38

Corpora and legal interpretation
Corpus approaches to ordinary meaning in legal interpretation

Stefan Th. Gries

Introduction

Legal texts (a constitution, laws/statutory texts, trusts, contracts, …) serve to provide guidelines regarding which actions/behaviors are permitted, which are not, and how violating these guidelines can/shall be sanctioned. This presupposes that the meanings of these texts are (correctly and consistently) interpreted by their readers: (i) people who are subject to these laws, so they can plan their actions with knowledge of their legal consequences and (ii) legal practitioners, so they can apply the legal texts in concrete situations. When it comes to judges engaged in legal interpretation, we can distinguish very different interpretive methodologies – textualism, intentionalism and purposivism – but it is probably fair to say that, when it comes to tackling the meaning of a legal text, all judges have more or less adopted a textualist practice of at least starting from the text of the relevant statute(s); as Justice Kagan stated in her Scalia Lecture Series, 'we are all textualists now' (https://youtu.be/dpEtszFT0Tg?t=508 accessed 16 July 2020).

While interpreting the meaning of (a word or part of) a legal text is a function of many things – e.g., legal canons such as *ejusdem generis* or *noscitur a sociis*, a statute's (pragmatic and historical) context, precedent and, potentially, legislative history – one of the main components to legal interpretation is the semantic meaning of the text, which, in turn, is a function of the meanings of the words and grammatical constructions (plus the above-mentioned interpretive rules specific to law). With regard to the semantic meaning of a statute, one of the most important guidelines regarding legal interpretation is the *plain meaning rule* and the *ordinary meaning doctrine* that it implies (even though the two notions are not synonyms, see Slocum 2015: 22f.). The ordinary meaning doctrine entails that words not defined in statutes are used in their plain/ordinary meaning (Scalia and Garner 2012: chapter 6; Slocum 2015: sections 1.1, 1.6; Hutton, Chapter 6, this volume.). This approach aims at deferring to the presumed intent of legislators, is compatible with the notice or fair-warning function of the law, helps assure consistency in legal interpretation and application, and protects reliance interests.

Given the centrality of the ordinary meaning doctrine, it is probably surprising that, '[c]urrently, the ordinary meaning doctrine is greatly undertheorized' (Slocum 2015: 30) or even 'ironically, we have no ordinary meaning of "ordinary meaning"' (Lee and Mouritsen 2018: 798). Specifically, legal practitioners and courts face the following interrelated questions (Slocum 2015: 2, chapter 3): the *constituent question* 'of what makes some meaning the ordinary one,' and the *evidential question* 'of how the determinants of the ordinary meaning of legal texts are identified'; see also Lee and Mouritsen (2018: 798) on the 'problem of operationalization or measurement.' Then how do courts deal with the vagueness and ambiguity that is inherent in legal texts?

(i) Judges have relied on their *intuition* as native speakers: if words not defined in, say, a statute are to be interpreted in their ordinary meaning, then surely judges should be able to rely on their understanding of the term;

(ii) Judges have used *dictionaries*, which have been created externally to the judge and the case at hand and should be authoritative sources regarding word meaning; Slocum (2015: 21) summarizes previous research showing that, while the United States Supreme Court's use of dictionaries was virtually non-existent before 1987, now as many as one-third of statutory decisions cite dictionary definitions (see Mouritsen (2010: 1920) for similar data on both the use of dictionaries and the invocation of the plain meaning rule);

(iii) Judges have used the *etymology* of words, essentially arguing that the contemporary use of a word is based on, or similar to, a word's (often Latin or Greek) origin;

(iv) Combining some of the above, judges have engaged in *morphological analysis*, as when the majority in *State v. Rasabout* (2015) parses *discharge* (v.) into the prefix *dis-* and the root *charge* to argue that *discharging a firearm* refers to 'firing one shot,' not 'emptying a complete magazine' (see also Hardaker, Chapter 42, and Hutton, Chapter 6, this volume, for discussion of this case).

Unfortunately, all of these practices are even more problematic than they are widespread. Linguists have known for decades and studied extensively the degree to which intuitions about language are often fallible, not shared, and volatile. Speakers are fairly good at determining whether a sentence is grammatical in their native language, but a huge amount of individual variation has been documented already since the 1970s (Labov 1975; Schütze 1993). However, speakers are even worse at determining the commonness of (meanings of) words. Judges sometimes seem to believe that they are experts in language/linguistics already. Lee and Mouritsen (2018: 866, their emphasis) flatly assert with some bravado that '[t]he inevitable fact of the matter is that judges and lawyers *are* linguists' – given how much they interpret and write language – but, just because someone deals with extremely unordinary language (Slocum 2015: 11f.) all day every day, that does not make them an expert on ordinary language. Plus, if all that was needed was good judicial intuitions plus the above tools, then how come even a group of speakers as homogeneous in background and education as nine justices on the Supreme Court of the United States cannot agree on the ordinary meaning of expressions as mundane as *use a firearm* or *carry a firearm*? Finally, relying on judges' intuitions regarding some expressions' meanings makes jurisprudence rely on something that is volatile, subject to unconscious cognitive heuristics, non-replicable, non-transparent, and likely to be 'off' when it comes to the ordinary meaning of a phrase (currently or when, say, a statute was enacted). As Lee and

Mouritsen (2018: 867) state, 'nothing is statistically worse than one data point – especially a biased one. The potential for motivated reasoning is evident,' and 'the interpretation of statutory language often involves judgments about far smaller distinctions than everyday life demands' (Solan and Gales 2017: 1333).

