

No. 18-956

---

---

IN THE  
**Supreme Court of the United States**



GOOGLE LLC,

*Petitioner,*

*v.*

ORACLE AMERICA, INC.,

*Respondent.*

---

*On Writ of Certiorari to  
the United States Court of Appeals  
for the Federal Circuit*

---

---

**BRIEF OF *AMICI CURIAE*  
EMPIRICAL LEGAL RESEARCHERS  
IN SUPPORT OF PETITIONER**

---

---

I. Elie Herman  
*Counsel of Record*  
Patrick Geraghty  
Patrick J. Lawless  
WILSON ELSEER MOSKOWITZ  
EDELMAN & DICKER LLP  
150 East 42nd Street  
New York, New York 10017  
212-490-3000  
elie.herman@wilsonelser.com

*Counsel for Amici Curiae*

January 13, 2020

---

---

## TABLE OF CONTENTS

Table of contents.....	i
Table of authorities.....	ii
Statement of interest.....	1
Introduction and summary of argument.....	2
Argument.....	3
A. The declarations copied by Google are analogous to words and short phrases.....	5
B. Allowing a copyright in these words and phrases would create an unreasonable constraint on innovation and interoperability.....	8
Conclusion.....	13
Appendix.....	A1

## TABLE OF AUTHORITIES

### Cases

<i>Hibbs v. Winn</i> , 542 U.S. 88, 101 (2004).....	4
<i>Loomis v. Wisconsin</i> , 881 N.W.2d 749 (Wis. 2016).....	4
<i>McCutcheon v. FEC</i> , 572 U.S. 185 (2014) .....	8
<i>Oracle America, Inc. v. Google Inc.</i> , 750 F.3d 1339, 1349 (Fed.Cir. 2014) .....	6
<i>Southco, Inc. v. Kanebridge Corp.</i> , 390 F.3d 276, 285 (3d Cir. 2004).....	13

### Other Authorities

BJARNE STROUSTRUP, <i>THE C++ PROGRAMMING LANGUAGE</i> 463, 586 (2000) .....	7
BRIAN W. KERNIGHAN & DENNIS M. RITCHIE, <i>THE C PROGRAMMING LANGUAGE</i> 136 (2d ed. 1988).....	6
DONN FELKER, <i>ANDROID APPLICATION DEVELOPMENT FOR DUMMIES 2</i> (2011).....	4
JAMES GOSLING ET AL., <i>THE JAVA PROGRAMMING LANGUAGE</i> 531-32 (4th ed. 2005).....	6, 7
JAMES GOSLING, ET AL., <i>THE JAVA LANGUAGE SPECIFICATION</i> 20 (3d ed. 2005) .....	4, 6
James Gosling, <i>The Feel of Java</i> , 30 <i>COMPUTER</i> 53 (June 1997).....	7
Nathaniel Popper, <i>The Robots Are Coming for Wall Street</i> , <i>N.Y. TIMES MAGAZINE</i> , Feb. 28, 2016 at 56.....	4

PETER SESTOFT, PROGRAMMING LANGUAGE CONCEPTS 8 (2d ed. 2017) .....	7
Robert M. Cover, <i>Violence and the Word</i> , 95 YALE L. J. 1601 (1986) .....	4

IN THE  
**Supreme Court of the United States**

---

No. 18-956

GOOGLE LLC,

*Petitioner,*

v.

ORACLE AMERICA, INC.,

*Respondent.*

---

On Writ of Certiorari  
to the United States Court of Appeals,  
for the Federal Circuit

---

**BRIEF OF AMICI CURIAE EMPIRICAL LEGAL  
RESEARCHERS IN SUPPORT OF PETITIONER**

---

**STATEMENT OF INTEREST<sup>1</sup>**

The amici are attorneys and empirical legal researchers who frequently write computer software in order to conduct research. James E. Daily is Head of

---

<sup>1</sup> Petitioner has filed a blanket consent letter with the Court. Respondent has consented to the filing of this brief. No counsel for any party authored this brief in whole or in part, and no person or entity other than the amicus, its members, or its counsel made a monetary contribution intended to fund the brief's preparation or submission.

