICTIONARY* 57 (1962).

³⁹ Samuel A. Thumma & Jeffrey L. Kirchmeier, *The Lexicon Has Become a Fortress: The United States Supreme Court's Use of Dictionaries*, 47 BUFF. L. REV. 227, 242 (1999).

"[t]he editors of a descriptive dictionary describe how a word is being used and, unlike their prescriptive counterparts, do not decide how a word should be used."⁴⁰

Historically, American dictionaries invoked proper usage. "Lexicographical prescriptivism in the United States is exactly as old as the making of dictionaries, because of the role played by the dictionary in a society characterized by a great deal of linguistic insecurity."⁴¹ That changed in the 1960s when Webster's Third International Dictionary controversially shifted to defining words based on actual usage.⁴² Thus, American dictionaries before the 1960s are less useful for determining how people actually used language.⁴³ Hence, dictionaries from the 1700s and 1800s may reflect more about their editor's sense of what is proper American English as opposed to how people ordinarily used and understood the language.

c. Dictionary piracy and idiosyncrasy

Dictionaries produced around the American Founding have additional weaknesses not found in modern dictionaries that flow from the fact that usually just one or two people created most dictionaries at the time. The epitome of these are the two most famous founding-era dictionaries: the dictionaries of Samuel Johnson and Noah Webster. As one scholar noted, "Johnson and Webster stand as the ultimate personifications of the solo artiste."⁴⁴ While "Johnson had his amanuenses" and "Webster had a single proof-reader, enlisted toward the end of the project[,] . . . these assistants were secondary figures. In neither case did the man whose name adorns the title page allow such helpers to influence his end product."⁴⁵

This solo-artiste style of dictionary-making creates at least two problems. First, it makes founding-era dictionaries rather

⁴⁰ *Id.*

⁴¹ HENRI BÉJOINT, TRADITION AND INNOVATION IN MODERN ENGLISH DICTIONARIES 116 (1994) (citation omitted).

⁴² See Norman E. Isaacs, *And Now, the War on Words*, THE LOUISVILLE TIMES, Oct. 18, 1961, reprinted in JAMES SLEDD & WILMA R. EBBIT, DICTIONARIES AND THAT DICTIONARY 79 (1962) (reporting that the editor-in-chief of Webster's Third stated that "the dictionary's purpose was to report the language, not to prescribe what belonged in it").

⁴³ True, to the extent people rely on dictionaries, even a prescriptive definition could somewhat reflect how people understood language, or influence how people used language, though it is second-best evidence.

⁴⁴ JONATHON GREEN, CHASING THE SUN: DICTIONARY-MAKERS AND THE DICTIONARIES THEY MADE 15 (1996).

⁴⁵ *Id.*

idiosyncratic. Thus, any particular definition may reflect more on that individual's understanding than society's understanding and usage of the language. After all, "dictionaries do not emerge from some lexicographical Sinai; they are the products of human beings. And human beings, try as they may, bring their prejudices and biases into the dictionaries they make."⁴⁶

The second problem is plagiarism. Perhaps because creating an entire dictionary oneself is a Herculean task, dictionary writers often plagiarized earlier dictionaries.⁴⁷ As one author described, "[t]he history of English lexicography usually consists of a recital of successive and often successful acts of piracy."⁴⁸ This tendency to plagiarize "can create a false consensus whereby it looks like all of the dictionaries independently agree, and thus reflect contemporaneous linguistic reality, but in actuality only reflect the views . . . of a few dictionary makers."⁴⁹ Plagiarism also means that later dictionaries may miss linguistic drift because they are copying definitions from dictionaries published as long ago as a century previously. Hence, idiosyncrasy and plagiarism undermine the utility of founding-era dictionaries in discerning how the ordinary public understood and used words.

