Nos. 17-1618, 17-1623, 18-107 VIDED

IN THE

Supreme Court of the United States

◆◆◆

No. 17-1618

GERALD LYNN BOSTOCK,

Petitioner,

—v.—

CLAYTON COUNTY, GEORGIA,

Respondent.

(Captions continued on inside cover)

ON WRITS OF CERTIORARI TO THE UNITED STATES
COURTS OF APPEALS FOR THE ELEVENTH, SECOND AND SIXTH CIRCUITS

**BRIEF FOR *AMICI CURIAE* CORPUS-LINGUISTICS
SCHOLARS PROFESSORS BRIAN SLOCUM,
STEFAN TH. GRIES, AND LAWRENCE SOLAN
IN SUPPORT OF EMPLOYEES**

MELINDA BOTHE
DANNA SELIGMAN
ALLEN & OVERY LLP
1101 New York Avenue, NW
Washington, DC 20005
(202) 683-3800

ANDREW RHYS DAVIES
  Counsel of Record
JUSTIN ORMAND
REBECCA DELFINER
GIDEON DUKE-COHAN
CHANGHEE HAN
ALLEN & OVERY LLP
1221 Avenue of the Americas
New York, New York 10020
(212) 610-6300
andrewrhys.davies@allenovery.com

Attorneys for Amici Curiae

TABLE OF CONTENTS

TABLE OF AUTHORITIES

PAGE(S)

Cases

STATEMENT OF IDENTITY AND INTEREST[1]

This amicus curiae brief is submitted on behalf of amici Professors Brian Slocum, Stefan Th. Gries, and Lawrence Solan. Amici are linguists, professors, and scholars of corpus linguistics.

Professor Brian Slocum is a professor at the University of the Pacific McGeorge School of Law. He has published several books, law review articles, and other peer-reviewed reports about the application of linguistics to statutory interpretation.

Professor Stefan Th. Gries is Professor of Linguistics in the Department of Linguistics at the University of California, Santa Barbara, and Chair of English Linguistics at the Justus-Liebig-Universität Giessen. Between 2013 and 2017, he was a Visiting Chair of the Centre for Corpus Approaches to Social Science at Lancaster University. Between 2007 and 2019, he was a Visiting Professor at the Linguistic Society of America Linguistic Institute, and in the spring semester of 2017, he was the Leibniz Professor at the Research Academy Leipzig of the University of Leipzig.

Professor Lawrence Solan is the Don Forchelli Professor of Law at Brooklyn Law School and has also served on the faculty at Yale Law School. He has published six books on language and the law, and three articles on the use of corpus linguistics to interpret legal texts.

Drawing from large corpora of texts collected from real-world sources, such as books, magazines, and newspapers, corpus linguistics provides information about the public understanding of the meaning of language in specific locations at particular times, and therefore helps determine what the ordinary meaning of words was at those times. Amici's work in the field of corpus linguistics has been cited in judicial opinions and published in several legal journals.

Amici's interests are in the philosophy of language, and in how corpus linguistics can assist judges and legal practitioners in reliably divining the meaning of undefined statutory language through rigorous, empirically-based analysis. Amici submit this brief to assist the Court in understanding the ordinary meaning of "sex" in Title VII through a corpus linguistics analysis of how the words "sex" and "gender" were used, and were not used, in the 1960s when Title VII was enacted.

SUMMARY OF THE ARGUMENT

As relevant here, Title VII of the Civil Rights Act of 1964 prohibits discrimination "because of . . . sex." 42 U.S.C. § 2000e-2(a)(1). Several of the lower courts that have considered these and similar cases have opined that the term "sex" could not have been understood in 1964 to encompass discrimination against employees for being lesbian, gay, bisexual, or transgender.[2] The employees demonstrate in their briefs that they should prevail even if "sex" as used in Title VII means only male or female, but Amici demonstrate below it is not true that "sex" had such a limited meaning in 1964.

Corpus linguistics is a study of words in their context. It provides reliable evidence of what particular words and phrases meant at certain times and places in history. Corpus linguistics is more rigorous and therefore more reliable than other modes of interpretation, such as an individual jurist's intuition or even a dictionary. That is because corpus linguistics analyzes how words were actually used in everyday settings. Here, Amici's corpus-linguistics analysis shows that "sex" did not have the limited meaning that the employers and some of the judges below

---

[2] *See, e.g., Zarda v. Altitude Express, Inc.*, 883 F.3d 100, 145 (2d Cir. 2018) (Lynch, J., dissenting, joined by Livingston, J.) ("'In common, ordinary usage in 1964—and now, for that matter—the word "sex" means biologically *male* or *female*; it does not also refer to sexual orientation.'") (quoting *Hively v. Ivy Tech Cmty. Coll. of Ind.*, 853 F.3d 339, 362-63 (7th Cir. 2017) (Sykes, J., dissenting) (emphasis in original)).

ascribe to it. Rather, in the 1960s, when Title VII was enacted, the term "sex" encompassed a diverse set of referents and could have encompassed the contemporary conceptions of sexual orientation and transgender status.

