STATUTORY INTERPRETATION — INTERPRETATIVE TOOLS — UTAH SUPREME COURT DEBATES JUDICIAL USE OF CORPUS LINGUISTICS. — *State v. Rasabout*, 356 P.3d 1258 (Utah 2015).

Courts have long sought to find the ordinary meaning of words and phrases in statutes,[1] enlisting a variety of tools, such as dictionaries,[2] canons of interpretation,[3] and the common sense of an English-language speaker.[4] Recently, in *State v. Rasabout*,[5] the Supreme Court of Utah considered a novel tool for statutory interpretation: corpus linguistics, "the study of language based on examples of 'real life' language use."[6] In recent decades, linguistic programs at universities and institutes have assembled *corpora* (or bodies) of language — vast computer databases cataloguing written and spoken language.[7] These databases can easily be searched to retrieve examples of how words or phrases have been used in different contexts at different times.[8] In *Rasabout*, the majority and concurrence debated the legitimacy of using this linguistic tool. The court unanimously held that the phrase "unlawful discharge of a firearm" in a criminal statute referred to each individual shot fired.[9] To arrive at this decision, the majority relied upon traditional tools of statutory interpretation.[10] But a concurring justice found these tools wanting and informed his judgment by searching for the word *discharge* in contemporary news articles and a linguistics database.[11] The majority argued that this research was inappropriate largely because corpus linguistics is an unfamiliar, scientific tool and its proper use requires an expertise judges lack.[12] Corpus linguistics is indeed novel. But in service of the traditional task of considering how a word is commonly used, jurists are capable of searching an online database for examples of how a word has been used. By providing externally generated examples, corpus linguistics can be a helpful double check against a judge's intuitive understanding of a word or phrase.

---

1 *See, e.g.*, Nix v. Hedden, 149 U.S. 304, 306 (1893).

2 *See* Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L.J. 275, 277 (1998).

3 *See, e.g.*, South Carolina v. Catawba Indian Tribe, Inc., 476 U.S. 498, 509–10 & n.22 (1986).

4 *See, e.g.*, Smith v. United States, 508 U.S. 223, 241–43 (1993) (Scalia, J., dissenting).

5 356 P.3d 1258 (Utah 2015).

6 TONY MCENERY & ANDREW WILSON, CORPUS LINGUISTICS 1 (2d ed. 2001).

7 *Id.*

8 *See, e.g.*, *Introduction*, CORPUS OF CONTEMP. AM. ENG., http://corpus.byu.edu/coca (last visited Feb. 2, 2016).

9 *Rasabout*, 356 P.3d at 1260–61.

10 *Id.* at 1262–65.

11 *Id.* at 1278–79, 1281–82 (Lee, A.C.J., concurring in part and concurring in the judgment).

12 *Id.* at 1264–66 (majority opinion).

Eight years ago, Andy Rasabout fired twelve shots from a car into the home of a rival gang member.[13] A jury convicted him of twelve separate counts of unlawful discharge of a firearm under a Utah statute that makes it illegal to "discharge any kind of dangerous weapon or firearm . . . from an automobile . . . ; from, upon, or across any highway; . . . or . . . within 600 feet of . . . a house."[14] Because the shots were part of a single criminal episode, before sentencing the trial court merged the twelve separate counts into one conviction.[15]

The Utah Court of Appeals reversed, holding that Rasabout must be convicted and sentenced for each discrete shot fired.[16] The court of appeals examined the text of the criminal statute to determine if *discharge* meant that the legislature intended a separate conviction for each shot fired or one conviction for the whole episode.[17] Looking to dictionary definitions, the court determined that, in this context, discharge meant to "fire a weapon"[18] or to "shoot."[19]

The Utah Supreme Court granted certiorari and unanimously affirmed the court of appeals's decision, finding that *discharge*, in the context of a "dangerous weapon or firearm,"[20] referred to "each discrete shot" and, as such, each of Rasabout's twelve discrete shots constituted a criminal violation.[21] To arrive at this conclusion, Justice Parrish, writing for the majority,[22] looked to the structure of the word *discharge*, the dictionary definition of *discharge*, the accompanying language in the statute, and common sense.[23] After observing that the root of the word *discharge* — *charge* — has noun and verb meanings related to the amount of gunpowder used in a single shot and that the dictionary definition of *discharge* included the meaning "to shoot," Justice Parrish concluded that "the clearest reading of the statute" is that *discharge* refers to each shot.[24]

---

[13] *Id.* at 1261.

[14] UTAH CODE ANN. § 76-10-508 (West 2007).

[15] State v. Rasabout, 299 P.3d 625, 627 (Utah Ct. App. 2013).