2. Non-Systematic Usage Sampling

To avoid the weaknesses of dictionaries, one can sample actual usage from the relative time period. But one would need to do so in a systematic way and in sufficient numbers to have confidence in the results. However, much like dictionaries, examples of contemporaneous usage of a term in question often suffer from the same defect of relying on legislative history—looking out among the crowd and calling on one's friends. Put another way, there is a temptation to cherry-pick usage examples that support one's position. The methods below help overcome these shortcomings.

⁴⁶ *Id.* at 11.

⁴⁷ See, e.g., ALLEN REDDICK, *THE MAKING OF JOHNSON'S DICTIONARY, 1746–1773*, at 11 (Cambridge Univ. Press rev. ed. 1996); Gregory E. Maggs, *A Concise Guide to Using Dictionaries from the Founding Era to Determine the Original Meaning of the Constitution*, 82 GEO. WASH. L. REV. 358, 383–84 (2014).

⁴⁸ SIDNEY I. LANDAU, *DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY* 35 (1984).

⁴⁹ James C. Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S. Constitution: A Corpus Linguistic Analysis of American English from 1760–1799*, 59 S. TEX. L. REV. 181, 191 (2017).

II. A BRIEF INTRODUCTION TO CORPUS LINGUISTICS⁵⁰

Corpus linguistics is the empirical study of language using samples (or bodies) of texts called corpora (in the plural). A corpus is constructed in order to study a particular register (variety of texts associated with a situational context) or speech community (group of language users who share the same dialect or language norms).⁵¹ Corpus linguistics is premised on the idea that “the best way to find out about how language works is by analy[z]ing real examples of language as it is actually used.”⁵² In studying naturally occurring language use, corpus linguistics can avoid the observer’s paradox—the phenomenon whereby people tend to change their behavior when they are aware they are being studied (i.e., the Hawthorne Effect).⁵³

Corpus linguistics is founded on two premises: (1) that a corpus of texts can be constructed to be sufficiently representative of a particular register or speech community, and (2) that one can “empirically describe linguistic patterns of use through analysis of that corpus.”⁵⁴ So corpus linguistics “depends on both quantitative and qualitative analy[sis].”⁵⁵ And corpus linguistics results “in research findings that have much greater generalizability and validity than would otherwise be feasible.”⁵⁶ Because “a key goal of corpus linguistics is to aim for replicability of results, [researchers and] data creators have an important duty to discharge in ensuring . . . the data they produce is made available to analysts in the future.”⁵⁷

A corpus can be made of any kind of naturally occurring texts. Common examples include collections of samples of newspaper articles, books, or legal documents. The utility of a corpus will depend on the degree to which it represents the target language

⁵⁰ Part II draws somewhat upon Phillips & Egbert, *supra* note 31, at 219–24.

⁵¹ See TONY MCENERY & ANDREW HARDIE, *CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE* 1–2 (Cambridge Univ. Press 2012).

⁵² PAUL BAKER ET AL., *A GLOSSARY OF CORPUS LINGUISTICS* 65 (Edinburgh Univ. Press 2006).

⁵³ See HENRY A. LANDSBERGER, *HAWTHORNE REVISITED: MANAGEMENT AND THE WORKER, ITS CRITICS, AND DEVELOPMENTS IN HUMAN RELATIONS IN INDUSTRY* 14, 23 (Cornell Univ. 1958).

⁵⁴ THE CAMBRIDGE HANDBOOK OF ENGLISH CORPUS LINGUISTICS 1 (Douglas Biber & Randi Reppend eds., Cambridge Univ. Press 2015).

⁵⁵ Douglas Biber, *Corpus-Based and Corpus-driven Analyses of Language Variation and Use*, in THE OXFORD HANDBOOK OF LINGUISTIC ANALYSIS 159, 160 (Bernd Heine & Heiko Narrog eds., Oxford Handbooks Online 2015).

⁵⁶ *Id.* at 159.