ARGUMENT

I. CORPUS LINGUISTICS IS A HIGHLY EFFECTIVE TOOL FOR DIVINING THE ORDINARY MEANING OF A STATUTORY TERM

The ordinary-meaning canon dictates that an undefined statutory term—such as the word "sex" used in Title VII—be given its ordinary, everyday meaning.[3] Statutory interpretation involves a quest for the meaning a reasonable person would understand the author to be conveying by using a given term in a given context. The question is not what the drafter subjectively meant to convey through the words chosen, but rather, "what those words would mean in the mouth of a normal speaker of English, using them in the circumstances in which they were used."[4]

This ordinary-meaning doctrine accords with the nature and use of a statute—the employment of natural language to accomplish a statutory purpose.[5] The basic premise of the ordinary-meaning doctrine is that statutory language

---

[3] *Encino Motorcars, LLC v. Navarro,* 138 S. Ct. 1134, 1140 (2018); *Taniguchi v. Kan Pac. Saipan, Ltd.,* 566 U.S. 560, 566 (2012); *see also* Antonin Scalia & Bryan A. Garner, *Reading Law: The Interpretation of Legal Texts,* 70, 435 (2012).

[4] *See* Oliver Wendell Holmes, *The Theory of Legal Interpretation,* 12 Harv. L. Rev. 417, 417-18 (1899).

[5] *See generally* Heikki E. S. Mattila, *Comparative Legal Linguistics* (Christopher Goddard trans., 2d ed. 2013) (examining the functions and characteristics of legal language and the terminology of law).

should be interpreted in light of the standards of communication that apply outside the law.[6] If successful communication is the goal, as it must be when a legislature uses statutory language, that language should be understandable in the same way by everyone who is subject to the statute.[7] The ordinary-meaning canon, fundamental to legal interpretation, thus reflects a presumption that legal language corresponds to language used in a non-legal context. Statutory terms should therefore be interpreted by reference to general principles of language usage that apply equally outside the law.

Corpus linguistics offers a highly and uniquely effective tool for divining the ordinary meaning of statutory words. That is because corpus linguistics provides the interpreter with context that is wholly missing when a term is read in isolation. A corpus linguistics analysis determines the context in which a term was actually used in the relevant place at the relevant time, and thereby more precisely informs the meaning of a term than other methods of statutory interpretation.

---

[6] *See* Brian G. Slocum, *Ordinary Meaning: A Theory of the Most Fundamental Principle of Legal Interpretation* 3 (2015).

[7] As Herman Cappelen, *Semantics and Pragmatics: Some Central Issues, in Context Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics* 19 (Gerhard Preyer & Georg Peter eds., 2007), explains, "[w]hen we articulate rules, directives, laws, and other action-guiding instructions, we assume that people, variously situated, can grasp that content in the same way."

A. Corpus Linguistics Offers a Reliable, Empirically-Based, Contextual Guide to the Ordinary Meaning of Statutory Terms

Corpus linguistics is a scientific discipline at the intersection of linguistics, digital humanities, computer science, and statistics and information theory.[8] It is a branch of linguistics based on the statistical analysis of data from a corpus.[9] A corpus[10] is a compilation of written and transcribed spoken language used in authentic communicative contexts, such as in newspapers or novels, that is placed into a machine-readable database. The basic premise of using corpus linguistics as a tool of interpretation is that by analyzing real examples of language as it was actually used at a specific point in time in a particular location, the researcher can reveal facts about how a certain term was ordinarily used and understood in everyday settings.[11]

The Corpus of Historical American English ("COHA") is the largest structured corpus of historical American English, and it contains a compilation of written language as used in the United States from 1810 to 2000.[12] Corpus

---

[8] *See generally* William J. Crawford & Eniko Csomay, *Doing Corpus Linguistics* 5-11 (2016).

[9] *See id.* at 6-7.

[10] The plural of "corpus" is "corpora."

[11] *See* Crawford & Csomay, *supra* note 8, at 5-9.