Legal Data Science at Skopos Labs, Inc. Adam Feldman is the principal of Optimized Legal Solutions and author of the Empirical SCOTUS blog.

The amici have no personal interest in the outcome of the case. As software developers they have a general professional interest in an outcome that favors innovation, open interfaces, and interoperability between computer programs. The amici have previously written programs using the languages and platforms at issue in the case, but they do not do so currently.

### **INTRODUCTION AND SUMMARY OF ARGUMENT**

The facts of this case are unusually technical and threaten to obscure the legal issues. As attorneys who are also software developers, the amici offer an analogy between the computer languages, with which the Court may be unfamiliar, and a kind of language with which the Court is uniquely familiar: the text of Supreme Court opinions.

Authors rely on the availability of language to express ideas. Limiting the use of language can lead to harmful repercussions for both authors and for the public: authors in what they can write, and the public in what it can consume. Limiting language is especially problematic when dealing with languages and subjects where even seemingly minor nuances can be critically important. Computer languages and the language of judicial opinions are two prime examples.

While computer languages are not used for the same purpose as regular prose, both lose a great deal of functionality when language choice is restricted. This is especially true when authors are precluded

from using simple, frequently used terms. To show the importance of freely available language, this brief presents an analogy from computer language to Supreme Court opinions. Both rely on precise terminology and phrases, rely on reusing and building upon language used by others, and have significant effects on the lives and livelihoods of millions of people.

In these contexts, small changes in language can greatly alter meanings. Software developers and Justices alike would be hamstrung if they were limited to a universe of language based on words and phrases not used in the past. Requiring every computer programmer or Justice to reinvent common phrases and concepts would make new works difficult to write and confusing to read. Likewise, it would prevent the cumulative development of new ideas and concepts that is vital to innovation in computer science, just as it is in the law.

This brief illustrates the point through the results of experiments showing what would happen if Supreme Court Justices were foreclosed from using language from prior opinions written by others. This demonstrates the hardships that software developers would face if Oracle were allowed to claim a copyright in the Java declarations at issue in this case.

### **ARGUMENT**

Judicial opinions and computer programs are both written using highly specific, exacting language in which each word must be carefully considered. A judge's choice or interpretation of a single word can mean the difference between a person gaining or los-

ing their freedom or property.<sup>2</sup> In our highly automated society, a computer programmer’s word choice can have similarly consequential effects.<sup>3</sup> In both of these contexts, even ordinarily synonymous words must be distinguished, and every word should be considered meaningful.<sup>4</sup>

Of course, many computer programs are unimportant or even frivolous. But just as a judge must give both important and unimportant cases due consideration, so too are the same strict rules applied to every computer program.

Allowing computer language creators, like Oracle, to claim a copyright in the sort of simple and funda-

---

<sup>2</sup> Robert M. Cover, *Violence and the Word*, 95 YALE L. J. 1601 (1986).

<sup>3</sup> *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016) (appealing, *inter alia*, the sentencing recommendations of the automated COMPAS risk assessment tool), *cert. denied*, 137 S.Ct. 2290 (2017); Nathaniel Popper, *The Robots Are Coming for Wall Street*, N.Y. TIMES MAGAZINE, Feb. 28, 2016 at 56 (“Decisions about loans are now being made by software that can take into account a variety of finely parsed data about a borrower”).