⁵⁷ MCENERY & HARDIE, *supra* note 51, at 66.

domain of interest. Corpus representativeness depends on two key considerations—“what types of texts should be included in the corpus and how many texts are required.”⁵⁸ What is true for computing is true for corpus linguistics: “garbage in, garbage out,” as corpus-based results can be no better than the corpus being used (and it can be worse if the corpus data is not properly analyzed).⁵⁹ If a corpus does not adequately represent the texts used within the register or by the speech community one wants to make observations about, then other features of the corpus, such as its size, will make little difference.

One tool often used in corpus linguistic research is collocation. Some words “co-locate” more frequently than other words. One can think of this phenomenon as “word neighbors.” These semantic patterns of word association can sometimes be intuitive: we expect *dark* to appear more often in the same semantic environment as *night* than with *perfume*. But sometimes the patterns are surprising. This linguistic phenomenon has long been recognized in the law in the canon of construction called *noscitur a sociis*: “it is known by its associates.”⁶⁰ Linguists just put it a slightly different way: “[y]ou shall know a word by the company it keeps.”⁶¹

By seeing which words are collocates of each other, we can sometimes get additional insight into how people understand those words. This can be done in a corpus by searching for a word and indicating (1) how many words to the left or right (or both) of the search term one wants to examine, and (2) which statistical measure (e.g., frequency, MI score, T score) will be used to measure the strength of association.⁶² In this way, researchers are able to estimate how common it is for words to co-occur in close proximity. We can also use collocate analysis to see how usage patterns change. For instance, as I and a co-author showed in a paper, the top five collocates (in raw frequency) of the term *domestic violence* from 1760-1979 were (1) *against*, (2) *state(s)*, (3)

⁵⁸ JESSE EGBERT ET AL., DESIGNING AND EVALUATING LANGUAGE CORPORA: A PRACTICAL FRAMEWORK FOR CORPUS REPRESENTATIVENESS i (2022).

⁵⁹ *United States v. Esquivel-Rios*, 725 F.3d 1231, 1234 (10th Cir. 2013) (“Garbage in, garbage out. Everyone knows that much about computers: you give them bad data, they give you bad results.”).

⁶⁰ *Noscitur a sociis*, BLACK’S LAW DICTIONARY (10th ed. 2014).

⁶¹ John Rupert Firth, *A Synopsis of Linguistic Theory, 1930–1955*, *STUD. LINGUISTIC ANALYSIS* 1, 11 (1957).

⁶² See JESSE EGBERT ET AL., DOING LINGUISTICS WITH A CORPUS: METHODOLOGICAL CONSIDERATIONS FOR THE EVERYDAY USER 25–29 (2020).

protect, (4) *convened*, and (5) *invasion*.⁶³ This reflects the sense, as used in the Constitution, of a rebellion or insurrection within a state. But the top five collocates of *domestic violence* from 1980-2009 showed a radical shift: (1) *women*, (2) *abuse(d)*, (3) *honor*, (4) *national*, and (5) *victims*.⁶⁴ These collocates reflect the sense of violence against a member of one’s household. However, collocation tends to be more exploratory than confirmatory in nature. Why words are collocating with each other is not explained by the fact that they are doing so.

Another corpus method commonly used in legal interpretive research is concordance line analysis. Concordance lines can be used for qualitative analysis or in order to obtain frequency data. Concordance lines are excerpts from texts centered on a search term. In cases where there are many hits resulting from a corpus query, researchers can extract a random sample of concordance lines from the corpus.

To get meaning out of the concordance lines often requires classifying or categorizing (often referred to as “coding”) the search results. For instance, one could search for a particular word, then classify each result presented in a concordance line according to a particular sense of that word. Additionally, if greater context than one sentence is needed, one can expand the size of the text excerpt surrounding the search hit to account for more context. In this way, one could analyze the results to determine something a dictionary cannot usually convey: which sense is more common in a given context (i.e., the distribution of senses). This particular exercise, using concordance lines to classify senses, has proven to be an effective method for addressing questions regarding the meaning of words and phrases in legal texts. Further, the nature of the search results prevents one from cherry-picking examples. Of course, classifying senses involves a measure of subjectivity in considering the context to properly classify (or code) a sense. To mitigate this subjectivity, one can either use multiple, independent coders or one can make public ones, coding so that anyone can check it.