[12] *See* Corpus of Historical American English, https://www. english-corpora.org/coha/ (last visited June 18, 2019).

databases like COHA can be used to research the use of a term during a specified period of time in the United States, and the search results provide the term in context, as it was used by the public during the specified time period.

For the ordinary meaning of a term to be understood, the term must in some sense be generalizable across contexts, and not be shaped by legal considerations alone. As such, the meaning of a term should be subject to empirical verification. Corpus linguistics provides such empirical verification through a systematic and neutral method of investigating the meaning of a given term.

Corpus linguistic analyses are "based on the evaluation of some kind of frequencies." [13] Frequency of use is a crucial aspect of what distinguishes an ordinary meaning from some meaning that is perhaps conceivable but unordinary. In other words, frequency is an indicator of how ordinary a given meaning or usage of a term is or was. Corpus linguistics can illustrate the number of senses, or meanings, that a linguistic expression may have and the most frequently used meaning. [14] But it is not merely a matter of counting examples and pointing to the most frequent usage. Other important factors provide clues as to what the

---

[13] Stefan Th. Gries, *What is Corpus Linguistics?*, 3 Language & Linguistics Compass 1225, 1226 (2009).

[14] *See id.* at 1225, 1228.

prototypical meaning of a term was—including dispersion, which is discussed below.[15]

In this way, corpus linguistics brings the advantages of empirical testing to statutory interpretation. Corpus analysis is replicable,[16] generalizable,[17] and transparent: any researcher with Internet access can access an online corpus and retrieve the same data as the original researcher. Although analyses of the data and inferences drawn from them may vary, application of basic linguistic or mathematical principles will lead to replicable results.

The increasing availability of scientifically-based research tools, such as corpus linguistics, has caused some scholars to suggest that these tools have the potential to transform the exercise of statutory interpretation into an "empirical" inquiry.[18] Although legal interpretation will never be simply an empirical inquiry, some aspects of interpretation are empirical in nature. For instance, interpretation cannot proceed without some consideration of the conventional

---

[15] The assumption underlying most corpus-based analyses is the so-called distributional hypothesis, that formal differences reflect, or correspond to, functional differences (*i.e.*, semantics). *See id.* at 1228.

[16] Tony McEnery & Andrew Hardie, *Corpus Linguistics: Method, Theory and Practice* 66 (2011).

[17] Douglas Biber, *Corpus-Based and Corpus-Driven Analyses of Language Variation and Use, in* The Oxford Handbook of Linguistic Analysis 159, 159 (Bernd Heine & Heiko Narrog eds., 2010).

[18] *See generally* Thomas R. Lee & Stephen Mouritsen, *Judging Ordinary Meaning*, 127 Yale L.J. 788 (2018).

meanings of words.[19] Corpus analysis can provide this information. That is, corpus analyses can bring empirical data and objective scientific methods to bear on difficult questions such as which sense or meaning of a term is ordinary, which meaning is possible but infrequently used, and which meaning is extremely rare. With corpora from different time periods, this kind of analysis can also be performed for different points of time such as when a statute was enacted, when it was amended, and the present time.

Corpus linguists engage in both qualitative and quantitative analyses that follow valid principles of language research. Corpus linguistics research is thus a systematic and neutral method of researching language usage and meaning. [20] The methodology employed in corpus linguistics research allows any results to be tested and replicated, isolating the analysis from subjective influences.[21]

B. Due to Its Objective, Empirical Basis, Corpus Linguistics Is Preferable to—or at Least Provides a Valuable Supplement to—Other Commonly Used Methods of Interpretation

By providing information about how language was actually used, corpus linguistics offers a more objective and reliable interpretation than other modes of interpretation, such as an

---

[19] *See generally* Slocum, *supra* note 6.

[20] Biber, *supra* note 17, at 160.

[21] Mcenery & Hardie, *supra* note 16, at 66.

individual jurist's subjective interpretation, or the use of dictionaries that provide definitions devoid of context.

For instance, a reasonable judge may ask what a "reasonable person" would deem to be the ordinary meaning of a term. The reasonable-person standard, however, provides dubious externality when a court purports to apply the standard without additional external determinants. Corpus linguistics provides empirical verification of those external determinants. A corpus linguistics analysis contextualizes a term by organizing empirical, replicable evidence about the term's meaning and context into a framework that represents valid linguistic choices.[22]

Dictionaries are the primary alternative source of information about the conventional meaning of language. Judicial reliance on dictionary definitions has increased significantly since the 1980s due in part to the judiciary's increased focus on linguistic meaning.[23] Many judges appear to believe that dictionaries provide an expert, neutral, and external standard for the ordinary meaning of words. Yet, the use of dictionaries to determine the ordinary meaning of statutory language may result in inaccuracies.