[16] *Id.* at 635.

[17] *Id.* at 631–32.

[18] *Id.* at 632 (quoting *Discharge*, MACMILLAN DICTIONARY, http://www.macmillandictionary.com/dictionary/american/discharge [http://perma.cc/X69V-LDQP]).

[19] *Id.* (quoting *Discharge*, MERRIAM-WEBSTER, http://www.merriam-webster.com/dictionary/discharge [http://perma.cc/3QCX-LP2T]).

[20] *Rasabout*, 356 P.3d at 1263 (quoting UTAH CODE ANN. § 76-10-508(1)(a) (West 2007)).

[21] *Id.* at 1261.

[22] Justice Durham and Judge Harris joined the majority opinion in full. Chief Justice Durrant and Associate Chief Justice Lee each wrote an opinion joining the majority in part.

[23] *See Rasabout*, 356 P.3d at 1263–64.

[24] *Id.* at 1263. Elsewhere in the statute, firearms and pistols were defined by their ability to fire a single shot, UTAH CODE ANN. § 76-10-501(9), (12), which confirmed for the majority that *discharge* also referred to a single shot, *Rasabout*, 356 P.3d at 1264. Finally, the majority decided that "it was reasonable for the Legislature to criminalize each shot fired because each shot carries an independent harm." *Id.*

Next, the majority admonished Associate Chief Justice Lee's concurrence for its reliance on corpus linguistics as a tool for statutory interpretation,[25] contending that his research was unfair to the parties because the rationale did not appear in the parties' arguments[26] and because judges should not decide cases by conducting their own "independent scientific research."[27] Justice Parrish contended that judges lack the expertise to conduct this research because "[l]inguistics is a scientific field of study that uses empirical research to draw findings," and judges are generalists, not scientists.[28] To illustrate this problem, Justice Parrish pointed out that professional linguistic studies published in reliable journals are subject to the rigors of peer review to ensure that the findings are reliable; in comparison, court judgments lack this systemic oversight.[29] Lastly, Justice Parrish criticized the concurrence's methodology by questioning the statistical significance of the findings and claiming that the more appropriate data set of language to analyze would have been the text of the Utah Code.[30]

Associate Chief Justice Lee concurred in part and concurred in the judgment.[31] He diverged from the majority because he did not find that dictionaries fully resolved the meaning of *discharge*.[32] While one definition, "shoot," would confine the meaning of *discharge* to each shot fired,[33] another definition, "empty of a cargo: UNLOAD," could include "unloading or emptying of the contents of a weapon."[34] Under Utah precedent, when neither of two meanings can be eliminated, the court opts for the more ordinary meaning.[35] Associate Chief Justice

---

[25] *Rasabout*, 356 P.3d at 1264–66.

[26] *Id.* at 1264–65.

[27] *Id.* at 1265.

[28] *Id.*

[29] *Id.* at 1265–66. The majority also argued that requiring district judges and litigants to conduct "ad hoc linguistics research" would overwhelm the court system and close its doors "for all but the most affluent." *Id.* at 1265.

[30] *Id.* at 1266. The majority concluded by affirming the court of appeals's dismissal of ancillary claims. Having found the plain meaning of "discharge," the majority also held that the rule of lenity did not apply because the statute was not ambiguous. *Id.* Lastly, Utah's single-criminal-episode statute did not apply because that statute does not "dictate the merger of offenses," *id.* at 1268, and the state's single-larceny rule did not apply because larceny was not a part of the case, *id.* at 1268–69.

[31] *Id.* at 1271–90 (Lee, A.C.J., concurring in part and concurring in the judgment). Chief Justice Durrant also wrote a concurring opinion. He found that the dictionary definitions supplied the ordinary meaning in this case, but he tentatively endorsed exploring the use of corpus linguistics in the future. *Id.* at 1269–70 (Durrant, C.J., concurring in part and concurring in the judgment).

[32] *Id.* at 1271 (Lee, A.C.J., concurring in part and concurring in the judgment).

[33] *Id.* at 1273 (quoting *Discharge*, MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 356 (11th ed. 2012)).

[34] *Id.* (quoting *Discharge*, WEBSTER'S THIRD INTERNATIONAL DICTIONARY 644 (3d ed. 2002)).

[35] *Id.* at 1272 (citing Hi-Country Prop. Rights Grp. v. Emmer, 304 P.3d 851, 855–56 (Utah 2013)).

