Title
Corpus-linguistic approaches to lexical statutory meaning: Extensionalist vs. intensionalist approaches

Journal
Applied Corpus Linguistics, 4(1)

ISSN
2666-7991

Authors
Gries, Stefan Th
Slocum, Brian G
Tobia, Kevin

Publication Date
2024-04-01

DOI
10.1016/j.acorp.2023.100079

Applied Corpus Linguistics 4 (2024) 100079

Contents lists available at ScienceDirect

Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp

Articles

Corpus-linguistic approaches to lexical statutory meaning: Extensionalist vs. intensionalist approaches

Stefan Th. Gries *, Brian G. Slocum, Kevin Tobia

University of California, Santa Barbara & Justus Liebig Universität Giessen

ARTICLE INFO

Keywords:
Statutory interpretation
Intensionalist semantics
Prototype theory
Word embeddings

ABSTRACT

Scholars and practitioners interested in legal interpretation have become increasingly interested in corpus-linguistic methodology. Lee and Mouritsen (2018) developed and helped popularize the use of concordancing and collocate displays (of mostly COCA and COHA) to operationalize a central notion in legal interpretation, the **ordinary meaning** of expressions. This approach provides a good first approximation but is ultimately limited. Here, we outline an approach to ordinary meaning that is **intensionalist** (i.e., 'feature-based'), top-down, and informed by the notion of **cue validity in prototype theory**. The key advantages of this approach are that (i) it avoids the which-value-on-a-dimension problem of extensionalist approaches, (ii) it provides quantifiable prototypicality values for things whose membership status in a category is in question, and (iii) it can be extended even to cases for which no textual data are yet available. We exemplify the approach with two case studies that offer the option of utilizing survey data and/or word embeddings trained on corpora by deriving cue validities from word similarities. We exemplify this latter approach with the word *vehicle* on the basis of (i) an embedding model trained on 840 billion words crawled from the web, but now also with the more realistic application (in terms of corpus size and time frame) of (ii) an embedding model trained on the 1950s time slice of COHA to address the question to what degree Segways, which didn't exist in the 1950s, qualify as vehicles in this intensional approach.

1. Introduction

1.1. Ordinary meaning, and how do courts deal with the vagueness or ambiguity in statutes?

While interpreting the meaning of (a word or part of) a legal text is a function of many things – legal canons such as *ejusdem generis* or *noscitur a sociis*, a statute's (pragmatic and historical) context, precedent and, potentially, legislative history and intention – one of the undeniably 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); see Eskridge (1998:1557). With regard to the semantic meaning of a statute, one of "the most fundamental principles of legal interpretation", the sub-title of Slocum (2015), is the ordinary, or plain, meaning rule/standard 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 & Garner, 2012: Chapter 6; Slocum, 2015: Sections 1.1, 1.6; Eskridge, 2016:35; Hutton, 2020:79). Put differently, the ordinary meaning standard "focuses on how an average reader – the typical member of the public – would understand the relevant language, as opposed to the legislature's intent or purpose in creating it" (Eskridge et al., 2021:1516f.). This approach aims at deferring to the presumed intent of legislators, is compatible with the notice or fair-warning function of the law,[1] helps assure consistency in legal interpretation and application, and protects reliance interests.

This kind of view has been around for a long time: Justice Holmes famously opined that the interpreter's role is not to ask what the author

* Corresponding author.
E-mail address: stgries@linguistics.ucsb.edu (S.Th. Gries).

[1] The notice or fair-warning function of the law refers to the notion that, for instance, statutes should define crimes such that ordinary, reasonable people understand what activities are prohibited by the law. In *McBoyle v. U.S., 283, U.S. 25*, Justice Holmes wrote that a criminal statute must give "fair warning [...] in language that the common world will understand, of what the law intends to do if a certain line is passed. To make the warning fair [...] the line should be clear." (p. 283).

https://doi.org/10.1016/j.acorp.2023.100079
Received 1 September 2023; Received in revised form 7 November 2023; Accepted 15 December 2023
Available online 19 December 2023

meant to convey but instead to determine "what those words would mean in the mouth of a normal speaker of English, using them in the circumstances in which they were used." (Holmes, 1899:417). But how do legal practitioners – especially courts – put this approach into practice? The following four approaches seem to be most widespread: Judges have

- relied on their intuitions as native speakers: if words not defined in a statute are to be interpreted in their ordinary meaning, then judges should be able to rely on their understanding of the term (see Eskridge & Nourse, 2021 and Tobia et al., in press, for much discussion of the problems coming with 'intuition'-based approaches to legal interpretation);
- used dictionaries, which have been created externally to the judge and the case at hand and are intended to 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;
- 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, as when Justice Breyer's opinion in Muscarello v. U. S. motivates his 'analysis' of carry based on Latin carum ('car, 'cart');
- combined some of the above and 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'.[2]

But there are many problems with any and all of these approaches, as has been documented well by Mouritsen (2010; 2017) and Lee and Mouritsen (2018); see also Gries (2020) for a recent overview. From the point of view of linguistics, for instance, speakers' intuitions as to what can and cannot be said are fallible and subject to a great degree of individual variation. From the perspective of the current authors, particularly influential publications making this point in linguistics are Labov (1975), who showed how volatile intuitions about the acceptability of sentences can be even as they are cited as making or breaking important theoretical points, and Schütze (1993), a book-length treatment of the huge amount of both subject/speaker-related and task-related factors that can affect acceptability judgments of sentences and that are often left completely uncontrolled outside of the most controlled experimental settings.[3] Similarly, it is well known that speakers' intuitions as to what is likely or frequent can be even worse, as can be intuitions with regard to subtle meaning differences of the kinds that might arise in a courtroom. To use a seemingly trivial example, consider *-ic/-ical* adjectives. The first author has been asking native speakers for their intuitions as to when they would use *electric* or *electrical, arithmetic* or *arithmetical, botanic* or *botanical*, and *symmetric* or *symmetrical* for 20 years and has yet to find a single native speaker who has intuitions that are strong, reliable, and jibe well with lexicographic or corpus-linguistic treatments of when which of the two forms of a pair is used or whether any systematic meaning differences might exist between the members of pairs.

From the point of view of legal interpretation, the problems are even more obvious because many heated discussions regarding recent court opinions and the way they relate to precedent have been based on the very fact that even a homogeneous group of speakers like SCOTUS justices, who share many characteristics (education, profession, socio-economic status, etc., again see Eskridge & Nourse, 2021), can differ massively in their interpretations of lexical items relevant to recent cases. In addition, legal scholarship has long criticized the ways in which judges have misused dictionaries (e.g. by assuming dictionaries rank senses; see Mouritsen, 2010; 2017) or committed dictionary-shopping (Aprill, 1998; Hobbs, 2011; Brudney & Baum, 2013). Recent corpus-linguistic work on ordinary meaning in legal/statutory interpretation has rightfully questioned the notion that dictionaries are a suitable tool for matters of *ordinary* meaning, arguing instead that dictionaries are better considered as conveying *possible* meaning. Lastly and in the same vein, the use of etymologies for determining ordinary meaning is usually of no use to interpreters, given how they involve historical knowledge of word meaning and use that is undoubtedly not shared by most ordinary native speakers of a language and not relevant after however many hundreds of years of use; clearly, the currently ordinary meaning of *gay* in American English does not relate much to its original meaning of 'joyful, carefree, etc.' anymore.

1.2. Corpus linguistics 'to the rescue'

Ever since Solan (2006), which predated the influential papers by Mouritsen and Lee & Mouritsen, a growing number of legal and linguistic scholars have brought the tools of corpus linguistics to bear on matters of legal interpretation. Corpus linguistics is the method of studying language use, acquisition, processing, etc. with **corpora** (singular: *corpus*, stress on the first syllable for both) to address linguistic questions. The notion of a corpus is a prototype category, meaning there are very good, central, universally-agreed-upon examples of corpora such as the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), or the Corpus of Historical American English (COHA), but there are also less good or central examples of corpora. Technically and to anticipate some of the discussion below, the prototype of a category is defined as an abstract entity – i.e., an abstract notion that may not actually exist in the real world – that exhibits the features (or properties) $f_{1\text{-}n}$ that have the highest cue validity for a category, where **cue validity**, a term to be discussed in more detail below, can in turn be defined as follows: A feature/property (e.g., 'having a beak') has a high cue validity for a category (e.g., 'birds') if

- many members of the category have the feature (and yes, all birds have beaks);
- most non-members of the category do not have the feature (most non-birds have no beaks – animals that are not birds but have beaks make up a small set of some turtles, octopuses, platypuses/echidnas, and pufferfish).

Armed with that definition, a prototypical corpus such as the BNC or COCA is

- machine-readable, i.e. it consists of text files (with a few thousand words up to many billions of words) with language use from natural communicative settings (meaning, the language in the corpus was created for natural/regular communicative purposes);
- representative for a certain language, variety, dialect, topic, … (at a certain time) such that all parts of a language, variety, dialect, topic existing in the population are also represented in the corpus;
- balanced, such that the sizes of corpus parts representing different parts of a language, variety, etc. are proportional to the amounts these parts make up of the population.

Corpora, or the data they provide, are usually studied with one or more of the following main corpus-linguistic methods:

[2] A fifth approach, or interpretive guideline, might be 'legal norms and traditions', which includes statutory precedent or interpretive canons like 'masculine terms include the feminine'.