---

[22] *See* Mark C. Suchman, *The Power of Words: A Comment on Hamann and Vogel's Evidence-based Jurisprudence Meets Legal Linguistics—Unlikely Blends Made in Germany*, 2017 BYU L. Rev. 1751, 1758 (2018).

[23] *See* 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, 486-87 (2013).

A dictionary is a highly abstract construct that presents words individually and takes them "away from their common use in their customary settings." [24] Although dictionaries can be useful as a general matter, "the listing of words as a set of isolated items can be highly misleading if used as a basis of theorizing about what words and their meanings are." [25] In addition, dictionary compilers are often unable to develop meaningful orderings of senses, thus intentionally leaving out the very information that judges have assumed dictionaries to provide. [26] As a result, different dictionaries may present different definitions of the same term, and selection of a particular dictionary over another might unduly sway the resulting interpretation. These factors have caused some courts to conclude that dictionaries simply do not always offer a helpful or reliable means to interpret statutory language. [27]

---

[24] M. A. K. Halliday & Colin Yallop, *Lexicology: A Short Introduction* 24-25 (2007).

[25] *Id.* at 25.

[26] *See generally* Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. Rev. 1915 (2010).

[27] *See United States v. Costello*, 666 F.3d 1040, 1044 (7th Cir. 2012) ("Dictionary definitions are acontextual, whereas the meaning of sentences depends critically on context, including all sorts of background understandings."); *see also Am. Bankers Ass'n v. Nat'l Credit Union Admin.*, 306 F. Supp. 3d 44, 68 (D.D.C. 2018) (noting that "a term does not necessarily mean the sum of its parts" and turning to corpus linguistics along with early 1930s-era judicial opinions—after starting with 1930s era

Corpus linguistics can help remedy these shortcomings. Unlike dictionaries that present acontextual definitions, corpus linguistics allows for the meaning of a term to be investigated in relation to other words with which the term co-occurs. Corpus linguistics is a method for studying language in use, and can thus account for context in ways that dictionaries cannot.

C. Courts Have Recognized That Corpus Linguistics and the Methods Underlying It Can Be Valuable When Interpreting Statutory Language

Judges have long recognized that empirical evidence about the meaning of words can have value when interpreting statutory language. For instance, this Court and at least one court of appeals have consulted online resources and newspaper articles to help identify the ordinary meaning of various statutory terms.[28] Moreover, federal and state judges, including individual members of this Court, have specifically

---

dictionaries—to determine the meaning of a phrase from a 1934 statute).

[28] *See, e.g.*, *Muscarello v. United States*, 524 U.S. 125, 128–130 (1998) (considering newspapers' use of the term "carries" as evidence of its ordinary meaning as used in the firearms chapter of the federal criminal code); *Costello*, 666 F.3d at 1044 (interpreting the statutory term "harboring" by, *inter alia*, performing internet searches to identify the common objects of that term); *see also Tex. Dep't of Hous. & Cmty. Affairs v. Inclusive Communities Project, Inc.*, 135 S. Ct. 2507, 2534 & n.2 (2015) (Alito, J., dissenting, joined by Roberts, C.J., Scalia, J., and Thomas, J.) (considering a newspaper's use of the statutory expression "because of" as evidence of its ordinary meaning).

recognized the potential usefulness of corpus linguistics for that purpose.[29]

II. IN THE 1960s, THE ORDINARY MEANING OF "SEX" WAS NOT LIMITED TO BINARY MAN/WOMAN DISTINCTIONS

Amici here used the Corpus of Historical American English ("COHA") to examine how the statutory term "sex" was ordinarily used in the 1960s, when Title VII was enacted. That

---

[29] *See Carpenter v. United States*, 138 S. Ct. 2206, 2239 n.4 (2018) (Thomas, J., dissenting) (citing to the BYU Corpus of Historical American English and the BYU Corpus of Founding Era American English, as evidence that the expression "expectation(s) of privacy" was not in common use at the founding); *Lucia v. SEC*, 138 S. Ct. 2044, 2056 (2018) (Thomas, J., dissenting, joined by Gorsuch, J.) (citing law review article that performed a corpus-linguistics-based analysis of the meaning of "Officers of the United States" at the founding); *People v. Harris*, 885 N.W.2d 832, 838-39 (Mich. 2016) (citing the Corpus of Contemporary American English as evidence of the ordinary meaning of "information" as used in a state statute); *see also id.* at 850 n. 14 (Markman, J., concurring in part and dissenting in part) (relying on the COCA, but disagreeing on the meaning of the statutory term).