⁶³ See Lee & Phillips, *supra* note 36, at 298–300.

⁶⁴ See *id.*

III. CORPUS LINGUISTIC ANALYSIS OF “*POSSESSIONS*”

A. Corpus of Founding Era American English (COFEA)

This study will use the Corpus of Founding Era American English, or COFEA (rhymes with “Sophia”). COFEA is the only large corpus of American Founding Era materials in existence, with over 127,000 texts and 138 million words.⁶⁵ COFEA “covers the time period starting with the reign of King George III, and ending with the death of George Washington (1760-1799).”⁶⁶ The corpus “contains documents from ordinary people of the day, the Founders, and legal sources, including letters, diaries, newspapers, non-fiction books, fiction, sermons, speeches, debates, legal cases, and other legal materials.”⁶⁷ Most of the texts “have been pulled from the following six sources: the National Archive Founders Online; William S. Hein & Co., HeinOnline; Text Creation Partnership (TCP) Evans Bibliography (University of Michigan); Elliot’s Debates; Farrand’s Records; and the U.S. Statutes-at-Large from the first five Congresses.”⁶⁸

B. Collocate Analysis

To perform collocate analysis for this study, I searched in COFEA over the years 1760-1776 for all collocates six words to the right and left of *possessions*, eliminating what linguists call “stop words.”⁶⁹ As can be seen, *lands* is the second most frequent collocate.

| Rank | Collocate | Frequency | Rank | Collocate | Frequency |
|------|-----------|-----------|------|-----------|-----------|
| 1. | rights | 37 | 6. | America | 21 |
| 2. | lands | 35 | 6. | crown | 21 |
| 3. | titles | 31 | 8. | own | 20 |
| 4. | great | 29 | 9. | confirmed | 18 |
| 5. | majesty | 24 | 10. | part | 18 |

However, being frequently collocated with *possessions* does not explain the connection between *lands* and *possessions*. It could be that lands are a type of possession, or it could be that *lands* are

⁶⁵ See *About the Corpus*, CORPUS OF FOUNDING ERA AMERICAN ENGLISH (COFEA) <https://lawcorpus.byu.edu/cofea/concordances> [https://perma.cc/J2HR-YBS8] (last visited Jan. 30, 2024).

⁶⁶ *Id.*

⁶⁷ *Id.*

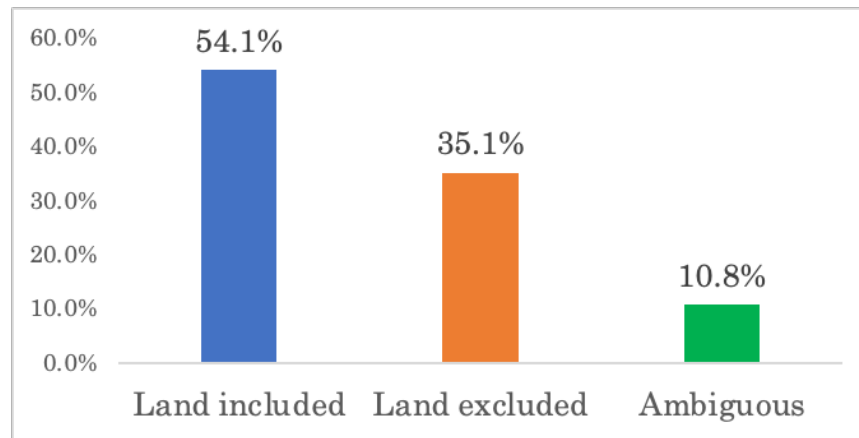
⁶⁸ *Id.*

⁶⁹ Stop words are words deemed insignificant to meaning, such as *the*, *is*, *which*, etc. See Kavita Ganesan, *What Are Stop Words?*, OPINOSIS ANALYTICS, <https://www.opinosis-analytics.com/knowledge-base/stop-words-explained/> [https://perma.cc/D9A8-6DV4] (last visited Feb. 1, 2024). There is no agreed upon list of such words. *Id.*

contrasted with *possessions* and thus excluded. To explore this relationship, one must look at concordance lines.