As for dictionaries, judges are using them for their *definitional function* (i.e. for what an expression means) and their *instantiating function* (i.e. for confirming whether a certain meaning has been recognized as valid or is a possible meaning). However, *how* judges have been using dictionaries is just as problematic as their use of intuition in a variety of ways that are discussed in an exemplary fashion in Mouritsen (2010). A first central 'meta'-problem is that courts seem to attribute to dictionaries a third function, namely an *ordinary-meaning identification function*. That is, judges use a list of *possible* senses/meanings of a term in a dictionary entry to support claims about its *ordinary* meaning, although that is not the dictionary's goal; dictionaries 'are not in the business of establishing *ordinary* usage.' Instead, they seek to 'instantiate the realm of *permissible or possible* usage' (Mouritsen 2010: 1945, my emphasis). Thus, this is a 'meta'-problem, because everything else that courts subsequently do with the information from a dictionary is already tainted by that misunderstanding. A second problem aggravating the situation is that courts often commit what Mouritsen (2010: 1036) calls the *sense-ranking fallacy*, i.e. believing that the order of senses/meanings means something (such as attributing more ordinariness to the first- or earlier-listed senses than to later ones) even when dictionaries such as *Websters Third* or the *OED* clearly state otherwise in their front matters or explanatory notes; and even dictionaries such as the *Random House Dictionary*, which lists senses 'generally' in order of frequency and degree of specialization are not much more suitable for the purposes of identifying ordinary meanings, given that one does not know when the above 'general' guidelines are not applied and whether the frequency information that is supposedly adhered to is accurate.

As for etymology, a criterion often used based on dictionaries, here, too, courts are employing a criterion that flies in the face of 'them being linguists.' As Mouritsen (2010: 1940) succinctly summarizes, '[t]he notion that we may accept a given meaning as valid simply because its etymology is consistent with our proffered meaning is unsustainable because it would lead to absurd results: *December* would quite literally mean *October*, *anthology* would mean a *bouquet of flowers*' – to explain the meaning of modern-day *carry* in English with reference to Latin *carum* (which means 'car' or 'cart') as done by Justice Breyer in *Muscarello v. United States* (1998) is as absurd as it sounds *if* the stated goal is to characterize the ordinary meaning of the term.

Finally, there are more general problems with dictionaries. First, judges can pick whatever dictionary/dictionaries they wish, which allows them – even if only unconsciously – to select a dictionary whose sense divisions and definitions meet a particular purpose: a judge or their clerk(s) working with a dictionary in their chambers can, but need not, conduct an objective, controlled and replicable scientific search for the truth. Second, lexicographers are just as fallible as the rest of us when it comes to the notoriously difficult or even impossible tasks of distinguishing senses of a word or identifying their frequency or obsolescence; as Green (1997, quoted in Mouritsen 2010: 1916) states, 'dictionaries do not emerge from some lexicographical Sinai.' Finally, as Mouritsen (2010) argues, given (i) that the opinions individual judges arrive at are usually not independent, (ii) that the decisions lexicographers arrive at are usually not (completely) independent and (iii) dictionaries' characteristics and their uses by judges, it is far from obvious that an amalgamation of

'votes' regarding the controlling sense of a term (along the lines of the Condorcet Jury Theorem) will lead to the correct outcome.

In sum and as Hobbs (2011: 344) concludes,

> This is not to say that courts should dispense with dictionaries altogether; however, they should be used for their intended purpose, to define technical or unfamiliar words. […] Using the dictionary to define common words is at best unnecessary, and may contribute to strained interpretations of statutory texts by producing meanings that are not ordinary but extraordinary.

Corpus-linguistic methods and their application to legal interpretation

Partially in response to the above-described facts, the last few years have seen a move towards corpus data in legal interpretation. Following a brief mention in Solan (2005) that corpus linguistics might be relevant to statutory interpretation, the publication of Mouritsen (2010), and Goldfarb's (2010) brief to the Supreme Court (in *FCC v. AT&T Inc.* (2011)), this field has been growing via a series of annual conferences at the University of Chicago and especially Brigham Young University (under the stewardship of their Law School and Associate Chief Justice of the Utah Supreme Court, Tomas R. Lee) and has just culminated in Lee and Mouritsen (2018) and several papers in the *BYU Law Review*. Legal scholars and practitioners are now exploring ways in which corpus-linguistic methods may help make legal interpretation more objective, replicable and robust. This section first provides a brief introduction to corpus linguistics and its potential for legal interpretation, before I turn to its main methods and how they have been applied in scholarly work and actual cases.

Corpus linguistics and its potential advantages

Corpus linguistics is a scientific discipline at the intersection of linguistics, digital humanities, computer science and statistics/information theory. It is a branch of linguistics based on the (sometimes *highly*) statistical analysis of data from a corpus. The notion of a corpus in turn is – like probably most polysemous words – a radial/prototype category with a relatively firmly established center/core and less central members. A prototypical corpus

(i)   consists of *machine-readable* Unicode text files;
(ii)   is meant to be *representative* for a particular kind of speaker, register, genre, variety or language, which means the corpus's sampling scheme represents the variability of the population it is meant to represent;
(iii)   is meant to be *balanced*, which means the sizes of the subsamples are proportional to the proportions of such speakers, registers, varieties, etc. in the population the corpus is meant to represent;
(iv)   contains data from *natural communicative settings*, which means the data in the corpus were produced not for the corpus and that their production was untainted by their collection.

Thus, it is probably fair to say that the *British National Corpus* (BNC, www.natcorp. ox.ac.uk/) represents a prototypical corpus: its most widely used version, the BNC World

Edition XML, consists of 4,049 XML-annotated Unicode text files containing altogether approximately 100 million words intended to be representative of British English of the 1990s. Similarly, the *Corpus of Contemporary American English* (COCA, https://corpus. byu.edu/coca/) is also fairly close to the prototype (although its underrepresentation of truly spoken/conversational data would give pause to some corpus linguists, and it is a dynamic/monitor corpus, i.e. it is a corpus whose size is changing over time as new material gets added to it).