[3] Examples of the former kind of factors include field dependence, handedness, and level of education; examples of the latter kind of factors include instructions, order of presentation, repetition, context, meaning, and frequency effects

- frequencies: how often does something occur?
- dispersion: how widespread are the occurrences of something?[4]
- collocation: what are words occurring around an expression, which is often quantified using frequencies (how often does a word occur in the vicinity of another word?), conditional probabilities (how much of a word in % occurs around another word?), or with association measures (how much is a word attracted to another word?);
- concordance: what are the exact contexts of an expression?

2. Lee and Mouritsen (2018): collocations of *vehicle*

Let's look at Hart's (1958) famous hypothetical *no vehicles in the park!*,[5] specifically at *vehicles*. Lee and Mouritsen (2018) discuss two different approaches, namely collocations and concordance lines of *vehicle*. Their study of collocations of *vehicle(s)* involves three (parts of) corpora:

- the NOW corpus (because it is among the largest and most up-to date corpora modeling the speech community of contemporary speakers of American English, see p. 833f.);
- COHA 1950-1959 (because this is the decade during which Hart's hypothetical was published);
- COHA 1910-1939 (which we won't discuss here).

The collocates of *vehicle* that they provide are the following (our emphasis, see below):

- for the NOW corpus: ***electric***, *motor*, ***plug-in***, *unmanned, armored, connected, cars, aerial*, ***charging***, *pure, launch, owners*, ***hybrid***, *traffic*, ***fuel***, *driving, gas, autonomous, struck, operating, road, safety, accidents*, ***battery***, *ownership, emergency*, ***batteries***, *emissions, seat, advanced, driver, primary, demand, gmv, commandeered*, ***fuel-efficient***, *uavs, automakers, demonstrators, excluding, lunar, passenger, fleet*, ***gasoline***, *luxury, drove, parking, retirement, vehicles, infrastructure*;
- for the 1950s decade of the COHA corpus: *motor, space, trucks, moving, wheeled, tax, self-propelled, passenger, unit, tracked, orbit, test, b.g., launching, highways, tanks, license, robot, emergency, units, taxes, streets, equipment, manned, armored, vehicles, fees, vehicle, traveling, operate, loaded, fuel, commercial, driver, ride, traffic, designed, weight, speed, cars, carrying, operation, unsafe, horse-drawn, high-powered, amphibious, administrators, tactical, registration, delivery*.

They then discuss this in terms of two straightforward distributional phenomena (again, our emphases in the quotes):

- the presence of certain collocates:

  – "Many of the collocates of *vehicle* in the NOW Corpus strongly indicate *automobile* as a likely candidate for the **most common use** of the term." (p. 837);
  – "the collocates of *vehicle* strongly suggest that the **most common** use of *vehicle* is with reference to automobiles." (p. 839);
  – "To the extent that our notion of ordinary meaning has a **frequency component**, this data suggests that *automobile* is overwhelmingly the **most common use** of the word *vehicle* in the modern written American English represented in the NOW Corpus" (p. 842);

- the *absence* of certain collocates:

  – "*Airplane* **does not appear**, though two particular types of aircraft are attested in the collocates – unmanned aerial vehicles (drones) and spacecraft. Similarly, *bicycle* **does not appear** among the collocates of *vehicle* in contemporary usage." (p. 838);
  – "Again, **none** of the top fifty collocates of *vehicle* include the notions of *airplane* or *bicycles*" (p. 838);
  – "the absence of *airplane* and *bicycle* in the top fifty collocates of *vehicle* raises an important question for our frequency continuum. If we accept that the necessary and sufficient conditions of *vehicle* are '[a]ny means of carriage, conveyance, or transport' or 'a means of carrying or transporting something,- then there seems little question that both an *airplane* and a *bicycle* are possible readings of *vehicle*. But if *vehicle* is never used to refer to *bicycle* or *airplane* in the corpus data, then we may end up with an even further extension of our **frequency continuum** from possible but rare to possible but unattested." (p. 840).

Let's generate such collocation data ourselves for the same COHA decade of 1950–1959 but with different statistical methods.[6] If we really just go with co-occurrence frequency, the top 50 collocates of *vehicle(s)* in that decade of data are shown in Table 1 (sorted in descending order of frequency).[7]

Obviously, this is not helpful. No corpus linguist would want to infer much from a collocate display sorted/organized solely by frequency; there's a reason corpus linguistics has seen 30–40 years of research on **association measures**, i.e. measures that quantify the association between a node word of interest (here, *vehicle*) and each of its collocates. Judging from a description in a footnote describing a different search (note 242 on p. 849), Lee & Mouritsen seem to use *Pointwise Mutual Information* but we will heuristically compute one of the best association measures, the so-called **log odds ratio**.[8] We consider this measure 'one of the best' because, from a statistical perspective, it is what is called an effect size measure, meaning it is one that is least strongly correlated with co-occurrence frequency and, therefore, is a measure that does not just replicate co-occurrence frequency but provides genuinely complementary information (see Gries, 2019a; b). As a consequence of that methodological choice, we can then interpret such results on the basis of both how often collocates co-occur with the node word *vehicle* (using co-occurrence frequency) and how strongly collocates are attracted to the node word *vehicle* (using the log odds ratio as an association measure).

Fig. 1 represents the results with co-occurrence frequency on the *x*-axis and association strength on the *y*-axis. In addition, Table 2 shows the top 50 collocates of *vehicle(s)* based on a metric in the column ASSOC, which combines (i) the co-occurrence frequency, (ii) the association strength, and (iii) the degree to which collocates score highly on both dimensions (which rewards words like *tracked* over words like *power* (on the bottom right)).

[4] Note that this is *not* the same as frequency but minimally just as important a concept! If a student decides to study 20 hours for an exam, that is the amount of exposure to the course content and corresponds to frequency, but these 20 hours can be spread out over time differently often: The student might decide to study 1×20 hours (which would be called 'clumpy dispersion') or, hopefully, 10×2 (which would be called 'even dispersion').

[5] This famous hypothetical has been posed by Hart (1958), who wrote "A legal rule forbids you take a vehicle into the public park. Plainly this forbids an automobile, but what about bicycles, roller skates, toy automobiles? What about airplanes? Are these, as we say, to be called 'vehicles' for the purpose of the rule or not?" (p. 607).

[6] All corpus processing and all statistical calculations were performed using the open source programming language and environment called R (R Core Team, 2023), one of the main programming languages used in (corpus-)linguistic computing and statistical analysis (next to Python).

[7] The presence of the at sign @ is due to the words deleted from the corpus for copyright reasons.

[8] The (logged) odds ratio is not available from COHA's online interface and, thus, needed to computed by us, which was done with an R function written by the first author.

Table 1
Most frequent collocates around vehicle(s) in COHA 1950-1959.

RANK	WORD	FREQ	RANK	WORD	FREQ	RANK	WORD	FREQ
1	the	1763	18	as	176	35	which	93
2	.	1227	19	be	175	36	they	93
3	of	839	20	at	163	37	one	90
4	and	685	21	with	157	38	are	88
5	a	654	22	from	152	39	have	87
6	to	554	23	were	144	40	but	79
7	vehicle	494	24	–	129	41	their	77
8	in	464	25	by	125	42	out	77
9	vehicles	430	26	he	124	43	an	75
10	for	250	27	,	119	44	into	71
11	was	233	28	or	112	45	all	71
12	"	225	29	his	103	46	?	70
13	on	222	30	's	103	47	more	69
14	that	220	31	not	101	48	its	68
15	it	182	32	had	99	49	would	65
16	is	181	33	will	97	50	up	62
17	@	178	34	this	97			

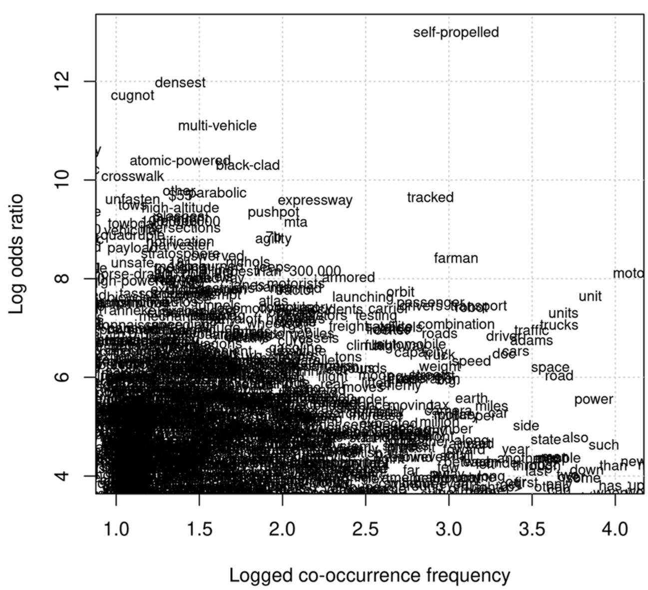

Fig. 1. Collocates of *vehicle* by co-occurrence frequency and log odds.