<sup>4</sup> *See, e.g., Hibbs v. Winn*, 542 U.S. 88, 101 (2004) (“A statute should be construed so that effect is given to all its provisions, so that no part will be inoperative or superfluous, void or insignificant”). In Java and Android, as in many programming languages, even uppercase and lowercase letters are distinguished. For example, two methods named *max()* and *Max()* would be regarded as completely different by the computer. JAMES GOSLING, ET AL., THE JAVA LANGUAGE SPECIFICATION 20 (3d ed. 2005); DONN FELKER, ANDROID APPLICATION DEVELOPMENT FOR DUMMIES 2 (2011).



mental computer code declarations at issue would unduly restrict both innovation and interoperability between computer programs. The amici demonstrate this by analogizing the declarations at issue to words from prior Supreme Court opinions. The analogy shows that, to avoid an undue restriction, the bar for copyrightability—or the shelter of fair use—must be extended further than Oracle would allow.

To be clear at the outset, judicial opinions are, of course, not copyrightable, and the amici are not claiming that a decision favoring Oracle would lead to such a conclusion. Rather, the analogy is a way for those less versed in computer programming to understand how a decision favoring Oracle would affect the development of new and innovative computer languages, platforms, and software.

**A. The declarations copied by Google are analogous to words and short phrases.**

The declarations that Google copied are not, by themselves, valid Java programs. In English they are analogous to a word or short phrase, such as “length” or “square root”. These are not complete thoughts, nor are they grammatical sentences. Even if one added enough code to turn these declarations into the bare minimum of a valid Java program, that program would not accomplish anything. It would be the equivalent of turning the phrase “square root” into the sentence “Square root is.” Technically grammatical, but devoid of substance.

These bare declarations can be given substance with implementation code, which is essentially a definition of what the word or phrase actually means, in

terms that a computer can understand. But Google did not copy Oracle’s implementation code, only the declarations.

In most computer languages, including Java, developers are free to write their own new methods rather than using existing ones. But the Java standard library, including the methods that Google copied, amounts to a “fixed vocabulary” of fundamental methods and concepts. As Oracle has conceded to some extent, many of these methods must be used “in order to make any worthwhile use of the language.”<sup>5</sup>

This fixed vocabulary includes many methods that predate Java and were included in older languages that Oracle acknowledges influenced Java.<sup>6</sup> The amici

---

<sup>5</sup> *Oracle America, Inc. v. Google Inc.*, 750 F.3d 1339, 1349 (Fed.Cir. 2014) (“The district court found, and Oracle concedes to some extent, that three of those packages—`java.lang`, `java.io`, and `java.util`—were ‘core’ packages, meaning that programmers using the Java language had to use them ‘in order to make any worthwhile use of the language.’”)

<sup>6</sup> The design of the Java language was influenced by the languages C and C++, among others. JAMES GOSLING, ET AL., THE JAVA LANGUAGE SPECIFICATION 1 (“The Java programming language is related to C and C++”). The C and C++ standard libraries have equivalents of many of the methods Oracle claims a copyright in. For example, the C and Java standard math libraries (called *math* in C and *Math* in Java) contain *sin()*, *cos()*, *tan()*, *log()*, and *sqrt()*, among many other common functions. BRIAN W. KERNIGHAN & DENNIS M. RITCHIE, THE C PROGRAMMING LANGUAGE 136 (2d ed. 1988); JAMES GOSLING ET AL., THE JAVA PROGRAMMING LANGUAGE 531-32 (4th ed. 2005). Similarly, C++ and Java both say that a string has a *length()* but a list has a *size()*. BJARNE STROUSTRUP, THE C++ PROGRAMMING LANGUAGE 463,

have found no evidence that Oracle took a license for this use.

This free borrowing of useful concepts and standard tools from other languages is common in computer science.<sup>7</sup> James Gosling, the chief designer of the Java language, acknowledged being influenced by at least seven predecessor languages.<sup>8</sup> These influences, Gosling notes, helped Java “feel[] very familiar to many different programmers.”<sup>9</sup> In this way, both judges and computer programmers rely heavily on terms previously defined by others, be it “compelling governmental interest” or *sqrt()*.

Given these similarities between computer language and legal language, one might imagine what it would be like if judicial opinions were as constrained in their use of language as computer programs would

---

586 (2000); GOSLING ET AL., *THE JAVA PROGRAMMING LANGUAGE* 51, 469.