Lee felt that *discharge* ordinarily refers to firing a single shot and not to unloading a firearm,[36] but rather than rely on intuition alone to choose one meaning over another, he turned to corpus linguistics.[37]

Associate Chief Justice Lee began by defending judicial use of corpus linguistics.[38] He argued that judges already use an introspective version of corpus linguistics to interpret statutes.[39] By searching their memories for how they have heard words or phrases used, judges are comparing the statute's language to a corpus of language in their minds.[40] The justice further argued that dictionaries themselves are "compiled from broader linguistic corpora."[41] And he noted cases where judges — including Justice Breyer[42] and Judge Posner[43] — have informed their interpretation of a statute by using search engines to acquire examples of the statutory language in question.[44]

To clarify the meaning of "discharge," Associate Chief Justice Lee performed a Google News search and a search of the Corpus of Contemporary American English (COCA).[45] He considered Google News a reliable source because published newspaper articles contain a wealth of natural language and the search engine allows the judge to search for phrases — a task that cannot be performed with a dictionary.[46] Associate Chief Justice Lee's Google News search of "discharge of a firearm" yielded favorable results for the majority's preferred definition.[47] While some articles were unclear, most articles used discharge to indicate the firing of a single shot and none used discharge to indicate the unloading of an entire gun.[48] But Associate Chief Justice Lee also acknowledged the deficiencies of a Google News search.[49] The algorithm is not transparent, and the results may be particularized for an individual user.[50] To avoid these defects, Associate Chief Justice Lee turned to COCA, which is free, is accessible on the Internet, "contains more than 520 million words of text[,] and is equally divided among spoken, fiction,

---

[36] *Id.* at 1275.
[37] *Id.*
[38] *Id.* at 1276–77.
[39] *Id.* at 1276.
[40] *Id.*
[41] *Id.*
[42] Muscarello v. United States, 524 U.S. 125, 129 (1998).
[43] United States v. Costello, 666 F.3d 1040, 1044 (7th Cir. 2012).
[44] *Rasabout*, 356 P.3d at 1277 (Lee, A.C.J., concurring in part and concurring in the judgment). Associate Chief Justice Lee has also performed corpus analysis in a few prior cases. *See* State v. Canton, 308 P.3d 517, 523 & n.6 (Utah 2013); Carranza v. United States, 267 P.3d 912, 918 (Utah 2011) (opinion of Lee, J.).
[45] *Rasabout*, 356 P.3d at 1277 (Lee, A.C.J., concurring in part and concurring in the judgment).
[46] *Id.* at 1278.
[47] *Id.*
[48] *Id.*
[49] *Id.* at 1280.
[50] *Id.*

popular magazines, newspapers, and academic texts."[51] Associate Chief Justice Lee searched for the word *discharge* within five words of *firearm*, *firearms*, *gun*, or *weapon*.[52] The examples he found overwhelmingly used *discharge* in connection with a single shot.[53] Given these results, he agreed with the majority that *discharge* referred to each shot fired.[54]

Associate Chief Justice Lee closed by responding to many of Justice Parrish's criticisms. He argued that his research was an appropriate legal investigation into the meaning of a law — not an inappropriate factual investigation of evidence in the case.[55] While conceding that judges are not expert linguists, he argued that judges are experts in legal interpretation.[56] Because the judicial mandate is to say what the law is, judges should employ the best tools possible to judge with certainty.[57] Although judges may misuse scientific tools, he argued that the response to this risk should not be to abandon those tools, especially because traditional tools — intuition and the dictionary — "involve bad linguistics" and thus similarly risk incorrect determinations of how a word is commonly used.[58] Finally, Associate Chief Justice Lee claimed that corpus analysis "is not rocket science."[59] Analogously, judges are not historians but still attempt to unearth the historical meaning of words because the judicial role demands that they try.[60]

State v. Rasabout is unlikely to attract attention for its decision on the merits, but its lively debate about corpus linguistics may foretell future skirmishes over the legitimacy of interpreting statutes with the help of data-driven tools.[61] In *Rasabout*, the majority accused the concurrence of sua sponte scientific research beyond the judiciary's expertise.[62] The majority's charge reveals two concerns: first, that corpus linguistics is an unfamiliar tool for statutory interpretation and, second, that corpus linguistics demands a skill set that judges lack. Corpus linguistics is new. But as long as judges use it to provide a ready source of examples of a word or phrase in context — and not to

---

[51] *Introduction*, *supra* note 8.

[52] *Rasabout*, 356 P.3d at 1281–82 (Lee, A.C.J., concurring in part and concurring in the judgment).

[53] *Id.* at 1282.

[54] *Id.*

[55] *Id.* at 1284.

[56] *Id.* at 1285.

[57] *Id.* at 1286.

[58] *Id.* at 1285.