Corpora can be searched with dedicated corpus software or, better, programming languages such as R or Python to find 'things' in corpora; *things* is deliberately vague because it can refer to anything: letter sequences, morphemes, words, whole phrases, sentences, any kind of annotation (such as part-of-speech tagging or lemmatization) etc. That means, in the simplest possible corpus application, one can look for a word or lemma and then, according to Lee and Mouritsen (2018: 831f.), '[w]hether we regard the ordinary meaning of a given word to be the *possible*, *common*, or the *most common* sense of that word in a given context, linguistic corpora allows [sic] us to determine empirically where a contested sense of a term falls on that continuum.' That is, corpus analyses inject empirical data and scientific methods into what in a (hypothetical) worst-case scenario could be non-transparent, non-replicable, non-falsifiable and possibly selective/biased legal interpretation. (Obviously, the question of whether one *wants* an objective, empirical and replicable approach to meaning is an entirely different one, and see below for some discussion of criticism leveled against corpus linguistics in the legal domain.)

Given the ever increasing size and availability of the internet, the questions might arise to what degree one can consider and use the internet as a whole or Google (either just the search engine or Google Books) as a corpus. Within corpus linguistics, opinions on this are divided, but it is probably fair to say that, while the size of 'the internet' or Google Books is certainly an attractive characteristic, using the internet/Google also comes with many very serious problems, including massive problems regarding representativeness and sampling (Kilgarriff 2007).

Corpus-linguistic methods

Concordances

Maybe the most central corpus-linguistic method is the concordance. In a concordance, each match of a search term – typically a word – is shown with a certain user-defined context (such as a window of *n* words to the left and right, one or more sentences, the whole paragraph etc.); usually the matches are highlighted, e.g. with bracketing, tabs or other kinds of formatting. These displays are not always easily exploitable, but maximally comprehensive because of how everything the corpus contains about the context of the search term will be provided – not only linguistic context but also, depending on the annotation of the corpus, for instance who said the search term (to whom), how was it understood (as inferred from an interlocutor's reaction), how the search term was produced (with disfluencies? with certain voice qualities?) etc. Because of this richness, concordances are usually not analyzed completely automatically but require a human annotator's attention.

In many applications, legal or otherwise, concordance lines are annotated for features of interest, such as senses/meanings but also lexical, grammatical or other characteristics.

For semantic tasks including those of legal interpretation, Phillips and Egbert (2017) distinguish three different kinds of annotation approaches:

(i) a *minimalist approach*: one only annotates whether each use in one's concordance instantiates sense$_1$ or sense$_2$ of two relevant senses;
(ii) a *dictionary-driven approach*: one annotates in a top-down fashion each use for which of the senses in one or more dictionaries it instantiates;
(iii) the *grounded-theory approach*: one annotates in a bottom-up approach each use for the sense it instantiates, but the senses are developed inductively. Thus, if an analyst changes the sense inventory after having seen the first 100 matches, the first 100 matches need to be revisited to determine whether, now, some of them need to be reclassified.

A different analytical procedure, though one that is compatible with the grounded-theory approach, is presented in Goldfarb (2017: 1399f.), the so-called *corpus pattern analysis* by Patrick Hanks: concordance lines are first grouped into semantically motivated/similar grammatical patterns, before a meaning is associated with each pattern based on information from the linguistic context (in particular lexical and grammatical information); see the section on *Muscarello v. United States* (1998) for an application.

Collocations

A second corpus-linguistic method of overall minor importance is the collocation display, where words co-occurring with the search term are shown with their frequency of co-occurrence; for instance, a collocate display of *difference* for the Corpus of Contemporary American English shows that *between* is very frequent around *difference* (in particular in position R1, one word to the right of *difference*); similarly a collocate display for the L1 slot of *difference* restricted to adjectives in that same corpus shows that *significant* is most frequent directly in front of *difference*, followed by *big* or *huge* etc. Collocate displays *seem* much easier to interpret because of the focus they appear to provide compared to the apparent noisiness of a concordance. However, because of their lack of context, they often do not allow the finer distinctions legal interpretation may require. For instance, even if *significant* was indeed the most frequent adjective in front of *difference*, the collocate display does not allow the analyst to see which of the instances of *significant* were used with the technical sense 'statistically significant, i.e. <5% likely given some null hypothesis') and which were used in the non-technical sense of 'substantial' – for that, one would need a concordance.

Selected applications

This section discusses a small number of cases to which corpus-linguistic methods have been applied. The first case will be discussed in more detail, given how it has been claimed to be one of the great textualist moments of our period and how it exemplifies many of the above-mentioned problems; subsequent applications are presented more succinctly (see Hutton, Chapter 6, this volume for discussion of these cases too).

Muscarello v. United States, *524 U.S. 125 (1998)*

When Frank Muscarello was transporting marijuana to sell it, he had a handgun in the locked glove compartment of his truck; the question was whether this way of transporting the handgun is covered by 18 U.S.C. §924(c)(1), which requires a sentence enhancement for a person who 'uses or carries a firearm' 'during and in relation to' a 'drug trafficking crime.' Justice Breyer's majority opinion argued that this scenario is covered by 'carry a firearm.' The majority relied upon 'an extraordinary panoply of sources, including: the King James Bible, *Robinson Crusoe*, and *Moby Dick*, as well as two electronic newspaper databases and several unabridged dictionaries' (Mouritsen 2010: 1916). However, the court's use of dictionaries exhibits many of the above problems: for instance, Justice Breyer relies on 'the greatest of writers,' the etymology of *carry* (tracing it back to Latin and Old French), and sense rankings in dictionaries that explicitly discourage the use of the sense numbering. Also, Justice Breyer reports results of a search in newspaper databases – a resource at least comparable to a corpus – stating that 'random sampling suggests that many, perhaps more than one-third, are sentences used to convey the meaning at issue here, i.e., the carrying of guns in a car' (*Muscarello v. United States* (1998): 129).