<sup>7</sup> PETER SESTOFT, *PROGRAMMING LANGUAGE CONCEPTS* 8 (2d ed. 2017) (“Most new programming languages arise as a reaction to some language that the designer knows (and likes or dislikes) already, so one can propose a family tree or genealogy for programming languages, just as for living organisms.”). Sestoft provides a family tree demonstrating Java’s numerous influences, as well as several languages influenced by Java in turn.

<sup>8</sup> James Gosling, *The Feel of Java*, 30 *COMPUTER* 53 (June 1997) (“[Java] has an object-oriented flavor that derives from a number of languages—Simula, C/C++, Objective C, Cedar/Mesa, Modula, and Smalltalk.”).

<sup>9</sup> *Id.*

be if Oracle could claim copyright in the declarations at issue.

**B. Allowing a copyright in these words and phrases would create an unreasonable constraint on innovation and interoperability.**

For this experiment, the amici analyzed 15,942 opinions issued in 7,113 cases from 1946 to 2014. They broke the text of each opinion into individual words and two- and three-word phrases. For each of the 131 opinions issued in 2014, the amici determined the number of words, phrases, and sentences that had an antecedent in at least one opinion from a prior term, excluding opinions written by the same Justice.<sup>10</sup> Figure 1 shows an example of this analysis applied to Chief Justice Roberts's opinion in *McCutcheon v. FEC*.<sup>11</sup>

---

<sup>10</sup> For a more complete description of our methodology, see the Appendix, *infra*.

<sup>11</sup> *McCutcheon v. FEC*, 572 U.S. 185 (2014).



But perhaps individual words are unfairly reductive. After all, as Oracle points out, some of the declarations in Java are complex.<sup>12</sup> Considering three-word phrases greatly expands the potential creative universe.

electing Citizens  
 : They , vote,  
 , volunteer  
 s campaign.  
  
 The democracy through  
 .  
 . 424 U. S. 1–27 (1976) (per curiam).  
 ,  
 simply  
 Free Enterprise Club's Freedom Club PAC v. Bennett, 564  
 U. S. , (2011) (slip op., at 24–25).  
  
 Many latter objectives attractive:  
 see fewer television commercials  
 s accomplishments or disparaging an opponent's character.  
 Money in politics repugnant  
 vigorously protects. If  
 flag burning, funeral protests, and Nazi parades—  
 despite the profound offense such spectacles cause—  
 despite popular opposition.  
 491 U. S. 397 (1989); , 562 U. S. (2011);  
 , 432 U. S. 43 (1977) (per curiam).  
 ,  
 " "  
 , 401 U. S. 265, 272 (1971).

*Figure 2. McCutcheon's introduction after removing all three-word phrases found in prior opinions not written by the same author.*

This is an improvement, but the text is still unreadable as prose and unusable as a legal opinion. On average, 64.6% of these three-word phrases can be found in prior opinions.

Of course, many of these phrases predate the beginning of our dataset in 1946, just as equivalents of

---

<sup>12</sup> Oracle BIO 6.

the declarations that Oracle claims a copyright over can be found in languages that predate Java and that Oracle acknowledges influenced Java’s design.<sup>13</sup> Even allowing free use of any phrase that occurs in the first five years of opinions—on the theory that these phrases are likely to be “public domain”—the results are still striking.

There is [REDACTED] basic in our [REDACTED] right to participate in electing [REDACTED]. Citizens [REDACTED] in a variety of ways: They [REDACTED], vote, urge others to vote for [REDACTED], volunteer to work [REDACTED], [REDACTED] [REDACTED]’s campaign. This case [REDACTED] [REDACTED].