[59] *Id.* at 1286.

[60] *Id.* (citing ANTONIN SCALIA & BRYAN GARNER, READING LAW 400 (2012)).

[61] For an argument that corpus linguistics could and should supplant traditional tools of statutory interpretation, see Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*, 13 COLUM. SCI. & TECH. L. REV. 156 (2011).

[62] *Rasabout*, 356 P.3d at 1264.

provide conclusive empirical proof of statutory meaning — corpus linguistics can be a helpful double check against a judge's intuitive understanding of statutory language.

Associate Chief Justice Lee's research was not a foreign, scientific approach to statutory interpretation; he employed a new tool for a traditional task. Just as dictionaries once were,[63] corpus linguistics is a novel device for statutory interpretation. Clearly, linguistic databases are useful for many scientific endeavors. But the tool does not define the task. Both linguists and jurists work with language, and they can each use corpus linguistics differently to accomplish their different goals. Associate Chief Justice Lee's efforts were worlds removed from the peer-reviewed research that linguists conduct.[64] He adopted a rudimentary linguistic approach of identifying two possible meanings of a word and observing each meaning's frequency within a sample.[65] Required to decide the ordinary meaning of *discharge*, Associate Chief Justice Lee used a linguistics tool for the classic judicial undertaking of checking his initial understanding against real-world examples.[66]

Even if corpus linguistics can be used for a traditional judicial task, corpus linguistics can be an appropriate aid to statutory interpretation only if the judiciary can effectively use it. Judge Easterbrook reminds us: "Judges are overburdened generalists . . . . Methods of interpretation that would be good for experts are not suitable for generalists.

---

[63] *See* 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, 244–46 (1999) (chronicling the early history of dictionary use in the Supreme Court).

[64] The justice neither mentioned nor attempted to employ common linguistic and statistical methodologies, which include tests of statistical significance, proportional statistics, factor analysis, and multivariate tests, among others. *See* MCENERY & WILSON, *supra* note 6, at 81–98.

[65] *See id.* at 82.

[66] *See* Smith v. United States, 508 U.S. 223, 241–43 (1993) (Scalia, J., dissenting) (considering examples of how the word "use" is commonly used to determine its meaning in a criminal statute); Nix v. Hedden, 149 U.S. 304, 307 (1893) (looking to "the common language of the people" to determine whether a tomato is a fruit or a vegetable under a statute). As Associate Chief Justice Lee pointed out, this practice has been defended in the scholarly literature as well. *Rasabout*, 356 P.3d at 1272 (Lee, A.C.J., concurring in part and concurring in the judgment) ("[W]e assume that Congress uses common words in their popular meaning, as used in the common speech of men." (quoting Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527, 536 (1947))); *id.* ("We should look at the statutory structure and hear the words as they would sound in the mind of a skilled, objectively reasonable user of words." (quoting Frank H. Easterbrook, *The Role of Original Intent in Statutory Construction*, 11 HARV. J.L. & PUB. POL'Y 59, 65 (1988))). Use of corpus linguistics would conform to Professor Gary Lawson's understanding of how the judicial interpretative process *actually* functions: when judges describe a chosen interpretation as clear, they are making a probabilistic determination that most reasonable people would find that the word or phrase conforms to this understanding. *See* Gary Lawson, *Proving the Law*, 86 NW. U. L. REV. 859, 877–80 (1992).

Generalists should be modest and simple."[67] As Associate Chief Justice Lee argues, with digitized corpora on the Internet, corpus linguistics now makes modest and simple demands of a jurist, requiring an effort and expertise similar to that required by other search engines.[68] After filling out a few text boxes and clicking "search," a judge is provided with real-world examples of statutory language more quickly than she could imagine such examples on her own.[69] Corpus linguistics need not "require expertise in fields in which [judges] have no training,"[70] but can be used as a generalist's assistant.

In addressing the judiciary's lack of expertise, the majority raised a valid concern that the judiciary could inexpertly use corpus linguistics to produce misleading results about a word's meaning. Setting aside the majority and concurrence's back-and-forth over the specter of statistical significance,[71] common sense alone tells us that quick Internet searches can lead to distorted results. For example, if Associate Chief Justice Lee had instead searched for *discharge* in proximity to *Glock*, *pistol*, and *magazine*, the resulting examples would have been different and Associate Chief Justice Lee could have drawn a different inference.[72] Associate Chief Justice Lee took a quick look at a sample of real-world examples. These examples do not exhaustively or definitively represent how *discharge* is commonly used. The majority's critique rightly shows that Associate Chief Justice Lee's research was a glimpse at the terrain rather than a full geological survey.[73] In light of the risk of an unrepresentative sample, Associate Chief Justice Lee's counting also invites the criticism that a precise analysis of a rough list is falsely precise.[74] Because *discharge*, in the COCA examples, was "almost always used in the sense of a single shot,"[75] we can conclude that *dis-*

---

[67] Frank H. Easterbrook, *Text, History, and Structure in Statutory Interpretation*, 17 HARV. J.L. & PUB. POL'Y 61, 69 (1994) (footnote omitted); *cf.* RICHARD A. POSNER, THE FEDERAL COURTS 147–60 (1985).