C. Concordance Line Analysis

For this study, I next examined all instances of *land* occurring within six words of *possessions* in COFEA from 1760-1776.⁷⁰ This resulted in 37 instances, because besides *land* that search formulation included *lands* and *landed*. I coded the results into one of three categories: encompassing land, excluding land, or ambiguous. I found no overwhelmingly dominant category, though the land-included category was the majority of usage:



The following 20 passages I coded as the land-included category (I have copied and pasted from COFEA, including with typos, non-standardized spelling, and punctuation errors):

1. persdhs who have had the occupation , or have been in the quiet possession , of any houses , lands , tenements , or other **possessions** , for the space of three whole years next before , and his or their estate or estates therein , is not ended
2. any other manner , by due form of law , that he or they entered into his house , lands , tenements or other **possessions** , by force . Provided always , that this act shall not extend to any person or persdhs who have had the occupation
3. assistant or assistants , jnstice or justices reside ; or of any wrongful detainer of any such houses , lands , tenements , or other **possessions** , by force , or strong handst that is to say , by , or with , such violent words or actions , as have a

⁷⁰ As this is not a sample, but all instances, there is no need to report a confidence interval for the following results.

4. the same are held by force ; then such assistants , or justices , shall cause the same houses , lands , tenements , or other **possessions** , to be re - seized , and the party to be put in possession thereof , who in such manner was put or
5. new Canadian Subjects, That nothing in the Order of that Date contained, should affect the Property of such as had **possessions** under proper Titles in Lands on the South Side of the Line, the Dominion of which was not disputed on
6. if it be found on such enquiry , that a forcible entry hath been made into houses , lands , tenements , or other **possessions** , or that the same are held by force ; then such assistants , or justices , shall cause the same houses , lands , tenements
7. or more assistants or justices of the peace , of any forcible entry made in any house , lands , tenements , or other **possessions** , lying within the county where such assistant or assistants , justice or justices reside ; or of any wrongful detainer of any
8. to cause to come before them , eighteen sufficient and indifferent persons , dwelling near unto the houses , lands , tenements , or other **possessions** , so entered upon or held as aforesaid , whereof fourteen shall be sworn well and truly to enquire of such forcible
9. government of this state , for the time being , touching or concerning the ratifying , confirming and quieting any titles to , or **possessions** of , lands within the district aforesaid , in cases not provided for by this act , and of and concerning the mode
10. happen, the farmers will be better off, than other people. Many of those that made up the Congress have large **possessions** in land, and may, therefore be looked upon as farmers themselves. Can it be supposed, they would be careless about
11. and House our assessed by the assessors appointed by the said Proprietors , and n 125 . r ; ce for quieting the **possessions** of such Persons who hold Lands there saw , as 2 ; under the said sales , notwithstanding there were some circumstanæ
12. much, as to render it well worth the testator's attention to change his will when he changes his landed **possessions** , and to be too great to be thrown into a sweeping residuary clause. Now these reasons are applicable to property
13. inhabitants were too few for the country, and want of people and money gave men no temptation to enlarge their **possessions** of land, or contest for wider extent of ground, are little more than generals of their armies; and though they
14. gentry, and an infantry of the commons. See 13 Edw. I. cap. 6, for arming the people according to their **possessions** in lands *. In the •ower are the records of the militia grants for cu••dy
15. collecting. My thoughts therefore on the subject you propose will be merely extempore. The opinion that our lands were allodial **possessions** is one which I have very long held, and had in my eye during a pretty considerable part of my

16. both lasting and scarce, and so valuable to be boarded up, there men will be apt to enlarge their **possessions** of land, were it never so rich, never so free for them to take: for I ask, what would a