At least two justices of the Utah Supreme Court have supported the use of corpus linguistics to ascertain the ordinary meaning of statutory terms. *See Brady v. Park*, --- P.3d ----, 2019 WL 2052350, at *29 n. 109 (Utah 2019) (Lee, C.J., concurring in part and dissenting in part); *Fire Ins. Exch. v. Oltmanns*, 416 P.3d 1148, 1164 & n.9 (Utah 2018) (Durham, J., concurring in part and concurring in the result); *State ex rel. J.M.S.*, 280 P.3d 410, 419 & n.3 (Utah 2011) (Lee, J., concurring); *In re Adoption of Baby E.Z.*, 266 P.3d 702, 724 & n.21 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment). The full court has so far been "divided on the viability and utility of this sort of empirical analysis." *Craig v. Provo City*, 389 P.3d 423, 429 & n. 3 (Utah 2016).

research demonstrated that "sex" had a broad, inclusive meaning, and that it was not limited to a strictly binary or biological meaning.

A. A Corpus-Linguistics Analysis Reveals That the Word "Sex" Was Not Used in Only a Binary Male/Female Sense in the 1960s

Amici performed a corpus-linguistics analysis of *sex*[30] as used in the 1960s. That analysis strongly suggests that, at that time, *sex* was the one word that was employed for what today is expressed with sex, gender, and sexual orientation.

A concordance for the case-insensitive string *sex* in the 1960s portion of COHA returned approximately three thousand hits. Amici studied a pseudorandom sample[31] of those hits to determine the degree to which *sex* in the 1960s was used at that time in a strictly binary male/female classification sense, and the extent to which *sex* was used in the 1960s potentially for a more diverse set of referents. Amici concluded that, insofar as *sex* was used to refer to the act of sex, it was not limited to heterosexual sexual activity. And insofar as *sex* was used as a mode of classification, it was not limited to a binary male/female classification.

---

[30] Italics are used herein to indicate that a word is being mentioned (meta-linguistically) as opposed to only being used. *Compare* "The word *car* has 3 letters," *with* "I am looking to buy a new car."

[31] The sample was the first approximately 350 examples from a randomly re-ordered set of the full 3,000 results.

Amici's analysis assumed that the noun *sex* is polysemous with, minimally, two senses:

- **sex₁** references the "act of sex" or "having sex" (maybe with an intercourse prototype);[32] and

- **sex₂** references the classificatory "biological sex" reading that could theoretically be defined genetically or chromosomally and that perhaps has male/female as prototypical categories and intersex as a special or hybrid type.[33]

Amici's pseudorandom sample included 347 instances of sex: 258 instances of sex₁ and 89 instances of sex₂. Amici annotated the former category by reference to the information provided about the participants in sex₁. The data showed that the 258 instances of sex₁ (i.e., the act of sex) involve:

- 107 cases of sex₁ where the context makes clear that what is referred to is sexual activity between a man and a woman;

- Three instances of sex₁ between two men;

- One instance of sex₁ between members of

---

[32] Examples (from the data in Amici's pseudorandom sample) include "the devout, who must abstain from food, drink, sex from dawn to sundown," "some young people use sex as an instrument of rebellion," "I'm against using sex as a weapon under any circumstances," and compounds such as "sex education," "sex crime," "sex criminals," "sex drive," which all seem to invoke sex₁.

[33] Examples (from the data in Amici's pseudorandom sample) include "some women have more female sex genes than others," "a woman could or should find quite as much pleasure with her own sex as she does with men," or "twenty gamblers of both sexes pressed up against the green baize."

the same but unspecified sex;

- 147 instances that did not provide enough information to decide.

Crucially, although sex between a man and a woman is the most frequent *classifiable* use[34] in the data, it is not the most frequent attested use. The most frequent attested use reflects the 147 cases that do not specify the sexes of the persons involved in sex_1. Thus, the existence of cases referring to same-sex sex_1 and the large number of cases in which sex_1 is used without specifying the sexes of the participants suggest that, even in the 1960s, the meaning of sex_1 already was not clearly limited to heterosexual sex. In turn, these findings do not support an argument that the term "sex," as used in the 1960s, should be interpreted as limited to heterosexual sexual activity.

As for the 89 instances of sex_2 (i.e., sex used as a classification), thirty make a clear reference to a two-way classification of sex (i.e., male versus female), while 58 are compatible with a more fine-grained classification because their context does not provide evidence that only a two-way classification interpretation was intended.[35] In other words, although the most frequent classifiable uses involve a binary classification (30 of the 89 total), the more frequent kind of attestations do not commit to a simple binary classification (59 of the 89 total). This indicates

---

[34] A use is classifiable when it is possible to determine the word sense from the context in which the word is used.