However laudable this precursor of a corpus analysis may seem, it also underscores the pitfalls when lawyers/judges, not linguists, do the analysis. Justice Breyer searched for sentences containing the words *carry*, *vehicle* and *weapon* (or variations thereof) in journalese. Of course he would find many instances where these co-occur and instantiate the majority's sense of *carry* – the relevant search would have been to search for *carry+weapon* and count how often these matches do (not) mean 'transporting in a car.' Since only about a third of the matches for his already biased search mean 'transporting in a car,' Justice Ginsburg's dissent rightly wonders 'what meaning showed up some two thirds of the time.' Plus, and this is the ordinary meaning doctrine coming into play, to the extent that the court documented anything at all, it is that 'transport a firearm in a car' is a *possible* meaning of *carry a firearm*, but not that it is a *common* let alone the *most common* or the *ordinary* meaning, even though, in the eyes of e.g. Justice Scalia (*Chisom v. Roemer* (1991): 410), the court's job 'is not to scavenge the world of English usage to discover whether there is any possible meaning' – 'our job is to determine … the ordinary meaning.'

Mouritsen's (2010) is a much better, truly corpus-based analysis pertinent to *Muscarello*. He analyzes 500 randomly sampled concordance lines from COCA and, while his ultimate goal is to contrast $carry_1$ ('carrying as conveyance') and $carry_2$ ('carrying upon one's person'), he counts other senses as well; his study therefore is a hybrid of the dictionary-based approach and corpus pattern analysis. He finds that $carry_2$ strongly outnumbers $carry_1$, but is careful to discuss what this implies:

> My contention is not that because $carry_2$ is far more common than $carry_1$, § 924(c) ought to be interpreted with the $carry_2$ meaning. Such a reading would be arbitrary. There are undoubtedly circumstances in which Congress employs the less frequent of two senses of a word. However, these data do serve to undermine the Court's contention that $carry_1$ is the word's primary or ordinary meaning, while $carry_2$ is a 'different, rather special way' to use *carry*. Instead, the circumstances are reversed. When it comes to the support or movement of a physical object, $carry_2$ is the primary or common use of *carry*, while $carry_1$ is the specialized sense of the term.

This conclusion is then reinforced by an analysis of data from the Corpus of Historical American English (COHA, https://corpus.byu.edu/coha/ accessed 16 July 2020).

Goldfarb (2017) applies a corpus pattern analysis to 901 instances of the lemma *carry* from COCA. He finds that

> corpus lines categorized as HUMAN CARRY OBJECT IN VEHICLE [...] were greatly outnumbered by those categorized as simply HUMAN CARRY OBJECT. [...] All of which gives reason to believe that phrases following the pattern [*human*] *carry* [*object*] are not used to express the meaning HUMAN CARRY OBJECT IN VEHICLE unless the IN VEHICLE part is explicitly encoded or otherwise supported by something in the context.

Additional analyses on subsets of his concordance lines do not change the picture; just as carefully as Mouritsen, Goldfarb concludes that

> *carry* is used more frequently to talk about acts of personally carrying objects (for instance, in one's hands or arms, or strapped to one's back), than about events in which objects are transported or carried in a vehicle. That may or may not suggest that the former use is more 'basic' or 'primary' than the latter, but it certainly does not suggest the opposite.

> (Goldfarb 2017: 1409)

In sum, Muscarello is a very instructive case: it showcases just about everything that can go wrong when improper dictionary use and an improper 'corpus-like' search are used to determine not ordinary, but a possible, meaning, but it also reveals how different kinds of proper corpus-based approaches come to a conclusion that provides strong evidence for what the ordinary meaning of *carry a firearm* might actually be (if we operationalize ordinariness with frequency).

United States v. Costello, *No. 11–2917 (7th Cir. 2012)*

Deanna Costello was charged with violating 8 U.S.C. § 1324(a)(1)(A)(iii); specifically, the issue was whether she was guilty of 'harboring' an illegal alien and, according to Lee and Mouritsen (2018: 805) 'whether the ordinary meaning of the verb *harbor* required proof of concealment.' Interestingly, Judge Posner argues against an overreliance on dictionaries and instead performs Google searches, again a forerunner of corpus-linguistic analysis. Much like Justice Breyer in *Muscarello*, however, his Google search is flawed: first and, perhaps, overly strictly speaking, his Google search does not target the time period when the relevant statute was enacted (1917). However, it is neither obvious that the verb *harbor* underwent substantial semantic change, nor is it obvious that the ordinary meaning of a statute requires looking for at the time it was enacted. Second, Google searches do not return frequencies of matches, but numbers of pages with matches, and, even worse, they are not replicable: because of Google's proprietary algorithm, search results differ from one day and person to the next; because of the impact of one's search history, it can even make a difference whether one is logged into Google's services or not. Most importantly, though, like Justice Breyer, Judge Posner restricts his data on *harbor* (v.) to a preselected set of direct objects, biasing the results.

Lee and Mouritsen study *harbor* (v.) both in the time period 1910–1919 and in current usage, and find that it is mostly used with emotion/cognition nouns as direct objects in both periods. Lee and Mouritsen then focus on direct objects relevant to the statute, but find the data inconclusive: of those matches whose precise relation to 'sheltering' and 'concealing' was identifiable, both senses were relatively equally frequent, but most matches could not be categorized.