17. was not the genius of the feudal policy to encourage cities, or to shew any regard for their **possessions** and immunities, these lands had been seized, and shared among the conquerors. The barons to whom they were grant

18. or Household, such necessaries, as may be raised on the Lands of the Company. 5. That any Lands, or other **possessions** of the said Philip Mazzei, which he may at this time have, and which are proper for the purposes of

19. in such manner that their location would remain permanent, fixed, and certain, would prevent disputes, differences, and law suits, quiet **possessions**, and of course render lands more valuable. Section I. Be it therefore enacted by the Senate and House of Representatives

20. happen, the farmers will be better off, than other people. Many of those that made up the Congress have large **possessions** in land, and may, therefore be looked upon as farmers themselves. Can it be supposed, they would be careless about

There are several ways that it becomes clear that the use of the term *possessions* in the excerpts above included land. Sometimes, the term was modified by an adjective indicating that it was only referring to possessions of land, such as “allodial possessions” or “landed possessions.” Likewise, sometimes the term was part of a larger phrase with the same effect as the adjectives just noted: “possessions of land” or “possession in land.” Of course, the term as used in the Pennsylvania Constitution neither has an adjective before nor is followed by a prepositional phrase that clarifies that it is only referring to land.

A third way the excerpts above show that land was included in *possessions* was by context, where reading additional material before and after shows that the type of *possessions* being referred to are landed ones. The fourth way *possessions* included land was when it was the last item in a list that included lands, usually preceded by the term “other.” The Pennsylvania Constitution does not include “other” before *possessions*, though it may be implied.

Finally, some of the excerpts above seem particularly relevant to the context of unreasonable searches, wherein they speak of forcible entry by government officials.

The 13 results below I placed in the land-excluded category:

1. not chargeable or liable, nor have not been bounden, charged or hurt of their bodies, liberties, franchises, lands, goods, nor **possessions** within the same county, but by such laws as they have agreed unto—and also, they have no knights, citizens

2. the hoidfe , lands , or possessions of any other person , or being entered peaceably , thall forceably hold the house , lands or **possessions** of any other person within this government , thall be proceeded against and punished as by the severall statutes made against
3. Thou shalt not covet, &c.” He coveting, did take the Saw-Mill-Logs, Boards, and also, the Lands, Labours, **possessions** , Farms, Tenements, &c. &c. from the rightful Owners, Proprietors, and first Occupants thereof, without
4. be convinc’d, it seems, that the Bishops of the English church, ought to enjoy the church “livings, lands and **possessions** ,” have seats in parliament, and even become members of the board of TRADE. We doubt not indeed, but
5. namely the manner of government, from time to time, to be used, the ordering and disposing of the lands and **possessions** , and the settling and establishing of a trade there, or such like, there shall be held and kept, every year
6. It is storied in their own history, that when the Emperor Constantine endowed the church with lands and **possessions** , the voice of an Angel was heard in the air, crying, Hodie venenum infunditur in ecclesiam. This day
7. do right upon writs of assise brought before them by such as are wrongfully thrust out of their lands and **possessions** , &c. 4. Of nisi prius, directed to the judges and clerk of assise, by which civil causes grown to
8. minds Religion was precious in their eyes; they were willing to leave houses and lands, and many dear and valuable **possessions** , for the sake of enjoying it in its purity. But they were men, and like other good men they were
9. to distress him by all the ways they can think of, such as the seizing on his Castles, Lands, and **possessions** , provision being only made for the safety of the persons of the King and Queen, and of their children.
10. namely the manner of government, from time to time, to be used, the ordering and disposing of the lands and **possessions** , and the settling and establishing of a trade there, or such like, there shall be held and kept, every year
11. have Means by which they might effectually perform these Duties.—Considering these Things, they liberally set apart Livings, Lands and **possessions** , for the Use of the Church; and tho’ these have been invaded by the avaricious Hand of Persecution, and mangled
12. of any of the inheritors or inheritance of the said county, of their bodies, liberties, franchises, goods, lands, tenements or **possessions** , being within the said county. For if any such act should be made, it were clean contrary to the liberties
13. failing of redress, shall lawfully distress and aggrieve the king all manner of ways, as by taking his castles, lands, **possessions** , &c. till redress is granted. After the restoration comes the corporation-act, and declares all resistance unlawful. The same