[35] One example was unclear.

that sex$_2$ did not specify a mere binary interpretation in the 1960s. In turn, these findings do not support an argument that the term "sex," as used in the 1960s, should be interpreted as limited to a binary male/female classification.

B. "Gender" Was Not a Commonly-Used Word in the 1960s

To explore whether the term *gender* was used in the 1960s as it is today, and to consider whether the answer to that question provides an explanation for the use of the word *sex* in the broad sense described immediately above, Amici performed a case-insensitive search for the string *gender* in the 1960s portion of COHA. Amici's analysis of the results of that search reveals that *gender* was a very rarely used word in the 1960s, which may well explain why *sex* had a broad meaning, encompassing what today we would call *gender*.

Tellingly, Amici's search for *gender* returned only 101 hits (as compared with over 3,000 hits for *sex*). These hits were extremely unevenly distributed in this section of the corpus data. The results are below:

• Twenty-one of the total 101 instances of *gender* were from a single source: a 1965 science-fiction novel about a genderless alien from a planet whose inhabitants have no concept of individuality;[36]

• Four instances were from a Jean-Paul

---

[36] Hortense Calisher, *Journal from Ellipsia* (1965).

Sartre book [37] and four instances were from a Christy Borth book;[38]

- Two sources with six instances each;

- Three sources with two instances each;

- Fifty-four sources with one instance each.

The extremely uneven distribution of *gender* as a noun or a verb is even more pronounced than the above distribution suggests. This is because the above search actually returns a number of hits that are arguably irrelevant. Specifically, 72 of the 101 hits reference the verb *to engender* (47 *engendered*, 11 *engenders*, 10 *engender*, four *engendering*).

This means that the instances of *gender* that are truly relevant to the current discussion are only either 29 instances (all matches that do not instantiate the verb *engender*) or 26 instances (all instances of the noun *gender*). To take the higher number, 29, the distribution of the corpus is extremely skewed or clumpy, both in terms of where they occur and the intended meaning. This is because of these 29 instances:

- Twenty are from the novel *Journal from Ellipsia*;

- Two are from TIME magazine;[39]

- Two are from John P. Hughes's novel *The Science of Language*;[40] and

---

[37] Jean-Paul Sartre, *The Communists and Peace* (1968).

[38] Christy Borth, *Mankind on the Move: The Story of Highways* (1969).

[39] *The Future of Swearing*, Time, Sept. 15, 1967.

- The remaining instances occur only a single time in their corpus files.

Moreover, eight of the twenty-nine instances actually refer to the notion of grammatical gender (including the above TIME magazine and *The Science of Language* examples), which, in spite of its name, is not at all the same as gender in the 'sex/male-vs-female' sense. One of the twenty-nine is a proper name and another cannot be included because it is a TIME magazine reference to gender in legislation and, thus, gives rise to the very issues we are discussing.[41]

In sum, *gender* in the relevant sense occurs only nineteen times in the 1960s portion of COHA: *gender* (16), *genders* (1), *genderless* (1), and *gendering* (1). All but one instance of these are from the very specialized source mentioned above, an avant-garde science fiction novel about genderless aliens.

This assessment of the absolute rarity of the word *gender* in the 1960s can be supported both linguistically and statistically. As for the former, one can identify words that have the same frequency in the corpus data of that time period (although raw-frequency comparisons, while widespread, are in fact too coarse an approach).[42]

---

[40] John P. Hughes, *The Science of Language* (1962).

[41] *The Congress: Now the Talking Begins*, Time, Feb. 21, 1964.

[42] For example, the following is a list of random words (one beginning with each letter of the alphabet) that have the same frequency as the noun *gender* in the 1960s COHA data (all homogenized to lower case): *avanti, bailing, callas, darien, explication, fightin, garters, hard-headed, idolized, jailing,*

As for the latter, Amici propose a more advanced and precise statistical analysis of what is called *dispersion.*

Dispersion is a statistic that quantifies the way a word is distributed in a corpus in a way that goes far beyond frequency. A word x can be distributed very evenly in a corpus, which means that the chance of seeing x in a randomly chosen part of the corpus (such as a file or a text) is high. Conversely, x can be distributed very clumpily, which means that the chance of seeing it in a randomly chosen part of the corpus (such as a file or a text) is very low. Examples of the former include most function words such as determiners (*the,* *a*), prepositions (*of,* *in*), conjunctions (*and,* *or*), etc. Examples of the latter include highly specialized terms of art (*potassium permanganate*), rare proper names, or even typos (such as *seperate* or *commisisoner*).