We can see that this approach already affords us much more precision: We can see which words are frequent or not so frequent around *vehicle* (on the right and on the left of Fig. 1 respectively), but we can also see somewhat independently which words are strongly attracted to *vehicle* and which are not (at the top and at the bottom respectively), and if we amalgamate co-occurrence frequency and association with a particular emphasis on words that score highly on both, the results look much more convincing.

Thus, in some way, such results seem to be impressive and to have a huge amount of post-hoc appeal. If one accepts the premises that

1. representative and balanced corpora reflect ordinary speakers' usage;
2. such usage is key to identifying the ordinary meaning of a term, which are both assumptions that work at the intersection of law and corpus linguistics has relied on,

then it's easy to look at them and feel that such corpus data are insightful. However, whatever quantitative sophistication notwithstanding, such results also come with problems, some of which have to do with methodological choices while others involve the specific argumentation offered; we will deal with both in the next section.

Table 2
Collocates most frequent with and attracted to vehicle(s) in COHA 1950-1959.

RANK	WORD	ASSOC	RANK	WORD	ASSOC	RANK	WORD	FREQ
1	tracked	1.116928	18	1936	1.063057	35	string	1.051211
2	farman	1.100597	19	tanks	1.062885	36	trains	1.050478
3	launching	1.095622	20	efficient	1.062818	37	paying	1.050370
4	accidents	1.086459	21	shelter	1.062504	38	affected	1.049981
5	orbit	1.084626	22	motors	1.060774	39	buses	1.049951
6	armored	1.081269	23	Mobile	1.057945	40	operate	1.049245
7	self-propelled	1.076680	24	highways	1.056838	41	compulsory	1.048933
8	legislators	1.074257	25	drivers	1.056651	42	tactical	1.047711
9	curve	1.072617	26	crossing	1.055491	43	controlled	1.046864
10	vessels	1.071580	27	frantic	1.055264	44	properly	1.046656
11	gasoline	1.069133	28	passenger	1.053227	45	satellite	1.046398
12	carrier	1.069071	29	provides	1.053212	46	wagon	1.044407
13	automobiles	1.066392	30	spokesman	1.052890	47	300000	1.044307
14	substitute	1.066262	31	operates	1.052069	48	steam	1.043986
15	freight	1.065847	32	revenue	1.051922	49	mines	1.043483
16	traveling	1.065063	33	craft	1.051698	50	tail	1.043477
17	testing	1.064678	34	thrust	1.051312			

3. Issues in collocate interpretation

3.1. Methodological issues

A first set of issues and problems has to do with methodological issues arising in the identification of what Lee & Mouritsen call "common" collocations:

- How is "common" defined in this analysis? Lee & Mouritsen's discussion cited above uses the term *common* a few times and seems to operationalize "common" with whatever COHA's web interface returns, but (i) that already conflates various notions because we have seen that COHA's collocate rankings are based on both co-occurrence frequency and the association measure PMI and (ii) that also muddies the water even further, given that notions such as 'prototypicality' might also contribute to the notion of 'commonness' (see Solan, 2020).
- How do we measure all the dimensions of information we decided to include? Lee & Mouritsen used PMI to find "common collocates" when, a bit ironically, PMI is an association measure that often ranks infrequent words very highly (see, e.g., Bourma, 2009), while our own analysis of the 1950s data uses a combination of (co-occurrence) frequency and a different association measure.
- How do we account for dispersion? The word *vehicle* is attested in 322 files of the 11,935 files of the 1950s COHA decade (i.e., ≈2.7 %). But all occurrences of the highly-ranked word *Farman* in the context of vehicle are in just a single file (i.e., ≈0.31 % of the ≈2.7 %).

However, the bigger issue with Lee & Mouritsen's discussion is concerned with semantic, rather than corpus-linguistic or statistical, issues, which is something we turn to now.

3.2. Argumentative issues

One problem is that Lee & Mouritsen's interpretation is selective in a way that could seem a bit self-serving because their argumentation is based on the logic that a collocate *C* frequently co-occurring with a node word of interest *N* means that *C* provides crucial information about the commonness (? frequency? ordinariness? typicality? …) of *N*. However, not every *pre*sence of a collocate is interpreted the same way and neither is every *ab*sence.

For instance, the presence of many car-related collocates (*motor, cars, owners, fuel, driving,* …) is used to argue that the 'automotive' sense of *vehicle* is the most common one. But if the strong presence of collocates is used to argue in favor of a certain sense of *vehicle* being the most "common", why is it that they do not also use the fact that *electric* and *plug-in* are collocates #1 and #3 of *vehicle* in the NOW corpus to argue that the most common sense of *vehicle* is an all-electric or partially electric/plug-in hybrid? We think *electric* is used before *vehicle* as frequently as it is precisely because the prototypical vehicle does *not* (yet) entail 'electric'! And how does their argument fit with the fact that we will hardly ever find $round_C$ $wheel_N$ or $quadruped_C$ $horse_N$ even though, clearly, the collocates *C* denote extremely common and prototypical features of the node words *N*! Instead, we will find $front_C$ $wheel_N$ and $brown_C$ $horse_N$ because these collocates *C* again denote things that are *not* already entailed by the node words *N*. That is to say, (i) collocates sometimes co-occur with a node word a lot because they denote something that is not (yet) a common or prototypical characteristic of the node word (as with *electric vehicle*), and (ii) extremely common and prototypical characteristics of a node word are often not at all reflected in collocates (as with *round wheel*).

Similar problems arise from the relative absence of collocates. For instance, the relative absence of aircraft- or bicycle-related collocates (in their data) is used to allude (though not commit) to the possibility that airplanes and bicycles are not (frequent? common? prototypical? ordinary?) vehicles, but why is the absence of *tire(s)* or *steering* not also used to argue that the most common sense of *vehicle* is one without tires/wheels and/or a steering wheel? In other words, and we believe this to be a very important point, collocates highlight **dimensions of meaning** relevant to the meaning of a node word – *electric, plug-in, charging*, but also *fuel* and *gasoline* highlight the dimensions 'engine as a means of propulsion' of *vehicle* – but they are in fact not useful for determining the exact **values of a word (or its referential prototype) on these dimensions of meaning**:[9] Does the prototype of a vehicle have an electric motor or a fuel/gasoline-powered motor? The collocates don't answer that question because collocates supporting each kind of vehicle are among the top. Thus, the use of collocations can be quite problematic if not in fact useless, in particular because one needs to distinguish very carefully whether, the collocates are used in entity- or feature-based reasoning, something that Lee and Mouritsen (2018) do not distinguish well and that Lee and Mouritsen (2021) only begin to discuss.

All this also holds true for the fairly recent turn towards word embeddings, methods like word2vec, GloVe, fastText, ELMo, BERT, and others that have taken corpus linguistics and esp. computational linguistics by storm.[10] These methods offer a degree of precision and power that often supersedes that of at least the simplest collocational approaches; in the context of the present discussion, they can all be used as a collocations-on-steroids kind of tool. If one downloads one of the largest freely available models from the GloVe website (https://nlp.stanford.edu/data/glove.840B.300d.zip, a model that has been trained on 840 billion words from a web crawl using the methods discussed in Pennington et al., 2014) and then takes words of interest – e.g., *vehicle* and *vehicles* – one can find the 50 words most similar to the combination of them, which are shown in Table 3:[11]

Also very impressive, yes, clearly extremely vehicle-related, and very suggestive of 'automobile' being a 'common' sense of *vehicle*, but all the computational sophistication in the world does not render this immune against the argumentative problems raised above, and in fact both approaches come with another, more fundamental shortcoming, which we discuss now.

4. The bigger problem

4.1. Extensionalist meaning/definitions

The much more fundamental problem might be the underlying approach to definitions in particular and meaning in general that is exhibited by nearly all work in the law-and-corpus linguistics (LCL) corner of the field of legal interpretation (and especially so by avowed textualists). Nearly all such work – in fact, much work in general corpus semantics – is based, virtually always implicitly, on an approach to meaning that among linguists/semanticists would be called an **extensional approach** to meaning, or an approach to meaning based on **reference** (see Textor, 2011). In such approaches, concepts/categories are 'defined by list', i.e. by providing a list of exemplars that instantiate a concept, exemplify a category, or are referred to by a term. As we said elsewhere, [w]ith an extensional approach to meaning, the interpreter must sort through the collocation, concordance, and other data and make a determination about whether the producers of the texts being

[9] This should be especially relevant to corpus linguists who endorse the strongest version of Firth's 'you shall know a word by the company it keeps', which this shows to be too vague.

[10] These methods, while quite different in some of their underpinnings, are all machine-learning methods involving so-called neural networks/transformers that are based/trained on very large amounts of language data to fulfill many different natural language processing tasks; see Gries (2021) or Choi (2023) for applications in legal contexts.

[11] R code to run this analysis and all the case studies from Sections 5 and 6 is provided at https://www.stgries.info/research/2024_STG-BGS-KT_IntensionalistLCL_ACorpLing.r.

Table 3
Collocates most similar to vehicle(s) in 840B words of webcrawl data.