Here we can tell that *possessions* does not include land because it is listed separately from it without something like “other” or “nor” preceding the term. And the sense of *possessions* seems to be something like personal property, as opposed to real property, as often other items in the list—castles, houses, tenements, lands, etc.—are real property.

Finally, I found four instances sufficiently ambiguous to categorize them as such:

1. in Edward the Confessor’s time.” So that the Norman King• claim no other right in the lands and **possessions** of any of their subjects, than according to English law and right. And so tender were they of p[ro]perty

While at first glance this seems to be an instance of land-excluded *possessions*, in reading the broader context before and after this excerpt, it could be referring to houses and tenements, so I coded it as ambiguous.

2. they were called to govern and protect. Our fathers would never have forsook their native land, delightsome habitations and fair **possessions** , and in the face of almost every danger and distress, sought a safe retreat, for the enjoyment of religious and

I could not tell whether *possessions* was being used as a synonym for *land* and *habitations* since the latter can mean a land where one dwells, and thus for rhetorical flourish it was a list of three very similar things. Alternatively, the word *habitations* could have been used in the sense of houses, and thus we have a list of three different things wherein *possessions* likely means personal property.

3. General Afembly met , and by the authority of the fame , That whosoever shall forcibly enter into the hoidfe , lands , or **possessions** of any other person , or being entered peaceably , thall forceably hold the house , lands or possessions of any other person

On the one hand, this seems to be treating *possessions* as distinct from real property, indicating a reference to personal property, but on the other hand, one cannot forcibly enter into personal property, so this could mean other real property, such as tenements.

4. Complaint for the Future , may remain , that any unjust Measures are used to defraud the Indians of their Lands or **possessions** ; and agreeable to their Request at said Treaty ; Be It Enacted by the . Authority ajorejaid , That no Sale , Conveyance , or

I could not tell whether this was referring to personal property, as would be implied by the phrasing, or whether the “or” was being used in the sense of “or, in other words, their,” whereby *possessions*

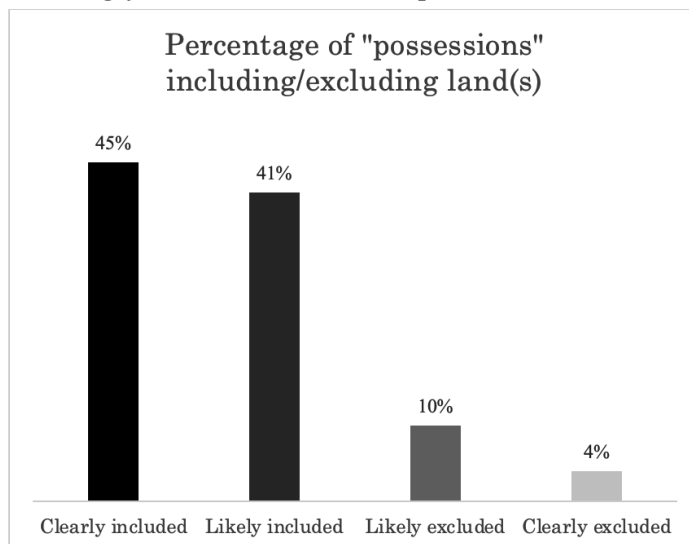
is just another way to describe lands. Perhaps my uncertainty was driven by a question of what possessions “Indians” had that they were defrauded of other than lands.