Gries and Slocum's (2017) analysis is essentially a combination of a grounded-theory kind of analysis and a corpus pattern analysis. They also find a strong prevalence of *harbor* to take emotion/cognition nouns as objects (e.g. *aspirations*, *joy*, *illusions*, *hope*, *doubts*, *a crush* etc.), which they argue supports the government's more inclusive interpretation along the lines of 'providing space/a habitat for something.' Also, they discuss *harbor*'s semantic prosody and compare it to *shelter* (v.), showing the kinds of objects that *shelter* takes (e.g. *a family*, *refugees*, *protesters*, personal pronouns referring to humans) are statistically significantly different from those of *harbor* (as mentioned above) and that *shelter*'s collocational behavior is more compatible with Judge Posner's analysis of *harbor*.

In sum, this is another case where justices (laudably) recognized the limits of a dictionary-based approach to ordinary meaning, but fell short in implementing their alternative.

Smith v. United States, *508 U.S. 223 (1993)*

John A. Smith tried to trade his MAC-10 for cocaine and, as *Muscarello*, was convicted of having 'used a firearm during and in relation to a drug trafficking crime'; his sentence involved a sentence enhancement because the court held that the statute does not require the firearm to have been used 'as a weapon.' This case is remarkable and by now infamous, because the majority claimed to approach this case from an *ordinary meaning* and a *textualist perspective* ('[when] a word is not defined by statute, we normally construe it in accord with its ordinary meaning,' (*Smith v. United States* 1993: 228), but then proceeds on the basis of extremely general dictionary definitions of *use* (v.) showing that 'trade' is a *possible meaning* of *use* with a very *intentionalist-looking perspective*. Correspondingly, Justice Scalia's dissent argues that the majority 'does not appear to grasp the distinction between how a word *can* be used and how it *ordinarily is* used' and that it fails to consider context in its interpretation: '[t]o use an instrumentality ordinarily means to use it for its intended purpose' and '[t]he ordinary meaning of "uses a firearm" does not include using it as an article of commerce' (*Smith v. United States* 1993: 242).

Gries and Slocum's (2017) analysis of *use* (v.) confirms Justice Scalia's argument. If ordinariness of meaning is operationalized via (relative) frequency of occurrence, 'trade' is not the ordinary meaning of either *use*+[weapon noun] or *use*+[any concrete object]: controlling for corpus dispersion (see below), they find not a single instance in more than 300 COCA matches. Revealingly, Justice Ginsburg's concurring opinion in the related case of *Watson v. United States*, 552 U.S. 74 (2007) belatedly agrees with Justice Scalia's dissent:

> For reasons well stated by Justice Scalia in his dissenting opinion in Smith, 508 U. S., at 241, I would read the word 'use' in §924(c)(1) to mean use as a weapon, not use in a bartering transaction. Accordingly, I would overrule Smith, and thereby render our precedent both coherent and consistent with normal usage.

Pointers to other relevant corpus-linguistic applications

One notable corpus application is *FCC v. AT&T Inc.*, 562 U.S. 397 (2011), in which Neal Goldfarb filed the first known brief in the U.S. Supreme Court using data from COCA regarding the meaning of *personal* (adj.). Another interesting application is Associate Chief Justice Lee's concurrence in *State v. Rasabout* (2015), where he uses corpus data from COCA regarding the meaning of *discharge* (v.). Finally, 2018 and 2019 saw the submission of multiple amicus briefs to the Supreme Court, which involve corpus analysis, namely Heilpern and Schaerr's amicus briefs in the cases *Lucia & Lucia Companies, Inc. v. SEC* (2018) and *Rimini Street, Inc. & Seth Ravin v. Oracle* (2018) as well as Slocum, Gries and Solan's brief in *Gerald Lynn Bostock v. Clayton County, GA* and Baron et al.'s brief in *New York State Rifle & Pistol Association Inc., Rommolo Colantone, Efrain Alvarez, & Jose Anthony Irizarry v. The City of New York and the NYPD License Division*.

Criticism against corpus-linguistic methods and some rebuttals

This section discusses criticism against the application of corpus-linguistic methods from the side of legal scholars and practitioners and from a (corpus-)linguistic perspective.

Criticism from a legal perspective

Corpora are not representative of anything and in fact unnecessary
Corpus-based approaches could be criticized on the grounds that corpora are finite (as opposed to the open-endedness of language) and do not represent the relevant speech community; this could be argued on a synchronic level ('corpus X does not represent all speakers or all registers') and/or on a diachronic level ('corpus Y does not represent the time period of the statute's enactment'); see Mouritsen (2010: 1966–1970). Plus, legal practitioners might consider themselves experts both because of their perceived generalist training and being native speakers of, here, English (see above for a first counterargument). This is partly a rehash of Chomsky's (1957: 15) critique of corpus data. Yet, while there are no infinite corpora and complete representativeness and balance are largely theoretical ideals, corpora still provide concrete empirical evidence of the behavior of, often, (tens of) thousands of speakers and hearers at the relevant time. Thus, for ordinary public meaning, corpus analysis permits a scientific discussion of pros and cons of corpus design, corpus choice, sampling decisions, annotation choice, rater reliability and suitability of statistical analysis, which is immeasurably better than, to use a polemic example, judge A saying 'I and my favorite dictionary think sense *x* is more ordinary,' to which judge B responds '[b]ut I'm a native speaker, too, and I think sense *y* is more ordinary, as does Shakespeare in *Hamlet*' – at least the scientific decisions leading to a certain corpus-based analysis are explicit and, thus, debatable – preferably by experts – in an adversarial system.