RANK	WORD	COSINE	RANK	WORD	COSINE	RANK	WORD	COSINE
1	vehicles	0.102678	18	Jeep	0.433788	35	Ford	0.495128
2	vehicle	0.119348	19	Hummer	0.453008	36	motor	0.496075
3	Vehicle	0.131827	20	vans	0.459445	37	minivan	0.497995
4	Vehicles	0.136627	21	driving	0.463261	38	Dodge	0.499722
5	cars	0.293984	22	Toyota	0.467617	39	Automobiles	0.500275
6	car	0.333747	23	passenger	0.469467	40	Hyundai	0.504281
7	SUV	0.341641	24	Motor	0.470363	41	Driving	0.505171
8	Cars	0.345422	25	Nissan	0.471827	42	Transportation	0.505231
9	automobiles	0.353168	26	motorcycles	0.477564	43	mileage	0.505398
10	automobile	0.358327	27	towing	0.479893	44	Prius	0.506942
11	Car	0.361752	28	parked	0.482190	45	motorcycle	0.507354
12	truck	0.364599	29	Chevrolet	0.482647	46	Chrysler	0.507437
13	trucks	0.368404	30	Volkswagen	0.485307	47	dealership	0.511247
14	Truck	0.387287	31	Volvo	0.487363	48	auto	0.513288
15	Automobile	0.412835	32	Automotive	0.488630	49	Passenger	0.514218
16	SUVs	0.413412	33	Rover	0.491758	50	off-road	0.515937
17	Trucks	0.420495	34	autos	0.494139			

searched demonstrated a belief (even if indirectly) that some concept falls within the scope of the category at issue. This determination will thus be based on the 'evaluation of some kind of frequencies.' There must therefore be some standard above which the frequency of instances can be said to represent category membership. If, for example, airplanes are not mentioned in the same contexts as "vehicles," the interpreter might conclude that airplanes likely do not fall under the 'vehicle' concept. But frequencies of co-occurrence alone are an insufficient basis on which to determine category membership." (Eskridge, Slocum, & Gries, 2021:1528)

While the last sentence of that quote again points to the problem of what frequencies of co-occurrence really mean, the more important and general problem is that any such approach runs into both (i) the above problem of having to decide – in a rigorous and principled way that prohibits motivated reasoning – which presences and absences of collocates 'mean something' and (ii) the problem of having to deal with cases when language and/or society change in ways that make it hard or impossible and/or controversial to decide on

- whether a certain term still applies to certain category members;
- whether a new object falls in the category of an existing term in a statute.

The latter case is the one more relevant here: For the hypothetical 1950s law prohibiting vehicles from the park, is a traditional wheelchair a vehicle or is a motorized one (which existed back then)? Is a Segway or a skateboard, or a motorized skateboard (onewheel or twowheel) (which did not exist back then)? How do we handle such cases in a principled, non-biased way? Methodologically, we won't be able to find (m)any hits in the 1950s data for terms whose referents did not exist in the 1950s, and, conceptually, textualists haven't always been clear on how to resolve such questions in a principled manner that stays true both to legal requirements and what linguists/semanticists know about (word) meaning.

For example, Justice Scalia, the godfather of original public meaning, was open to including new entities within the meaning of old legal provisions (e.g., he joined Justices Thomas and Alito in thinking that the Second Amendment protects thousands of firearms that did not exist in 1791 or 1868), but also approached the question of the meaning of *vehicle* in a way that is, ultimately, arbitrary and idiosyncratic: Scalia and Garner (2012) claim that their 'fair reading' method "requires aptitude in language, sound judgment, the suppression of personal preferences regarding the outcome, and, with older texts, historical linguistic research, plus it is also said to require an ability to comprehend the purpose of the text", However, it is neither clear how "aptitude in language" is operationalized (or that Scalia and Garner possess it, at least in regard to ordinary language, given that they are not exactly 'ordinary speakers'), nor do we know how they, or any judge for that matter, would be reliably able or even willing to "suppress personal preferences regarding the outcome" or would be competent in "historical linguistic research"; and of course it is even less clear how this approach would fare when it comes to deciding on whether a new object falls in the category of an existing term in a statute passed decades or longer ago.

When it comes to *vehicle* in particular, Scalia and Garner (2012) simply assert that the sense of *vehicle* relevant to a hypothetical sign "No person may bring a vehicle into the park." is 'sizable wheeled conveyance', but they do so

- without defining *sizable* and without explaining why bicycles or tricycles do not count (freight trikes certainly are sizable wheeled conveyances and can be much more sizable than, say, Segways, which, according to Scalia & Garner, are excluded from parks); and
- without explaining whether or not this is supposed to mean that vehicles using only continuous tracks as a means of propulsion (like tanks) or vehicles using both wheels and continuous tracks (like certain kinds of snowmobiles) are vehicles or not.

Clearly, we need something more principled than semantics by fiat, but also something that is more versatile in accommodating change in language (as when the meanings of words change over time) or change in the world (as when new objects or concepts arise for which one needs to establish how they relate to existing ones); for legal interpretation purposes, we suggest that intensional approaches to meanings are useful and, ultimately, indispensable.

4.2. Intensionalist meaning/definitions

The complementary approach to an extensional/reference approach to meaning is an **intensional approach to meaning**, or an approach to meaning based on **sense** (see, again, Textor, 2011). In such approaches, concepts/categories are 'defined by definition', and two approaches are most widely used and discussed. The first of these is the classical Aristotelian approach, which historically has meant providing a set of **necessary** ('only if') conditions that are jointly **sufficient** ('whenever'). To provide a classic example:

1. *only if* someone is male, and
2. *only if* someone is an adult, and
3. *only if* someone is unmarried, and

4. *only if* someone has never been married before, and
5. *whenever* conditions 1 to 4 hold, that someone is a bachelor.[12]

In this classical Aristotelian approach, conditions, or features/criteria, are typically all binary yes/no-features and, thus, all equally relevant when it comes to determining category membership – specific contextual/cultural features and/or their typicality/applicability have often not been considered, which makes the approach appear much less useful for capturing ordinary meaning in the sense of 'what an ordinary speaker would consider a definition/category to be': Not only do necessary-cum-sufficient definitions often fail to provide a good picture of what ordinary people think/know – who would seriously think defining humans as 'featherless bipeds' captures our ordinary knowledge of humans? – they also fail to capture fairly straightforward nuances of meaning; for example, the Pope and Tarzan meet all traditional criteria for bachelorhood but are not usually considered bachelors, see Lakoff (1987:68–70).

A much more promising option would, therefore, be an intensional approach based on **prototype theory** (see Taylor, 2004; 2011), a by design cognitively more realistic approach in which the importance of features can be graded or, more specifically, quantified via the notion of **cue validity**, which we mentioned earlier in Section 1.2. On top of the more abstract definition provided above, this notion of cue validity has received two quantifiable definitions, which have the same goal but differ in their calculation:

- an early definition in the context of prototype theory (e.g., Rosch, 1978): $p(C|f)$, according to which cue validity is the (conditional) probability of membership in category C given the presence of a feature f;
- a later definition in the context of the Competition Model (Bates & MacWhinney, 1982; 1989; MacWhinney, 2005), according to which cue validity is the product of
  - cue reliability ($p(C|f)$ (i.e. the above definition of cue validity) and
  - cue availability ($p(f)$).

For example, to compute either version of the cue validity of the feature 'has a beak' to the category 'bird', one needs to look at entities that are birds and at entities that are non-birds to see whether they have beaks or not and set up a table like Table 4 (i.e., one needs to set up a table like this for every feature under consideration). With this, the early version of cue validity would amount to $^{100}/_{102}$ while the later version would amount to $^{100}/_{102}\times^{102}/_{200}$; we will improve on this approach below.

This approach, first raised in connection with (psycho)linguistic approaches to legal interpretation by Solan (2006), would be a fantastic addition to LCL approaches to meaning because prototype theory has largely been concerned with ordinary meaning (as opposed to, say, classical necessary-cum-sufficient features definitions that are often applied in non-ordinary, technical contexts), and it has been very successful at explaining and modeling everyday categorization by ordinary speakers, the exact target of the ordinary meaning doctrine. However, intensional approaches and prototype theory, while sometimes alluded to in LCL approaches, are as yet hardly ever fully appreciated let alone used. This is likely because

- most legal practitioners lack the proper knowledge of (the stages of) prototype theory simply because law school curricula do not include training in linguistic semantics;
- as we will see below, it is much more difficult to implement because it is a social science kind of approach to something as volatile and intersubjective as meaning, which can require extensive manual analysis with annotation and follow-up diagnostics (e.g. for interrater reliability) and/or very high-powered statistical approaches using the kind of word embeddings approaches from above.

Table 4
The correlation of 'has a beak' and 'birdness'.

	category: *bird*	category: not *bird*	Sum
beak: *yes*	100	2	102
beak: *no*	0	98	98
Sum	100	100	200

For a cognitively more realistic prototype approach from an intensional perspective, it is necessary to have a list of features, or properties, conditions, or criteria, that one considers useful or even indispensable for the definition of a word/term (such as *vehicle*). In an LCL context, these could be based on many things, including but not limited to, dictionaries, regulations from official bodies (e.g., the DMV), criteria explicitly stated in previous legal opinions, survey research with naïve/untrained subjects. Possible features for the category 'vehicle' might include

- WHL: has wheels;[13]
- ENG: has an engine/motor;
- ICE: specifically, has an internal combustion engine;
- STR: has a steering wheel;
- PPG: can transport people and/or goods/freight;[14]
- ROA: designed for use on roads (as opposed to, say, rails or waterways).