[68] *Rasabout*, 356 P.3d at 1276–78 (Lee, A.C.J., concurring in part and concurring in the judgment).

[69] For example, the Brigham Young *Corpus of Contemporary American English* is free, accessible through a web browser, and functions as a search engine. *Introduction*, *supra* note 8.

[70] *Rasabout*, 356 P.3d at 1284 (Lee, A.C.J., concurring in part and concurring in the judgment) (quoting *id.* at 1265 (majority opinion)).

[71] Significance testing would not resolve the conflict here. Tests of statistical significance are quantitative attempts to measure the likelihood that an observed relationship within a sample was due to error and does not reflect the broader population from which the sample was drawn. *See* MCENERY & WILSON, *supra* note 6, at 78–81, 84–85; James P. Shaver, *What Statistical Significance Testing Is, and What It Is Not*, 61 J. EXPERIMENTAL EDUC. 293 (1993).

[72] The results of conducting this particular search on COCA, however, do not persuasively suggest an alternate meaning of *discharge*.

[73] *See Rasabout*, 356 P.3d at 1278–79, 1281–82 (Lee, A.C.J., concurring in part and concurring in the judgment).

[74] *Id.* at 1266 (majority opinion).

[75] *Id.* at 1277 (Lee, A.C.J., concurring in part and concurring in the judgment).

charge has been used in this sense. But because a quick database search may have missed examples of *discharge* used in a different sense, we cannot infer that the word *discharge* is confined to this meaning.

Given this risk, if judges use corpus linguistics for statutory interpretation, they should employ it not as a conclusive method for determining meaning but rather as a safety net to catch what intuition and the dictionary might miss. Perhaps the best example of the benefits of testing intuition with a quick corpus search comes from a dialogue with an opponent of corpus linguistics, Professor Noam Chomsky:

> Chomsky: The verb *perform* cannot be used with mass word objects: one can *perform a task* but one cannot *perform labour*.
>
> Hatcher: How do you know, if you don't use a corpus and have not studied the verb *perform*?
>
> Chomsky: How do I know? Because I am a native speaker of the English language.[76]

But Chomsky was wrong. A quick corpus search reveals that *perform* can be used with mass word objects — "[o]ne can *perform magic*, for example."[77] Using corpus linguistics as a complement to traditional tools of interpretation can temper this risk of deficient interpretations of a word or phrase. The majority correctly warned that a sample from a corpus may not accurately represent the entire English language,[78] and the concurrence rightly warned that our intuitive understandings of language may likewise be prejudiced or incomplete.[79] But each tool bears its flaws independent of the other. The judicial use of corpus linguistics, while not dispositive of meaning,[80] can reveal what lurks in the blind spots of traditional tools of interpretation.

State v. Rasabout could be a bellwether case. Despite the majority's admonishment that the concurrence relied "on scientific research that is not subject to scientific review,"[81] as an accessible, nontechnocratic check on traditional methods, corpus linguistics may well belong in judges' statutory interpretation toolkits.

---

[76] MCENERY & WILSON, *supra* note 6, at 11 (citing THE TEXAS CONFERENCE ON PROBLEMS OF LINGUISTIC ANALYSIS IN ENGLISH 29 (Archibald A. Hill ed., 1962)).

[77] *Id.*

[78] *Rasabout*, 356 P.3d at 1266.

[79] *Id.* at 1283 (Lee, A.C.J., concurring in part and concurring in the judgment). Dictionaries are likewise imperfect, and define words, not phrases. *See generally* Thumma & Kirchmeier, *supra* note 63.

[80] Statutory interpretation invites temporal and contextual questions as well. For example, should one look to the ordinary meaning at the time a statute was enacted or the ordinary meaning at the time of the criminal action itself? Should one look to the ordinary meaning understood by a legislator or the public? Both corpora and search results can be tailored to provide samples only from certain sources or certain time periods. For instance, COCA can filter results based on finely tuned parameters such as academic subject matter or newspaper section. *See Introduction*, *supra* note 8.

[81] *Rasabout*, 356 P.3d at 1264.