As for the criticism that corpora may not be useful for recovering the ordinary meaning of a text at the time of enactment, there are two counterarguments to be made. First, for many applications, it is possible to find corpus data for the relevant time period; for instance, one might use COHA to study the use of *harbor* (v.) around 1917, as Lee and Mouritsen (2018) did, one might study the use of *sex* and *gender* in Title VII of the Civil Rights Act of 1964 as in Slocum, Gries and Solan's amicus brief to the Supreme Court or

one might use the Corpus of Founding Era American English (COFEA, https://lawncl. byu.edu/ accessed 16 July 2020) for questions of constitutional meaning.

Second, it is not even clear that, for ordinary meaning, the meaning of the expression at the time of enactment should even be relevant. Historical meaning is certainly relevant from an originalist perspective of intended meaning, but how could it possibly be relevant for contemporary ordinary meaning and fair-notice considerations: would Deanna Costello, or anyone else for that matter, be supposed to read 8 U.S.C. § 1324(a)(1)(A)(iii) in 2012 but know how the meaning of *harbor* (v.) was different in 1917 from what it is now? The fair-notice requirement would seem to imply a contemporary ordinary reader perspective to the statutory text (Lee and Mouritsen 2018: 825).

Corpus analysis is not objective (either)

The annotation of, say, concordance lines by linguists/annotators could be criticized for being just as subjective as a judge's armchair analysis/parse of the meaning of some statutory text (e.g., Hessick 2017: 1525). However, such problems are well known in any social science and there are a multitude of precautions social scientists are aware of (and that are unavailable to a judge or lawyer not trained in social science methods). Therefore, a proper scientific analysis of such data involves multiple trained/supervised annotators, an annotation manual, a resolution-of-disagreement process, statistical analysis quantifying interrater agreement and the robustness/generalizability of the results etc. As mentioned above, all these characteristics allow the analysis to be explicit, be questioned and be discussed and improved in a way that a single judge's/lawyer's intuition or unsupervised (mis)use of dictionaries obviously does not.

Corpus results can be indeterminate, too

Corpus results could be criticized for not leading to completely clear-cut results if, for instance, two senses/meanings x and y in question are equally frequent in a corpus (sample) or if x is merely 10% more frequent than y, which might not be considered a decisive margin. However, it is not like the current non-scientific intuition-/dictionary-based approach is producing clear-cut results, as 5:4 Supreme Court decisions in the above cases have shown. Social science phenomena are hardly ever clear-cut, but if those results were to be arrived at in a proper, testable and falsifiable way, then ambiguous results are informative, too, and might lead a court to, for instance, conclude that the ordinary meaning of a statute is not clear and apply the rule of lenity or Chevron deference (i.e., interpret a statute in a way favorable to the defendant or accept, or defer to, a government agency's permissible interpretation of a statute respectively).

Corpus frequencies do not speak to ordinary meaning

Corpus results could be doubted as saying less about how the frequency of a sense represents something about a sense's ordinariness and more about how frequent the sense/ situation is in the real world; see Solan and Gales (2017) or Herenstein (2017). This is similar to Chomsky's famous but mistaken claim that corpora are useless because the fact that I live in New York may be more frequent in a corpus than I live in Dayton, Ohio has no linguistic relevance. However, Stefanowitsch (2005) has shown that the former sentence being more frequent than the latter is corpus-linguistically unproblematic: proper statistical analysis reveals that the former sentence is not significantly more frequent than the latter. Thus, part of this concern results from a lack of the kind of statistics that corpus linguistics requires from its practitioners; see below.

Second, as Lee and Mouritsen (2018: 874f.) point out, if something happens more often in the real world, then not only might it be more often attested in a corpus, but for that exact reason it might also be the 'first [sense] to come to mind,' which is how Justice Scalia's dissent in *Smith v. U.S.* 'operationalizes' the ordinary meaning of use a firearm (*Smith v. United States* 1993: 230). Thus, the fact that corpus frequency may be correlated with real-world frequency does not undermine corpus frequencies as one proxy towards ordinariness of meaning – if anything, it strengthens it.

Corpus analysis is not practical / too cumbersome
Some majority opinions of the Utah Supreme Court have argued that corpus analysis could 'put a strain on parties and courts' and Hessick (2017: 1523) laments that adoption of corpus-linguistic methods would make 'litigants […] submit dueling corpus analyses.' However, as discussed by Lee and Mouritsen (2018: 871f.), this argument is not persuasive. First, the number of cases benefiting from corpus analysis is comparatively small (although I disagree with their view that corpus analysis is only relevant in cases of lexical ambiguity/vagueness and/or determining a term's extension are also amenable to corpus analysis). Second, if the stakes are high enough, why would one *not* seek expert testimony? Expert testimony is not only required for ballistics, genetics and other hard/STEM forensic sciences but also for something that many legal practitioners have little exposure to: linguistic processing/comprehension by ordinary readers, and dueling expert testimony is common in such and other cases. For instance, in a recent case involving the death of a person during a canoe expedition, I was asked to determine whether the expression '[some number] or fewer' included 'zero,' which had implications for whether the victim would have been entitled to a life vest by the canoe's owner who was sued for damages – who would want to tell the plaintiffs that linguistic analysis for this question is too cumbersome and not worth expert testimony?

Various more general misunderstandings
Hessick (2017) criticizes the law and corpus linguistics movement for a variety of perceived threats and problems. She

(i)   asserts that 'the frequency with which a term is used does not give us [the] information' of 'how an ordinary member of the public would understand the statute' (1508) or that 'the ordinary meaning ought to be determined according to the frequency with which a word is used in a particular way' (1509);
(ii)   stipulates that proponents of corpus analysis want to 'change modern statutory interpretation' with 'a new interpretive theory called "corpus linguistics"' (1504);
(iii)   claims it is delusional to assume that legislators or the public would conduct corpus analysis to understand the meaning of a (draft of a) statute (1516);
(iv)   claims corpus linguistics wants to rob judges of their freedom/job/responsibilities (1511f., 1517) and that judges are not as arbitrary as is feared, given how they have to explain their decisions in opinions.