The reason dispersion is so important is that it is a better indicator of word commonness than is frequency. Words that have the same frequency can vary significantly in their dispersion, and dispersion is usually the measure that better matches native-speaker intuitions.[43] However,

---

kayano, leprosy, metromedia, nightgowns, oscillation, pan-american, ques, rhubarb, sambuco, three-cornered, untamed, vassall, widder, x2, yaks, and *zarzuela.*

[43] *See generally* Stefan Th. Gries, *Dispersions and Adjusted Frequencies in Corpora.,* 13 Int'l J. Corpus Linguistics 403 (2008); Stefan Th. Gries, *Dispersions and Adjusted Frequencies in Corpora: Further Explorations, in* Corpus Linguistic Applications: Current Studies, New Directions (Stefan Th. Gries, S. Wulff, M. Davies eds., 2010).

until approximately ten years ago, little was known about dispersion statistics, and their computation can be extremely labor-intensive because computing the dispersions of all words in a corpus can require many hours to complete, even on clusters of computers.

For the present issue, Amici computed the best dispersion statistic for all approximately 316,000 different word forms in the 1960s portion of COHA. This measure is called *DP* (for *Deviation of Proportions*) and theoretically ranges from nearly zero (words that are extremely evenly distributed such as *to*, *a*, and *and* in COHA 1960s) to nearly one (words that occur in only a single part of the corpus, such as *sociolinguistics*, *janizaries*, *bayonetting*, and *mooniness* in COHA 1960). The *DP*-value obtained for even the most generous version of *gender*, the twenty-nine cases that include the linguistic ones as well as the verbs and the adjective, is 0.9858641. This is a value that is extremely close to the theoretically possible maximal *DP*-value, which is indicative of extremely uncommon words.[44]

---

[44] This numerical result is strengthened by identifying words that have the same dispersion values (within rounding precision) in the corpus data of that time period. The following is a list of random words (one beginning with each letter of the alphabet) that have the same dispersion as { *gender, genders, gendering, gendered, genderless*} in the 1960s COHA data (all homogenized to lower case): *aky., brilliantp250that, caricatured, drambuic, emilythen, five-and-ten-cent, grittiness, homeroom, invitedher, jamaican-based, kllai, lepage, mlf, nierkusii, out-but, puses, quibbles, revealedp251by, supra- rational, topologically, unrhetorical, vincentdo, wiic-tv, x/2o, yearth*, and *zautla*. While there are somewhat ordinary words in this list (*caricatured* or

The above results can be visualized as follows. Figure 1 on the next page is a plot that represents on the x-axis the frequency of words (logged to the base of 10) and on the y-axis the *DP*-values of the same words. Each word is represented by a grey point. Words that are more frequent (on the right) at least tend to be more evenly dispersed, but the crucial finding is the red dot, which represents {*gender, genders, gendering, gendered, genderless*}. Clearly, the joint frequency of these expressions is already quite low, but their dispersion is truly minimal, indicating that, in American English in the 1960s, *gender* was an extremely uncommon word.

---

quibbles), it speaks to the rarity of {*gender, genders, gendering, gendered, genderless*} in the 1960s that it is as well dispersed in the data as are text-processing errors in the corpus: *brilliantp250that* (which should be *brilliant* [p. 250 of the book] *that*) or *revealedp251by* (which should be *revealed* [p. 251] *by*).

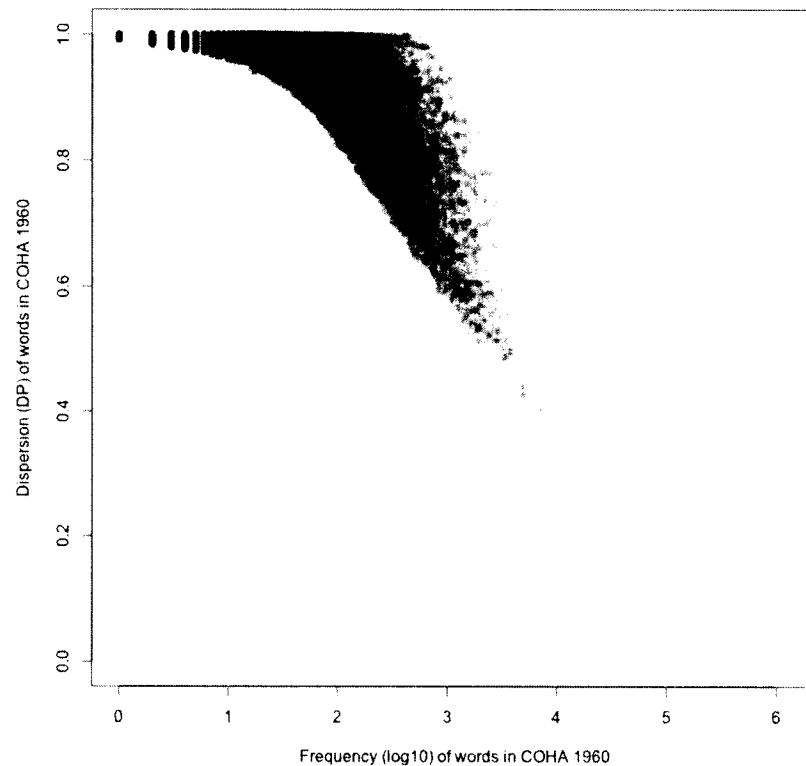**Frequency (log10) and dispersion (DP) of all approx 316K word forms in COHA 1960**