Possible additional criteria are ones that include questions of whether part of being a vehicle needs to involve something about a vehicle requiring licensing or something about the speed at which vehicles travel, but we will not deal with those here – our point is not the specific definition of *vehicle* one needs for a specific case but more generally theoretical and methodological. Nevertheless, let us point out that the quality of the list of criteria is less important than one might think because the empirical steps to be discussed presently will manage to see whether a proposed criterion is useful or not. That also means, the approach to be discussed cannot be gamed: If a lawyer were to force into an analysis a criterion generally considered useless but that they hope will decide a case in their favor, that criterion will likely be useless in the empirical part of our method and, thus, be downgraded and unable to sway the analysis.

5. A intensional approach with manual annotation

5.1. Annotation of three concordances

Let's use these features to exemplify what an intensional strategy to *vehicle* using prototype theory, features and their cue validity, and a corpus-based approach could look like. Remember that, according to

[12] This classical Aristotelian approach using necessary and sufficient conditions is also mentioned by Lee and Mouritsen (2018:840) – "[i]f we accept that the necessary and sufficient conditions of *vehicle* are '[a]ny means of carriage, conveyance, or transport' – but it is not utilized further in their discussion; similarly, Lee and Mouritsen (2021) briefly discuss necessary and sufficient conditions of 'vehicle' (e.g., p. 339f.) but do not advance a full theory or method of how features, especially non-binary features, might be utilized.

[13] We are simplifying here – a more accurate criterion might be 'has (and uses) wheels as its primary means of locomotion'.

[14] This list is partially inspired by McBoyle v. U.S. (1931) and Lee & Mouritsen, 2021:339).

what we said above, to quantify the cue validity of a feature *f* for a category *C* (such as 'vehicle'), we need to know

- how many members of the category *C* have each feature *f*;
- how many non-members of the category *C* do not have the feature *f*,

which in turn means we need examples of vehicles and examples of non-vehicles for which we can count the presences and absences of whatever features $f_{1\text{-}n}$ we want to consider in our analysis.[15] Accordingly, **step 1** is to select features for our analysis and we will use the six features listed earlier: WHL, ENG, ICE, STR, PPG, and ROA.

Step 2 would be to generate a concordance of *vehicle* (plus whatever morphological variants one would like to include) so that we have examples of members of the category *C* in our data. Given the hugely laborious steps we will discuss in a moment, we are not in a resource position to demonstrate this with actual data but for now demonstrate this instead with simulated concordance output that has a column for all matches of *vehicle* or *vehicles* in the corpus, together with preceding context on the left and subsequent context on the right; see Table 5:

Table 5
Concordance 1: things we know to be vehicles.

PRECEDING	MATCH	SUBSEQUENT
The police pulled the	vehicle	over for speeding. The driver got out of the car
Tanks are just one example of military	vehicles	extremely expensive to maintain
This E400 is an extraordinary	vehicle	on sale right now
…	vehicle	…
…	vehicle	…
…	vehicle	…
…	vehicle	…
…	vehicle	…
…	vehicle	…
…	vehicle	…

Step 3 would be to add to this spreadsheet

- a classification/label of the kind of vehicle each concordance line instantiates (here, a column called THING);
- a column for each feature $f_{1\text{-}n}$ we've decided to use in step 1, where annotators read each concordance line and decide whether the use of *vehicle* in the concordance line involves each of our features $f_{1\text{-}n}$. The simplest possible way to do this would be in a binary *yes*/1 vs. *no*/0 fashion (but see below). In Table 6 we simulate the former:

That is to say, the first concordance line refers to a car with wheels, an internal combustion engine, and a steering wheel which can transport people and goods, but doesn't travel on roads.[16]

Step 4 consists of computing from this for each feature how much on average it characterizes all vehicles. In other words, this would be the percentage of 1′s out of each column for a feature: the column ENG has eight 1′s and two 0′s so its score would be 0.8, the column WHL has six 1′s and four 0′s so its score would be 0.6, etc.

Step 5 is the most cumbersome part. So far, we obtained data for members of *C*, i.e. for things that *are* members of category *C*, 'vehicle'. But to compute cue validities, we now also need a 'control group' of sorts, i.e. we need non-members of *C*, non-vehicles, but the same kind of feature information for them as for vehicles. It might be best to develop such control group data in two steps.

A first part of our 'control group' could consist of a concordance of words that are co-hyponyms of *vehicle*, i.e. not vehicles but members of a superordinate category of *vehicle* (e.g., 'concrete object' or 'man-made object'), and, maybe ideally, that concordance would consist of as many hits as we had for *vehicle* (i.e., here another 10). This concordance could be generated, for example, from randomly-chosen seed words or from randomly-chosen nouns that (manual?) inspection indicates are concrete and/or man-made objects (the superordinate category of 'vehicle') but probably mostly not vehicles (the target category). Then, this first control group, the things that are not likely to be vehicles but other things – hence *something* in the column MATCH – would then be annotated for the same six features whose cue validity for 'vehicle' we are interested in; we are filling the right six columns with random 0s and 1s to get Table 7.

The second part of our 'control group' could consist of a concordance of words 'instantiating the features' $f_{1\text{-}n}$ we decided are relevant to *vehicle*. In the present context, these could be *wheel(s)* and/or *tire(s)* for WHL, *engine(s)* and/or *motor(s)* for ENG, *combustion* and/or *gas* for ICE, *steering* and/or *wheel* for STR, *person(s)* and/or *good(s)* and/or *freight* and/or *transport* and/or *payload* for PPG, and *road(s)* and/or *street(s)* for ROA. Again we would get matches with the preceding and their subsequent context in the usual format and then this second control group, the things referred to in the context of *vehicle*'s features, would then be annotated for three things (like in Table 8):

- the column THING, i.e. what the thing represented in the concordance line is: for example,

  – our first concordance line might be a match for the search term *engine* which referred to a yacht;
  – our second concordance line might be a match for the search term *wheel* which referred to a wheel of a hospital bed;
  – our third concordance line might be a match for the search term *wheel* which referred to a desk;

- the column MATCH, i.e. whether we consider them vehicles or not, and we note this in the column MATCH;

  – our first concordance line referred to a yacht, which we might consider a vehicle;
  – our second concordance line referred to a hospital bed, which we might not consider a vehicle;
  – our third concordance line referred to a desk, which we do not consider a vehicle;

- for the same six features whose cue validity for vehicle we are interested in, and we are again using largely random 0s and 1s for that.

5.2. Two important side remarks or clarifications

First, note that we do not actually need to use binary 0s and 1s only for any of these annotations. Theoretically, we could incorporate **interrater differences** into the analysis. For instance, if every concordance line was annotated by multiple annotators, then we could use the proportion of annotators who said that a certain feature is present in a certain concordance line rather than just a binary 0 or 1. Recall that above we said "a yacht, which we *might* consider a vehicle". That is, if 10 annotators read the first line of the last data frame just shown, …

[15] This strategy might remind readers of the very important distinction of semasiological vs. onomasiological approaches, which is extremely relevant to LCL. Semasiology is concerned with $p(\text{meaning}|\text{term})$ while onomasiology is concerned with $p(\text{term}|\text{meaning})$. In legal corpus linguistics circles, this has been discussed most impactfully by Solan & Gales (2017:1351ff.) (though somewhat surprisingly not with these established semantics terms but as "double dissociation").

[16] Recall that these are pseudorandom simulated data; this example might involve a kind of car that may only be used on the limited grounds of a business but not on general public roads, e.g. a prototype of a self-driving car which is being tested on the property of a development company but is not yet licensed to travel on public roads

Table 6
Concordance 1: things we know to be vehicles (with feature annotation).

PREC	MATCH	SUBS	THING	WHL	ENG	ICE	STR	PPG	ROA
[…]	vehicle	[…]	car	1	1	1	1	1	0
[…]	vehicle	[…]	tank	0	1	1	0	1	1
[…]	vehicle	[…]	car	1	1	1	0	1	1
[…]	vehicle	[…]	car	1	1	1	1	1	0
[…]	vehicle	[…]	rocket	0	1	1	1	0	1
[…]	vehicle	[…]	car	1	1	1	1	1	1
[…]	vehicle	[…]	bus	1	1	0	0	1	0
[…]	vehicle	[…]	truck	1	1	1	1	1	1
[…]	vehicle	[…]	missile	0	0	0	0	1	1
[…]	vehicle	[…]	sail boat	0	0	0	1	0	0

Table 7
Concordance 2a: things we know to be non-vehicles but concrete objects (with feature annotation).