However, all these points are fundamentally mistaken; in fact, most 'address' positions that were never even stated by anyone arguing for corpus-linguistic applications to ordinary meaning. As for (i), no corpus linguist would suggest that the mere frequency of a term in a corpus tells us anything; all studies I am aware of emphasize the relevance of linguistic *and* extralinguistic context to the question of how a statute should be and is actually understood. Her argument that recent media coverage of catastrophic flood events would

bias an ordinary-meaning corpus analysis of the word *flood* would only be meaningful if the analyst did not take context (e.g. publication outlet, distribution of *flood* across a corpus both in terms of documents and time periods etc.) into consideration. Thus, this critique, just as Herenstein's (2017), uses as a strawman a primitive corpus-linguistic 'analysis' that no corpus linguist would actually endorse.

As for (ii), no one argues corpus linguistics is a new interpretive theory; what is promoted is that it can sometimes help make judges aware of how an ordinary reader would understand a statute, which can reduce the risks of erroneous interpretation or motivated reasoning by judges. *Anyone* interested in fair notice should be deeply troubled by how *Muscarello* shows that an ordinary reader might have arrived at the same interpretation of 18 U.S.C. §924(c)(1) as four dissenting Supreme Court Justices and still be sentenced.

As for (iii), this, too, is a claim no one has made – the idea is that judges, lawyers or, better, linguists conduct corpus analyses to determine whether a reasonable defendant could have been expected to know what conduct was prohibited if they had applied their ordinary reading and comprehension skills.

As for (iv), we have seen in decisions like *Muscarello* and *Smith* that judges' argumentative methodology is often deeply flawed and that their opinions can be deceptively formulated. But no one wants to deprive judges of their constitutional obligations; the idea is to help judges determine the public meaning of a statute. Does Hessick really prefer that judges misuse dictionaries and etymologies for that purpose while pretending to consider public meaning? Once an ordinary meaning is determined, a judge can still reject it in favor of an unordinary meaning and proceed from there (see Slocum 2015 or Gries and Slocum 2017: 1431). The majority in *Chisom v. Roemer* (1991) construed the meaning of *representative* more broadly than the relevant statute's ordinary meaning (on intentionalist grounds and against Justice Scalia's textually-motivated dissent). Thus, claims that corpus linguistics makes 'judges […] not bear the responsibility' (Hessick 2017: 1518) for their opinions really just misunderstand the whole enterprise.

Finally, Tobia (2020) is methodologically the most astonishing critique of corpus-linguistic approaches. He conducts experiments where he correlates non-linguists' survey responses to a small number of corpus prompts with dictionary definitions, using a series of mixed-effects models, and 'concludes' that (i) corpus data speak more to prototypes, whereas dictionary data speak more to extensivist/permissible meanings and, thus, that (ii) corpus data are unreliable for ordinary meaning purposes. However, (i) is actually a foundational *premise* of most corpus work (legal, lexicographic or generally linguistic) rather than a *conclusion* and of course means that methods (dictionaries v. corpus linguistics) with different goals (identifying all permitted meanings v. identifying most frequent/prototypical meanings respectively) can simply not be compared that straightforwardly. Second, his logic is analogous to the following argument: 'the fact that all of Gries's surgery patients died shows that surgery is not a feasible medical procedure': no one ever claimed that lay people recruited from a crowdsourcing website (Amazon Mechanical Turk) would conduct good/insightful corpus-linguistic analysis, especially when the results from the corpus data are compared to the wrong gold standard (see (i) above) and when they are presented with just a fraction of the data a real corpus-linguistic analysis would consider. Finally, both the experimental sampling into groups and his statistical analyses are problematic: the former is random but not balanced and the latter are neither comprehensive nor done correctly.

Criticism from a (corpus-)linguistic perspective

Finally, some criticism against the current kind of corpus applications to legal interpretations has also been voiced by corpus linguists. Most of those revolve around the notion that current work underutilizes what quantitative corpus linguistics has to offer: nearly all current work relies on concordances of words and the relative frequencies of their senses; i.e. an extensionalist approach to meaning that operationalizes ordinary meaning as 'most frequent/common' (Phillips and Egbert 2017: 1608: Mouritsen 2010: 1965). However, here are some ways in which corpus linguistics has more to offer.

First, operationalizing ordinary meaning as 'the most frequent meaning' misses the well-known fact that, often, commonness is better operationalized as dispersion (Gries 2008) – i.e. the evenness of the distributions of something across a corpus – or as frequency *and* dispersion (see Slocum, Gries and Solan's amicus brief). Relatedly, in corpus-based work on legal interpretation, there are laudable references to the notion of prototypicality (Lee and Mouritsen 2018: 801f,), but they have not been integrated into a theory of ordinary meaning (though see Slocum 2015: chapter 5). For example, Gries and Slocum (2017: 1441f.) mention several ways to approach prototypicality: highest frequency, but also (i) most even dispersion, (ii) earliest acquisition, (iii) network centrality and (iv) the abstract combination of features that have a high/the highest cue validity of a feature for that category (i.e. features that predict category membership very well: the feature of 'having a beak' is highly predictive of the category birds, 'having eyes' is not). Not only are especially criteria (i), (iii) and (iv) promisingly related to the Supreme Court's informal criterion of 'what comes to mind first,' (iv) also allows us to extend corpus-linguistic methods to intensional approaches to meaning (via the features constituting a prototype). This conceptual step would address Solan and Gales's (2017: 1311) frequency-in-the-world critique: while *blue pitta* may not occur in a corpus (i.e. an extensionalist approach might fail to classify a blue pitta as a bird), an intensionalist approach would recognize that a blue pitta or a dodo are birds because they exhibit all the prototypical features of birds.