Figure 1: Frequencies and dispersions of words in COHA 1960; the red dot represents {*gender*, *genders*, *gendering*, *gendered*, *genderless*}

C. In the 1960s, the Word "Sex" Could Have Encompassed What We Now Call Gender and Sexual Orientation

As shown above, the data for *gender* in the 1960s is extremely sparse, and the examples demonstrate that *sex* in the 1960s subsumed what now is described as *gender*. The one use of *gender* that was not from *Journal from Ellipsia* is the following:

(1) although by her superior force she had overborne his visible reluctance, she, being a woman, or at all events of the *female **gender***, could never quite forget that she had done the wooing.[45]

Arguably, one could replace *gender* in (1) by *sex*, as the 1960s component of COHA contains multiple examples of *sex* not *gender*, which when preceded by *female* would be semantically completely compatible with the use of *gender* in (1). This is illustrated in (2), (3), and (4).

(2) Masterson himself had fewer sexual encounters than he boasted and most of these with the *female **sex***.[46]

(3) the uncanny ability of the *female **sex*** to see through the subtlest ruses of men.[47]

(4) Ezra had a weak stomach for alcoholic beverages and that his conquests of the *female **sex*** were largely, although not entirely, imaginary.[48]

Even in the (generally unrepresentative) novel *Journal from Ellipsia*, those uses of *gender* that are not highly literary (*see* (5) or (6)) do have straightforward analogous uses of sex (*see* (7) and (8) respectively):

[45] Kate Douglas Wiggin, *The Eventful Trip of the Midnight Cry* (1895).

[46] James Purdy, *Eustace Chisholm and the Works* (1967).

[47] Milford E. Anness, *Song of Metamoris* 229 (1964).

[48] George Garrett, *Do, Lord, Remember Me* (1965).

(5) was it possible that enmity *between the genders* here was such that the two never met at all?[49]

(6) character is unmixed with **gender**.[50]

(7) Felix Ungar (Jack Lemmon) is a casualty of the war *between the* **sexes**.[51]

(8) ethics and moral standards can be *combined with* **sex** information.[52]

This analysis confirms that *sex* in the 1960s was not limited to a binary man/woman distinction. Rather, in 1964, *sex* was broadly used to cover aspects of sex and sexuality that in 2019 may be represented by different terms, in particular *gender* and its variants.

---

[49] Calisher, *supra* note 36, at 155.

[50] *Id.* at 114.

[51] *The Odd Couple*, Time, Nov. 3, 1968.

[52] Joseph N. Bell, *Why the Revolt Against Sex Education?*, Good Housekeeping, Nov. 1969.

CONCLUSION

Amici curiae Professors Slocum, Gries, and Solan respectfully submit that this Court should reject arguments that the term "sex" as used in the 1960s was limited to a strictly binary or biological meaning or to heterosexual sex, and apply a corpus linguistics analysis to ascribe the broader meaning of the term "sex" that was actually applicable at that time.

July 3, 2019

Respectfully submitted,

ANDREW RHYS DAVIES
  Counsel of Record
JUSTIN ORMAND
REBECCA DELFINER
GIDEON DUKE-COHAN
CHANGHEE HAN
ALLEN & OVERY LLP
1221 Avenue of the Americas
New York, New York 10020
(212) 610-6300
andrewrhys.davies@allenovery.com

MELINDA BOTHE
DANNA SELIGMAN
ALLEN & OVERY LLP
1101 New York Avenue, NW
Washington, DC 20005
(202) 683-3800

Attorneys for Amici

BLANK PAGE