PREC	MATCH	SUBS	THING	WHL	ENG	ICE	STR	PPG	ROA
[…]	something	[…]	pen	1	0	0	1	0	1
[…]	something	[…]	computer	1	1	0	0	0	0
[…]	something	[…]	desk	0	0	0	0	1	1
[…]	something	[…]	table	1	1	1	0	0	0
[…]	something	[…]	toilet	0	0	0	1	1	1
[…]	something	[…]	postcard	1	0	1	0	0	1
[…]	something	[…]	pencil	0	0	0	0	1	0
[…]	something	[…]	pacifier	1	1	1	0	0	0
[…]	something	[…]	street sign	0	0	0	0	0	0
[…]	something	[…]	waffle iron	1	0	0	0	1	1

Table 8
Concordance 2b: things that may be (non-)vehicles and are related to features $f_{1\text{-}n}$ (with feature annotation).[17]

PREC	MATCH	SUBS	THING	WHL	ENG	ICE	STR	PPG	ROA
[…]	something	[…]	yacht	1	0	0	1	0	1
[…]	something	[…]	hospital bed	1	1	0	0	0	0
[…]	something	[…]	desk	1	0	0	0	1	1
[…]	something	[…]	baby stroller	1	1	1	0	0	0
[…]	something	[…]	baby stroller	1	1	0	1	0	1
[…]	something	[…]	shopping cart	1	1	1	0	0	1
[…]	something	[…]	pencil	0	0	1	0	1	0
[…]	something	[…]	forklift	1	1	1	1	0	0
[…]	something	[…]	penc. sharpener	0	0	0	1	1	0
[…]	something	[…]	hospital bed	1	0	0	0	1	1

Table 9
Concordance 2b, line 1: things that may be (non-)vehicles and are related to features $f_{1\text{-}n}$ (with feature annotation).

PREC	MATCH	SUBS	THING	WHEELS	ENG	INTCOMB ENG	STEER WHEEL	PPLn GOODS	ON ROADS
[…]	vehicle	[…]	yacht	1	0	0	1	0	1

(Table 9).

… but two of them thought the yacht *did have* an internal combustion engine and one person did not think the yacht had a steering wheel, this could be reflected in the numeric values like in Table 10.

Even more powerfully perhaps, we could incorporate **annotator uncertainty** into the analysis. An annotator could be offered the option of 0/*no* and 1/*yes*, but also every number in between as in a probabilistic judgment. That is, if 10 annotators read the first line of the last data frame and nine said they are absolutely certain that the yacht had no engine at all, but one said they are on the fence (50 % undecided) that the yacht had in fact an internal combustion engine, this could be reflected in the numeric values like in Table 11.

That way, even whether something mentioned in a concordance line is a vehicle or not could be expressed in a graded fashion. Statistically speaking, the column MATCH would then become a numeric column, expressing the proportion of annotators that considered the match a vehicle or the certainty with which an annotator considered the match a vehicle – we just kept a binary vehicle-vs.-not distinction here for ease of exposition. Thus, this approach is extremely flexible and can cover both disagreement between, and uncertainty of, speakers.

Before we continue, here is a second very important clarification: The concordance lines that we get from the concordancing for control group part 1 (co-hyponyms of *vehicle*) and for control group 2 (features of vehicles) – i.e. all the concordance lines represented in Tables 7 and 8 with the match *something* – will of course sometimes contain things we would consider vehicles, or likely vehicles, as well; for example, Table 8 contains an example of *forklift* as a THING, which one might consider a vehicle. This is not a problem as long as we get enough non-vehicles for the statistical analysis as well: Such examples just get added to the group of (likely) vehicles and will feature in the computation of cue validities just like everything else.

[17] Recall again that these are simulated annotation data: making up an internally coherent set of 6 features times 30 examples = 180 data points that yields comprehensible results is a daunting task we did not undertake for this largely programmatic article.

Table 10
Concordance 2b, hypothetical line 1 with annotator disagreement for three features.

PREC	MATCH	SUBS	THING	WHL	ENG	ICE	STR	PPG	ROA
[…]	vehicle	[…]	yacht	1	0.2	0.2	0.9	0	1

Table 11
Concordance 2b, hypothetical line 1 with annotator uncertainty for two features.

PREC	MATCH	SUBS	THING	WHL	ENG	ICE	STR	PPG	ROA
[…]	vehicle	[…]	yacht	1	0.05	0.05	1	0	1

5.3. The computation of cue validities

Now that the annotation of all concordance lines has been completed, we turn to how to compute cue validities from them. **Step 6**: we merge all data and determine for each feature its weight or predictive/discriminatory power (feature weight$_{\text{man}}$) for the category 'vehicle'. This can be done in different ways and we need research on this to explore the advantages and disadvantages of each. A first and simple approach would consist of

- computing the proportion of vehicle cases that exhibit a certain feature;
- computing the proportion of non-vehicle cases that exhibit a certain feature;
- computing the difference between the two for each feature.

These differences can theoretically range from –1 to +1 and the more a feature is present in vehicles but also absent in non-vehicles – exactly how we explained cue validity in Section 1.2 above – the higher the value of these differences will be. If we apply this to the present pseudorandom data, we obtain the following ranking, according to which, of the six features we considered, STR has the highest predictive power for something being a vehicle whereas WHL has the lowest; see Table 12.

A second possibility could be to use a more technical approach and determine how much each feature's presence predicts vehicle-ness with a machine learning method such as random forests (see Gries, 2021: Section 7.2, James et al., 2021: Section 8.2.2). In other words, a forest would be tasked to predict the column MATCH, i.e. whether something is a vehicle or not, based on the presences and absences of all six features in all the concordance lines and would return so-called variable importance scores that represent how important each feature is for that classification/prediction task (Table 13). If we apply this to the present pseudorandom data, we obtain the following ranking and, with this approach, PPG now has the highest predictive power for something being a vehicle. While the results of the two methods differ a bit, their overall rank-order correlation is quite high (Spearman's $\rho\approx0.841$).

The final step, **step 7**, would be to 'assign those features back' to all instances in the annotated corpus data, meaning we weigh each feature in the annotated corpus instances by its feature weight$_{\text{man}}$. Thus, the data frame containing the binary annotation for each feature for all data, the combination of Tables 6–8, here shown in Table 14 with its first six rows, …

… becomes a data frame where each 0 remains 0 and each 1, i.e. presence of a feature, is weighted by the importance of the feature; we use the feature weights from Table 12 and show again only the first six rows in Table 15:

This is because we can then aggregate the weighted features per THING so that we obtain, for each THING, an average value of each feature. For example, Table 15 contains all four instances of *car* in our made-up data and the aggregation – with a mean to accommodate the fact that some things (*car*) will be more frequent than others (*truck*) – converts them into the following results; note in particular how the value for STR for *car* is the average of three times 0.3800905 and one time 0:

- WHL for *car*: 0.04524887;
- ENG for *car*: 0.280543;
- ICE for *car*: 0.2624434;
- STR for *car*: 0.2850679;
- PPG for *car*: 0.280543;
- ROA for *car*: 0.07239819.

The vehicle-ness score$_{\text{man}}$ for *car* is then the mean of these aggregated features, and Table 16 shows the results for all levels of THING:

That means, given our six criteria for vehicle-ness and the annotation of corpus data containing both vehicles and non-vehicles, we now have a way to quantify how vehicle-y each 'thing' in our concordance data is, and it is done in a way that has three huge advantages (esp. over, say, Scalia & Garner's approach):

1. it is based on how many ordinary speakers use language in real-life naturalistic settings (namely the concordance data) rather than on what few, and arguably unrepresentative, speakers such as SCOTUS Justices claim about what a word means;[18]
2. the annotation is based on ordinary speakers who are disinterested with regard to the issue at hand: they just annotate concordance lines and don't even know what the issue at stake is, which rules out motivated reasoning on the part of anyone involved;
3. it is cognitively and linguistically well-founded in how it *directly* operationalizes the notion of cue validity from prototype theory our vehicle-ness scores depend on how predictive each feature is for vehicles across a wide range of corpus examples and other nouns.

5.4. Words that are not in the training data

There is a third huge advantage: Now that we know each feature's cue validity for 'vehicle', i.e. its importance for membership in the category 'vehicle', we can use this to decide whether words not in the training data, like a traditional wheelchair, a motorized wheelchair, or a golf cart, are vehicles. First, we would classify each with regard to the same six features we used in the analysis so far, and we could do generically as here in Table 17 or more probabilistically on corpus examples.

Then we compute for each object a vehicle-ness$_{\text{man}}$ score as before: We consider the features that these candidate words exhibit weighted by the features' cue validities/importances for the category 'vehicle'. The result one would obtain from this is that a golf cart scores a value of 0.164, a motorized wheelchair score 0.101, and a traditional wheelchair scores 0.054 (0.28054299 (PPG) + 0.04524887 (WHL) / 6 criteria).

We can use these scores in several ways: First, we can see where each of the three new items falls on the continuum (recall, based on pseudorandom data!) from the least vehicle-y thing (a street sign) to the most

[18] See Eskridge & Nourse (2021:1728) for a discussion of Scalia's applications of textualism.

Table 12
Feature $weights_{man}$ as determined by differences between proportions.

STR	ENG	PPG	ICE	ROA	WHL
0.38009050	0.28054299	0.28054299	0.26244344	0.14479638	0.04524887

Table 13
Feature $weights_{man}$ as determined by variable importances.

PPG	STR	ENG	ICE	WHL	ROA
2.5591951	2.3473575	1.7355744	1.2585811	1.0687770	0.9409168

Table 14
All concordance lines with all features (unweighted).

PREC	MATCH	SUBS	THING	WHL	ENG	ICE	STR	PPG	ROA
[…]	vehicle	[…]	car	1	1	1	1	1	0
[…]	vehicle	[…]	tank	0	1	1	0	1	1
[…]	vehicle	[…]	car	1	1	1	0	1	1
[…]	vehicle	[…]	car	1	1	1	1	1	0
[…]	vehicle	[…]	rocket	0	1	1	1	0	1
[…]	vehicle	[…]	car	1	1	1	1	1	1

Table 15
All concordance lines with all features (weighted).