The right to participate in democracy through [REDACTED] [REDACTED] by the [REDACTED] absolute. [REDACTED] held that Congress may [REDACTED] [REDACTED] protect against [REDACTED], 424 U. S. 1–27 (1976) (per curiam). At the same [REDACTED] made clear that Congress may [REDACTED] simply to reduce the amount of [REDACTED], or to restrict the [REDACTED] [REDACTED] Free Enterprise Club’s Freedom Club PAC v. Bennett, 564 U. S. \_\_\_, \_\_\_ (2011) (slip op., at 24–25).

Many [REDACTED] latter objectives attractive: They would be delighted to see fewer television commercials [REDACTED] [REDACTED]’s accomplishments or disparaging an opponent’s character. Money in politics may [REDACTED] repugnant [REDACTED] [REDACTED] Amendment vigorously protects. [REDACTED] Amendment protects flag burning, funeral protests, and Nazi parades—despite the profound offense such spectacles cause—[REDACTED] [REDACTED] despite popular opposition. See [REDACTED], 491 U. S. 397 (1989); [REDACTED], 562 U. S. \_\_\_ (2011); [REDACTED] [REDACTED], 432 U. S. 43 (1977) (per curiam). Indeed, as we have emphasized, the First Amendment “ [REDACTED] the [REDACTED] office.” [REDACTED], 401 U. S. 265, 272 (1971).

*Figure 3. McCutcheon’s introduction after removing all three-word phrases found in prior opinions not written by the same author, excluding phrases found in the first five years of our data (i.e. 1946 – 1951).*

Under this generous extension of the analogy, 32.3% of three-word phrases are redacted on average, and the text remains somewhere between unintelligible and unusably ambiguous.

---

<sup>13</sup> See n. 5, *supra*.

Similar results are obtained under several other kinds of analysis, including:

- Redacting majority opinions using only non-majority opinions, and vice versa (56.9% redaction of three-word phrases)
- Redacting opinions authored by one Justice using only opinions written by a single other Justice (45% redaction of three-word phrases)
- Redacting opinions that reference mobile phones, cell phones, or smart phones using only those that do not, and vice versa (59.4% redaction of three-word phrases)

A more complete description of these additional analyses may be found in the Appendix.

Finally, one might imagine how far one must go before the redaction no longer overwhelms the text or prevents the use of important terms of art. While, as discussed above, the Java declarations at issue in this case are not the equivalent of sentences, one could consider redacting only entire sentences that are found verbatim in a prior opinion. Under that strict rule, only 4.7% of sentences are redacted on average and none in the *McCutcheon* excerpt.

So, at last one finds the other side of the line: complete sentences, not words or short phrases. Not coincidentally, that begins to touch upon the minimum level of creative expression that has been held to be



covered by copyright.<sup>14</sup> Unreasonable restriction on judicial language can be avoided only by restricting the analysis to the point that it no longer covers the equivalent of what Oracle is claiming. In the same way, Oracle's claim to copyright in Java declarations would prevent Google and other software developers from doing precisely what Oracle itself did: creating innovative languages and platforms that build upon and interoperate with the work of others.

### CONCLUSION

Allowing copyright in Java declarations is tantamount to allowing copyright in words or short phrases. This would unreasonably restrict the development of new programming languages and platforms, and prevent the cumulative work that has been the hallmark of innovation in computer science—including Java itself. The Court should overturn the Federal Circuit's decision in this case and hold that Oracle cannot claim copyright in Java declarations.

---

<sup>14</sup> *Southco, Inc. v. Kanebridge Corp.*, 390 F.3d 276, 285 (3d Cir. 2004) (*en banc*) (excluding part numbers from copyright protection because they are analogous to short phrases).

14

Respectfully submitted,

I. Elie Herman

*Counsel of Record*

Patrick Geraghty

Patrick J. Lawless

WILSON ELSER MOSKOWITZ

EDELMAN & DICKER, LLP

150 East 42nd Street

New York, NY 10017

(212) 490-3000

elie.herman@wilsonelser.com

*Counsel for Amici Curiae*

January 13, 20&S

## **APPENDIX**

## APPENDIX

The data for the analysis consisted of the text of 15,942 opinions issued in 7,113 cases from 1946 to 2014, labeled by LEXIS citation, author, and opinion type (i.e. majority, concurrence, dissent). The dataset was created by downloading the case text and parsing it into separate opinions. The full dataset is available from the amici on request.

