

SIXTH CIRCUIT REVIEW

Corpus Linguistics: Just Another Tool in the Sixth Circuit Toolbox?

ANDRIA DORSTEN EBERT

Corpus linguistics use has been trending recently in [state courts](#), and reached the Sixth Circuit in two cases this summer. Judge Thapar indicated his interest in corpus linguistics, initially in a concurrence in [Wilson v. Safelite Group, Inc.](#), 930 F.3d 429 (6th Cir. 2019), in which he provided an extended analysis of this legal method. And, dropped into a footnote in [Wright v. Spaulding](#), No. 17-4257, 7 n. 1 (6th Cir. Sep. 19, 2019), Judge Thapar also indicated that he had asked counsel to provide him with an analysis of the text based on corpus linguistics methods.

[Corpus linguistics](#) is an approach to studying language that uses electronic collections of linguistic data known as corpora. These corpora are built from real-world language used in their initial context—in books, magazines, legal documents, and transcripts of spoken language. These digitized databases allow legal practitioners to analyze language for patterns of usage in a [more targeted and transparent way](#) than a mere dictionary definition can provide.

For example, a legal linguist can use the open-source [BYU corpora](#) to discover not only the dictionary definition of the word “personal,” but also how it is used as an adjective to modify other nouns. This [replicable search](#) then demonstrates to the court that the most common nouns that “personal” modifies include “personal life,” “personal experience,” “personal friend,” “personal appearance,” all to demonstrate that “personal privacy” should only apply to people, and not to corporate entities, despite a corporation’s status as a legal “person.” Using a corpora search instead of a dictionary can be more useful and help mitigate any bias associated with the acontextual nature of a dictionary definition.

A common thread that runs through both traditional methods of interpreting text and corpus linguistics is that [words have meaning](#). Using corpus linguistics as a legal tool is derived from older premises in statutory interpretation: [use of dictionaries](#) and their corollaries will aid in finding the ordinary meaning of a statute. Its proponents state that while a dictionary provides a static interpretation of a given word, corpus linguistics can give a much more dynamic interpretation.

However, [opponents of corpus linguistics](#) take a significantly

more pessimistic view. They argue that corpus linguistics is not the panacea of objectivity and transparency that its proponents claim, but instead leads legal interpreters to define a word [radically out of its context](#). For example, how does the use of a word in [Moby Dick, the King James Bible, or newspaper articles](#) clarify the meaning of a word in an ambiguous present-day statute? Instead, use of corpus linguistics is just as (or even perhaps more) subjective a method of statutory interpretation as the other tools available to judges. Despite its seemingly transparent and scientific nature, it is still exposed to [human subjectivity](#). Corpus linguistics still requires human judgment when choosing the corpus, the search terms, and in analyzing the results for interpretive application.

Even committed textualists understand that [context matters](#). And, discussing the need for context when interpreting ordinary language, Justice Stephen Breyer stated:

When I see the word “any” in a statute, I immediately know it’s unlikely to mean “anything” in the universe...When my wife says, “there isn’t any butter,” I understand that she’s talking about what is in our refrigerator, not worldwide. We look at context over and over, in life and in law.”

Corpus linguistics, used on its own, cannot distinguish between “in life and in law.” But recognizing the role of human judgment in corpus linguistics can aid in interpreting a statute and can keep ordinary language in context.

As Judge Thapar stated, corpus linguistics can be a valuable tool to “help courts as they roll up their sleeves and grapple with a term’s ordinary meaning.” *Wilson*, 930 F.3d at 445 (Thapar, J., concurring). Even so, courts should recognize that “corpus linguistics is one tool—new to lawyers and continuing to develop—but not the whole toolbox.” *Id.* at 440. Without the addition of human judgment to create the search parameters or add context and purpose to the terms used, corpus linguistics might not be “the most helpful tool in the toolkit.” *Wright*, No. 17-4257 at 7 n. 1. Instead, its use should be tempered by recognizing that it “brings us no closer to an objective method of statutory interpretation,” and involves human “judgment calls.” *Id.* at 448 (Stranch, J., concurring); *Id.* at 441 (Thapar, J., concurring). Using corpus linguistics with the full range of judicial tools, such as “historic and common-sense considerations—including the ‘text, structure, history, and purpose’ of a statute” can help guide the court and avoid corpus linguistics’ acontextual

2019]

OHIO STATE LAW JOURNAL ONLINE

limitations. *Id.* at 441 (Thapar, J., concurring).