Finally, corpus linguistics is a highly statistical discipline but the law and corpus linguistics community has so far only used simple percentages. We need to report dispersion measures, choose proper (random or balanced) sampling techniques, and analyze results statistically appropriately. For example, association is a central term in corpus-linguistic methodology, so, when Lee and Mouritsen (correctly) summarize that '[a] prototype is a sense, or example of a sense, that is viewed as most strongly associated with a given term in a given context,' then the word *associated* remains unhelpfully vague unless one specifies how association is measured, and (corpus and psycho-)linguistics has something to offer there; see Phillips and Egbert (2017) for more discussion.

Concluding remarks

The field of 'law and (corpus) linguistics' exhibits all the signs of a fledgling new discipline: a recognition that the status quo in legal interpretation is suboptimal (in terms of (i) theorizing and 'measuring' ordinary meaning and (ii) what that means for courts' decisions), the development of an exciting interdisciplinary approach/solution, and a sometimes fierce debate about its theoretical foundations and methodological implications. That being said, the potential for a less arbitrary textualist approach to legal interpretation and the resulting increase in fair notice and protection of reliance interests provide

strong motivations for further discussion and research in this exciting, but also hugely impactful, area.

Acknowledgement

Further Reading

Aprill, E. (1998) 'Dictionary shopping in the Supreme Court', *Arizona State Law Journal*, 30: 275–336.
Gries, St.Th. and Slocum, B.G. (2017) 'Ordinary meaning and corpus linguistics', *Brigham Young University Law Review*, 6: 1417–1472.
Hutton, C. (2014) *Word Meaning and Legal Interpretation: An Introductory Guide*, Basingstoke: Palgrave.
Mouritsen, S.C. (2017) 'Corpus linguistics in legal interpretation', *International Journal of Language and Law*, 6: 67–89.
Slocum, B.G. (2015) *Ordinary Meaning: A Theory of the Most Fundamental Principle of Legal Interpretation*, Chicago: The University of Chicago Press.
Solan, L. (2010) *The Language of Statutes: Laws and their Interpretation*, Chicago: The University of Chicago Press.

Legal sources

Chisom v. Roemer, 501 U.S. 380 (1991).
Federal Communications Commission and United States v. AT&T Inc. et al., 562 U.S. 397 (2011).
Lucia v. Securities and Exchange Commission, 585 U.S. (2018).
Muscarello v. United States, 524 U.S. 125 (1998).
Rimini Street Inc. v. Oracle USA Inc., 586 U.S. (2018).
Smith v. United States, 508 U.S. 223 (1993).
State v. Rasabout, 356 P.3d 1258 (Utah 2015).
Watson v. United States, 552 U.S. 74 (2007).

References

Chomsky, N.A. (1957) *Syntactic Structures*, The Hague: Mouton.
Goldfarb, N. (2017) 'A lawyer's introduction to meaning in the framework of corpus linguistics', *BYU Law Review*, 2017: 1359–1417.
Gries, St.Th. (2008) 'Dispersions and adjusted frequencies in corpora', *International Journal of Corpus Linguistics*, 13: 403–437.
Gries, St. Th. and Slocum, B.G. (2017) 'Ordinary meaning and corpus linguistics', *BYU Law Review*, 2017: 1417–1472.
Herenstein, E.J. (2017) 'The faulty frequency hypothesis: Difficulties in operationalizing ordinary meaning through corpus linguistics', *Stanford Law Review Online*, 70: 112–122.
Hessick, C.B. (2017) 'Corpus linguistics and the criminal law', *BYU Law Review*, 2017: 1503–1530.
Hobbs, P. (2011) 'Defining the law: (Mis)using the dictionary to decide cases', *Discourse Studies*, 13: 327–347.
Kilgarriff, A. (2007) 'Googleology is bad science', *Computational Linguistics*, 33: 147–151.
Labov, W. (1975) 'Empirical foundations of linguistic theory', in Robert Austerlitz (ed.), *The Scope of American Linguistics*, Lisse: The Peter de Ridder Press, 77–133.
Lee, T.R. and Mouritsen, S.C. (2018) 'Judging ordinary meaning', *The Yale Law Journal*, 127: 788–879.

Mouritsen, S.C. (2010) 'The dictionary is not a fortress: Definitional fallacies and a corpus-based approach to plain meaning', *BYU Law Review*, 2010(5): 1915–1978.

Phillips, J.C. and Egbert, J. (2017) 'Advancing law and corpus linguistics: Importing principles and practices from survey and content-analysis methodologies to improve corpus design and analysis', *BYU Law Review*, 6: 1589–1619.

Scalia, A and Garner, B.A. (2012) *Reading Law: The Interpretation of Legal Texts*, Eagan, MN: West.

Schütze, C.T. (1993) *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*, Chicago, IL: The University of Chicago Press.

Slocum, B.G. (2015) *Ordinary Meaning: A Theory of the Most Fundamental Principle of Legal Interpretation*, Chicago, IL: The University of Chicago Press.

Solan, L.M. (2005) 'The new textualists' new text', *Loyola of Los Angeles Law Review*, 38: 2027–2062.

Solan, L.M. and Gales, T. (2017) 'Corpus linguistics as a tool in legal interpretation', *BYU Law Review*, 6: 1311–1357.

Stefanowitsch, A. (2005) 'New York, Dayton (Ohio), and the raw frequency fallacy', *Corpus Linguistics and Linguistic Theory*, 1: 295–301.

Tobia, K.P. (2020) 'Testing ordinary meaning: An experimental assessment of what dictionary definitions and linguistic usage data tell legal interpreters', *Harvard Law Review*, 134, http://dx.doi.org/10.2139/ssrn.3266082.