The analysis software was written in the Python programming language using two standard open source libraries for natural language processing and data analysis.<sup>15</sup> Full source code is available online.<sup>16</sup>

The opinion text was first preprocessed to remove extraneous markup such as page numbers and “Footnote” / “End Footnote” markers. Then the text was parsed into sentences and each sentence into one, two, and three-word phrases. Sentences were skipped if they appeared to be direct quotations (i.e. started and ended with double quotation marks).

A word was defined as a token (i.e. unit of text) that was not a number, symbol, punctuation, space, or unrecognized token. Two and three-word phrases were defined as two or three tokens in a row, optionally separated by commas. This was intended to avoid phrases

---

<sup>15</sup> Python Software Foundation, <https://www.python.org/> (last visited Jan. 8, 2020); Explosion AI, <https://spacy.io/> (last visited Jan. 8, 2020); The Pandas Project, <https://pandas.pydata.org/> (last visited Jan. 8, 2020).

<sup>16</sup> James E. Daily, *Phrase Reuse*, <https://github.com/james-daily/phrase-reuse> (last visited Jan. 9, 2020).

that crossed semicolons, parentheses, and other boundary punctuation.

The analysis consisted of selecting opinions (generally all opinions from the most recent year in the data) and then comparing their constituent phrases to those used in opinions from earlier years and written by different authors. The “antecedent fraction” of an opinion was calculated as the number of unique phrases of a given length found in prior opinions divided by the number of unique phrases in the opinion. The “modern antecedent fraction” is similar, except that each phrase is only compared against phrases that do not exist in the first five years’ worth of opinions.

Table 1 shows a sample of the results of this analysis. Full results are available from the amici on request.

## A3

	1 Word	2 Words	3 Words	Sentence
Prior Opinions	99.0%	89.0%	64.6%	5.4%
Exclude 1st 5 Years <sup>17</sup>	5.9%	23.1%	32.3%	4.7%
Single Author <sup>18</sup>	97.3%	77.1%	45.0%	3.0%
By Opinion Type <sup>19</sup>	98.6%	85.4%	56.9%	3.1%
By Topic <sup>20</sup>	98.7%	86.6%	59.4%	3.6%

*Table 1. Average number of words, phrases, or sentences redacted across a variety of analyses.*

The redactions were produced by treating short excerpts of an opinion to the same analysis process described above. Each phrase found to have an antecedent was then wrapped in an HTML `<b>` tag and CSS was used to produce the redaction effect.

---

<sup>17</sup> Opinions were redacted using only words and phrases that do not appear in the first five years of data. This shows the opposite trend of the other analyses because most individual words are used within the first five years, whereas any given combination of words is less likely to have been used because there are so many more possible combinations. Despite this trend, however, the percentages of two and three-word phrases redacted are still much lower under this analysis than under the less constrained basic analysis.

<sup>18</sup> Opinions were redacted using only opinions written by a single author, John Paul Stevens. Stevens was chosen because he is the most prolific author in the dataset and was not on the Court in 2014.

<sup>19</sup> Majority opinions were redacted using non-majority opinions written by other authors and vice versa.

<sup>20</sup> Opinions were divided into those that mention cell phones, mobile phones, or smart phones and those that do not. Each group was used to redact the other, subject to the usual constraint of only using opinions written by different authors.

Amici distinguished between lowercase and uppercase versions of words. There are many occasions in English where two words are distinguished by case alone. For example, “Apple” the computer company is a very different word than “apple” the fruit. Amici considered these to be different words in order to keep the comparison to the case-sensitive Java language as fair as possible.