PREC	MATCH	SUBS	THING	WHL	ENG	ICE	STR	PPG	ROA
[…]	vehicle	[…]	car	0.04524887	0.280543	0.2624434	0.3800905	0.280543	0
[…]	vehicle	[…]	tank	0	0.280543	0.2624434	0	0.280543	0.1447964
[…]	vehicle	[…]	car	0.04524887	0.280543	0.2624434	0	0.280543	0.1447964
[…]	vehicle	[…]	car	0.04524887	0.280543	0.2624434	0.3800905	0.280543	0
[…]	vehicle	[…]	rocket	0	0.280543	0.2624434	0.3800905	0	0.1447964
[…]	vehicle	[…]	car	0.04524887	0.280543	0.2624434	0.3800905	0.280543	0.1447964

Table 16
Our simulated candidate words and their vehicle-ness.

RANK	WORD	VEHICLENESS	RANK	WORD	VEHICLENESS
1	truck	0.23227753	13	yacht	0.09502262
2	car	0.20437406	14	pen	0.09502262
3	rocket	0.17797888	15	waffle iron	0.07843137
4	tank	0.16138763	16	postcard	0.07541478
5	forklift	0.16138763	17	desk	0.07466063
6	toilet	0.13423831	18	missile	0.07088989
7	shopping cart	0.12217195	19	pencil	0.06862745
8	baby stroller	0.11990950	20	hospital bed	0.06636501
9	penc. sharpener	0.11010558	21	sail boat	0.06334842
10	bus	0.10105581	22	computer	0.05429864
11	table	0.09803922	23	street sign	0
12	pacifier	0.09803922			

Table 17
Some candidate terms and their feature annotation.

THING	WHL	ENG	ICE	STR	PPG	ROA
Traditional wheelchair	1	0	0	0	1	0
Motorized wheelchair	1	1	0	0	1	0
Golf cart	1	1	0	1	1	0

vehicle-y thing (a truck) and have that inform our view of how vehicle-y each of the three new items is. For example, a golf cart scores a $vehicleness_{man}$ score that would rank really high in our list, i.e. among things that are probably fairly uncontroversial vehicles (*truck* (1), *car* (2), *tank* (4), *bus* (10)). Second, we can see how the findings may relate to vehicle-y things that have already been dealt with in legal precedent. If, for instance, there were multiple things that had lower scores than *motorized wheelchair*, but that precedent had considered vehicles, a legal practitioner might say 'then a motorized wheelchair must be considered one as well', or instead they might arrive at the conclusion that the development of the term in question necessitates revisiting precedent.

5.5. *Interim summary*

In sum, the proposed method is a way to develop the intension of a category term (like *vehicle*). This way is admittedly quite laborious and, thus, restricted to cases where expert witnesses have funds to conduct potentially large-scale annotation of hundreds (or thousands) of concordance hits by, ideally, multiple annotators, but the proposed procedure

- targets ordinary meaning, because
  - it is based on corpora, i.e. language data from natural communicative settings;
  - it can involve annotation from many ordinary and impartial speakers of a language rather than a very small number of highly unrepresentative speakers;
- is cognitively grounded in psychological and linguistic research using prototype theory and cue validity;
- is methodologically versatile in how it can accommodate uncertainty or interrater disagreement;
- is powerful in how it can be applied in situations where especially originalism textualists run into problems, namely when linguistic or

societal change has resulted in a situation where it is not clear whether a statute that was passed at a certain point of time but does not include a readily applicable intensional definition of a relevant entity does in fact cover an entity that did not exist at the time the statute was passed.

At the same time, the approach does not lock legal practitioners in: Of course a judge can still follow precedent, follow the golden rule, avoid absurdity, etc. – all the method does is

- help a legal practitioner arrive at the importance of features in a way that is compatible with the ordinary meaning doctrine and the contemporary understanding of meaning by linguists/semanticists and, if so desired,
- use the importance of feature to determine the degree of membership of something in a category in a way that allows an expert witness to use such data for a principled decision untainted by motivated reasoning.

6. An intensional approach with embeddings

Given the potentially enormous amount of manual annotation required, is there a way to do this without the inordinate amount of annotation but with the 'distributional knowledge' captured in the kind of deep learning/word embedding models that are currently all the range in large parts of corpus/computational linguistics in particular, but also linguistics in general? It seems to us that there are two ways, depending on whether the word one wants the check for vehicle-ness is part of the training data or not.

6.1. Words that are in the training model

If we want to quantify vehicle-ness $\text{scores}_{\text{emb}}$ for candidate words that *are* in the training data like

- *car* and *truck* (whose vehicle-ness score should be high);
- *forklift* and *airplane* (whose vehicle-ness scores should be lower);
- *table* and *glove* (whose vehicle-ness scores should be much lower),

we can again follow a multi-step procedure, but one that avoids the need for potentially hundreds of hours of annotation of corpus data. This also means that the following demonstrations are not based on simulated data but on actual data just like one might use in an actual case. In one such procedure, **step 1** would be to generate one object that contains deep learning/embedding models for *vehicle* and one for each word that indexes, or represents, the six features we came up with. For example, we might return to the embeddings model used in Section 3.2 above and import it into an R object called w2v.model and compute the models for the features in the same way as we computed the model for *vehicle(s)*. **Step 2** would be to also compute such embedding models for each candidate word.

Step 3 would then be to just determine how similar each candidate word (via its model from step 2) is to each feature (via its model from step 1) to get the embedding-based equivalent of cue validities. **Step 4** is simple: we again aggregate the feature-similarity scores up per candidate terms – with a mean, as before – because then candidates that exhibit many of the most important features will score highest, and indeed they do; see Table 18.

In other words, in the earlier manual-annotation approach, we derived cue validities by computing from the annotation of the concordance lines which features made the biggest differences for whether something was a vehicle or not – in the current embedding approach, we derive cue validities from the similarity of embedding models of features to embedding models of candidate terms. As one can see in these now authentic data, the results strongly support the expectations one might have and that were formulated above: the vehicle-ness $\text{scores}_{\text{emb}}$ for *car* and *truck* are indeed highest, those for *forklift* and *airplane* are indeed lower, and those for *table* and *glove* are even lower. And one can use these vehicle-ness $\text{scores}_{\text{emb}}$ just like the vehicle-ness $\text{scores}_{\text{man}}$: We can decide for existing items how vehicle-y they are and we can infer the importance of features for the category (i.e., cue validities), but we can now do so without requiring huge amounts of manual annotation. However, what about the third application we discussed above for the manual approach, the application of this approach to words that are not in the training model? This situation can be addressed as follows.

6.2. Words that are not in the training model

For words that are not in the training data, a different approach avoiding manual annotation of concordance lines is available that still utilizes top-down knowledge of features. We can explore this by just pretending for a moment that the six candidate words from above – *car, truck, forklift, airplane, table*, and *glove* – were *not* attested in the training data. How would we proceed then? One approach is actually very simple: **Step 1** would be the same as before: we would load the same embeddings model w2v.model and we would generate the same list object with models for *vehicle* and all six features (as represented by terms indexing the features).

On to step 2. In the previous case, where the candidate terms of interest were attested in the model, steps 2 and 3 involved computing the similarity of each candidate term to the target category of 'vehicle'. Now, where we cannot do this because we are assuming the words are not attested in the data and we can therefore not compute a model for them, **step 2** means we briefly annotate every candidate term of interest with regard to the features we consider relevant to vehicle-ness (and, as above, this annotation may be binary or graded, i.e. with values in the interval of [0, 1]):

But now we need to again figure out how important each of these features is when it comes to determining a candidate term's vehicle-ness. Therefore, **step 3** is to compute, for each feature that we consider relevant to vehicle-ness, a feature weighting as above – but this time from how much the models for our six features, which we already computed in step 1 above, are correlated with the model of the category term in question, *vehicle*. The result from running the code will show that ENG, though not necessarily an internal combustion engine, and WHL are the most important of the six features we are considering. **Step 4** is then to do the same as above with the manual data: We apply these weights of features to Table 19, the annotation for each candidate word,

Table 19
Our candidate terms and their feature annotation.

Object	WHL	ENG	ICE	STR	PPG	ROA
car	1	1	1	1	1	1
truck	1	1	1	1	1	1
forklift	1	1	0	1	0	1
airplane	1	1	0	1	1	0
table	0	0	0	0	0	0
glove	0	0	0	0	0	0

Table 18
Six candidate terms and their embeddings-based vehicle-ness scores.

	car	truck	airplane	forklift	table	glove
score	0.5703084	0.5460508	0.3882583	0.3077917	0.1934678	0.1709745

Table 20
Six candidate terms and their embeddings-based vehicle-ness scores.

	car	truck	airplane	forklift	table	glove
score	0.4278753	0.4278753	0.3049727	0.2902798	0	0

and then aggregate them by averaging (Table 20).

Voilà: we get a ranking of candidate terms in terms of their vehicle-ness without any embedding model for any candidate term but by combining our annotation of candidate terms in terms of features with the similarity of each feature to the target term *vehicle*.