D. Sample of “Possessions”

Because *possessions* in the context of *land* differs from the language of the Pennsylvania Constitution, I next searched for all instances of *possessions* from 1760-1776 in COFEA.⁷¹ This returned 507 results. I downloaded these to a spreadsheet and then randomly selected a number⁷² for the first concordance line to code and coded every fifth line in order to sample 100 concordance lines.⁷³ I coded each line into one of four categories:

- (1) Clearly includes land
- (2) Likely includes land
- (3) Likely excludes land
- (4) Clearly excludes land

As can be seen in the following chart, the results were overwhelmingly in favor of the term *possessions* including land.

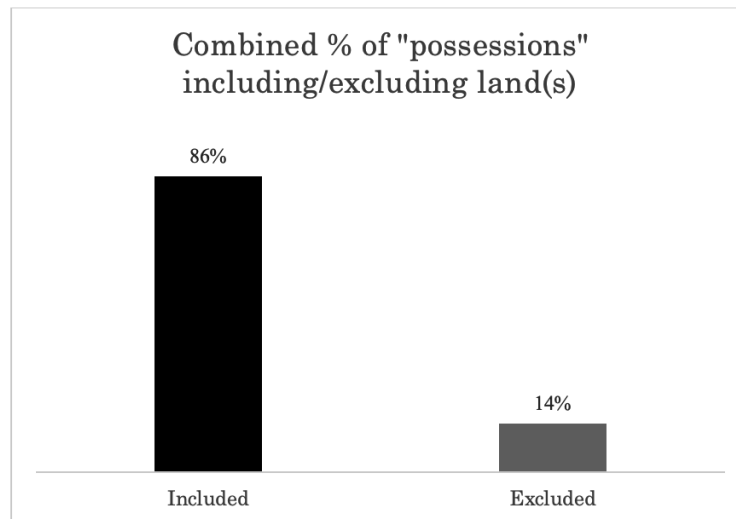


⁷¹ This time period covered 28,182,558 words and 19,342 documents.

⁷² See Appendix, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4780581 [<https://perma.cc/G8XQ-Q5SZ>].

⁷³ This resulted in 99 coded lines. So, I randomly selected a 100th line to make the sample complete (line 421). If a line to be coded was not readable or was quoting a similar constitutional provision, as a few were, I coded the next line.

In nearly half of the sample (45%), *possessions* clearly included land. And in another two-fifths of the sample (41%), land was likely included in the term *possessions*. Yet only 10% of the time was land likely excluded, and just one of every twenty-five instances, (4%) of the term results, was land clearly excluded. By collapsing the "clearly" and "likely" categories, the pattern becomes even more stark.



A full 86% of the time *possessions* likely or clearly included land and 14% of the time, *possessions* likely or clearly excluded land.⁷⁴ In other words, it was over six times more likely that *possessions* included land than that it did not (and over 11 times more likely when just looking at the instances where land was either clearly included or excluded).⁷⁵ In fact, sometimes the word *possessions* was being used as a synonym for *lands*.

This analysis is arguably more relevant than the first analysis above where the term *possessions* is used in the context of *land* since that is not how these constitutional provisions are worded. Thus, it appears in the sample taken of language usage from 1760-

⁷⁴ Clearly included (45%), likely included (41%), likely excluded (10%), clearly excluded (4%).

⁷⁵ With a confidence level of 95%, and assuming an equal population proportion, a sample size of 100 from a population of 507 yields an 8.79% margin of error. Thus, there is a 95% chance that the real value in the underlying population ranges from 77-95% including land and 5-23% excluding land.

1776, when the term *possessions* was used, it overwhelmingly included land.

CONCLUSION

Founding Era Americans tended to use the word *possessions* to include land one owned. In the context of the lemma *land*, a majority of the time the word *possessions* appeared to include land as property. More significantly, when looking more broadly at any instance of the term *possessions*, whether or not the lemma *land* was used nearby, early Americans used the term to include land approximately 86% of the time. This is evidence, then, that the Pennsylvania Constitution, and likely other early state constitutions, were originally understood to protect against unreasonable searches of one's land—thus providing broader protection than the U.S. Constitution's Fourth Amendment.