6.3. Segway in the 1950s COHA

As a brief final example, let us consider a question that Scalia and Garner (2012) also touch upon, namely the question of whether a Segway would be considered a vehicle, when we consider it in the context of a hypothetical statute regarding vehicles enacted in the 1950s, i.e. when Hart first discussed his famous hypothetical. This is an appropriate test case because Segways did not exist in the 1950s, meaning the word *Segway* is not attested in any corpus data covering that time period, so a court case involving Segways arising in, say, 2024 would have to determine whether the 1950s statute involving vehicles covers the only later invented category of Segways. If that court adopted the “original public meaning” approach currently popular among textualist and originalist judges in the United States, it would seek to give *vehicle* the meaning it had in 1950.

To test our approach, we first train a word2vec embeddings model on the complete COHA data for the decade 1950–1959; specifically, we used the R package wordVectors (Schmidt and Li, 2021) to train a 300-vector skip-gram model with a context window size of 4 words and a required minimum frequency of 3 and 35 iterations.

Once we have such a model, we proceed in the same way as before: **Step 1** is to create a list object that contains models for *vehicle* (our target category) and, for consistency’s sake, the same six features as before; however, now these are of course based on the relevant temporal data, the 1950s decade of COHA.

Then, **step 2** is to annotate every candidate term of interest with regard to the features we consider relevant to vehicle-ness. We use the same candidate terms and annotations as in Table 19 before but now also add Segways to the mix, which we consider to have wheels, an engine/motor that is not an internal combustion engine, a steering wheel, and which we consider to transport people and/or goods on roads (Tables 21 and 22).

Step 3 is again to compute a weighting for each feature that we consider relevant to vehicle-ness, but this time we do so from how much the models for the features are correlated with the model of the category term in question, *vehicle*. The result shows that PPG as well as ENG (and ICE to a lesser degree) are the most important of the six features we are considering in the 1950s decade of COHA.

Finally, **step 4** is then to do the same as above with the manual data: we apply these weights of features to the annotation for each candidate word and amalgamate them by averaging. The result is shown in Table 22:

Voilà: we get a ranking of all candidate terms in terms of their vehicle-ness and, to us at least, the result makes a lot of intuitive sense even though it was arrived at using a very simplistic ’featural analysis’ of candidate vehicles in general and Segways in particular: Segways rank fairly highly, below the arguably more prototypical vehicle types of cars and trucks, but also higher than the arguably less prototypical vehicle types of airplanes and forklifts. However, the by far most important conclusion in the present methodological context is the proof of concept: We succeeded in getting a vehicle-ness value for Segways that is again grounded in prototype semantics and natural communicative uses of language (as encapsulated in an embeddings model trained on the 1950s data) *even though* the corpus data do not actually contain the target word whose categorization we were interested in; in other words, some approach like this appears very promising for especially historical analyses that need to come to grips with various kinds of linguistic/societal change.

Table 21
Our candidate terms and their feature annotation.

Object	WHL	ENG	ICE	STR	PPG	ROA
Segway	1	1	0	1	1	1
car	1	1	1	1	1	1
truck	1	1	1	1	1	1
forklift	1	1	0	1	0	1
airplane	1	1	0	1	1	0
table	0	0	0	0	0	0
glove	0	0	0	0	0	0

7. Concluding remarks

Modern American legal interpretation is often textualist and originalist, and it often involves a problem with a particular structure: How should the original meaning of this (old) text apply to this new situation? For example, does a rule stipulating "no vehicles in the park" prohibit Segways from the park, when Segways did not exist at the time the rule became law? Does the rule’s original meaning prohibit drones? Does it prohibit other entities that will only be created in the future? The legal-interpretive literature has addressed these questions by shifting from "original expected expectations" to "original public meaning." We propose a similar shift for law and language, from extensional to intensional corpus linguistic approaches. Accordingly, the main two bottom lines of this paper are as follows.

First, we hope to have demonstrated that the LCL community and law and linguistics researchers more generally interested in corpus approaches to legal interpretation should refocus their attention away from extensional approaches to meaning and towards intensional approaches to meaning because it is only with intensional approaches that we can address problems that textualists regularly face when ’the world’ or ’the language’ changes. In addition, we also submit that the proposed approach can better address what has been called the ’blue pitta problem’ or the ’Nonappearance Fallacy’, the claim that the nonappearance of some use in the corpus generally indicates that this use is incompatible with ordinary meaning (as when blue pittas are completely uncontroversially birds but not mentioned as a bird in a corpus like COCA or when some corpora contain no examples of airplanes referred to as vehicles (see above and Gales & Solan, 2019; Tobia, 2021: Section I.3.E). With our feature-based method, something simply doesn’t have to be mentioned because the categorization/definition is based on intensional features, not extensional mentions.

Second, we hope to also have shown that the intensional approach to meaning can be implemented in different ways:

- one in which an additional human element (in the form of manual annotation) is involved and which allows for the incorporation of disagreements between raters/annotators and uncertainty in the judgments of raters/annotators;

Table 22
The candidate terms and their embeddings-based vehicle-ness scores.

	Car	truck	Segway	airplane	forklift	table	glove
score	0.3666656	0.3666656	0.3083438	0.2589892	0.2301910	0	0

- one in which embeddings can be used to approximate the kind of featural information required for intensional approaches to meaning; the example of whether a 1950s statute regarding vehicles that does not reference Segways would cover Segways or not seems to us to be a particularly instructive application.

As always, more exploration and validation is needed. The linguistically maybe most obvious next step is to test the current approach with more recently proposed and more context-sensitive embedding methods like BERT (Devlin et al., 2018) or ELMo (Peters et al., 2018) – while we do not anticipate that the results would change much, this is clearly an empirical question worth exploring. In addition, it is essential we test this approach for other target terms, especially more abstract ones. We see a lot of potential for other cases, for potentially any cases in which the previous extensional approach to meaning runs into the above-mentioned problems or inconsistencies or in cases where societal or language change (or vagueness and ambiguity) render traditional analysis implausible or even impossible, but also need to caution that the methods involved are highly technical and not available as off-the-shelf methods – their execution require *extensive* expertise in corpus and/or computational linguistics (including programming in Python or R and statistical methods).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aprill, E.P., 1998. Dictionary shopping in the Supreme Court. Arizona State Law J. 30, 275–336.
Bates, E., MacWhinney, B., 1982. Functionalist approaches to grammar. In: Wanner, Eric, Gleitman, Lila R. (Eds.), Language Acquisition: the State of the Art. Cambridge University Press, Cambridge, pp. 173–218.
Bates, E., MacWhinney, B., 1989. Functionalism and the competition model. In: MacWhinney, Brian, Bates, Elizabeth (Eds.), The Crosslinguistic Study of Sentence processing. Cambridge University Press, New York, NY, pp. 3–73.
Bourma, G., 2009. Normalized (pointwise) mutual information in collocation extraction. Proceed. German Soc. Comput. Linguist. Lang. Technol. 30, 31–40.
Brudney, J.J., Baum, L., 2013. Oasis or mirage: The Supreme Court's thirst for dictionaries in the Rehnquist and Roberts eras. William Mary Law Rev. 55 (2), 483–580.
Choi, J., 2023. Measuring Clarity in Legal Text. to appear. University of Chicago Law Review.
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding arXiv:1810.04805v2. https://arxiv.org/abs/1810.04805.
Eskridge, W.N., 1998. Textualism, the unknown ideal? Mich. Law Rev. 96 (6), 1509–1560.
Eskridge, W.N., 2016. Interpreting Law: A Primer on How to Read Statutes and the Constitution. Foundation Press, St. Paul, MN.
Eskridge, J., William, N., Nourse, V.F., 2021. Textual gerrymandering: The eclipse of Republican government in an era of statutory populism. New York Univ. Law Rev. 96 (6), 1718–1825.
Eskridge, W.N., Slocum, B.G., Gries, S.T., 2021. The meaning of *sex*: Dynamic words, novel applications, and original public meaning. Mich. Law Rev. 119 (7), 1503–1580.
Gales, T., Solan, L.M., 2019. Revisiting a classic problem in statutory interpretation: is a minister a laborer? Georgia Stqte Univ. Law Rev. 36 (5), 491–533.
Gries, S.T., 2019a. 15 years of collostructions: some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). Int. J. Corp. Linguist. 24 (3), 385–412.
Gries, S.T., 2019b. Ten Lectures on Corpus-Linguistic Approaches: Applications for Usage-Based and Psycholinguistic Research. Brill, Leiden & Boston.
Gries, S.T., 2020. Corpus approaches to ordinary meaning in legal interpretation. In: Coulthard, Malcolm, May, Alison, Sousa-Silva, Rui (Eds.), The Routledge Handbook of Forensic Linguistics, 2nd ed. Routledge, New York & London, pp. 628–643.
Gries, S.T., 2021. Corpus linguistics and the law: extending the field from a statistical perspective. Brooklyn Law Rev. 86 (2), 321–356.
Hart, H.L.A., 1958. Positivism and the separation of law and morals. Harv. Law Rev. 71 (4), 593–629.
Hobbs, P., 2011. Defining the law: (Mis)using the dictionary to decide cases. Discour. Stud. 13 (3), 327–347.
Holmes, O.W., 1899. The theory of legal interpretation. Harv. Law Rev. 12 (6), 417–418.
Hutton, C., 2020. The category of ordinary meaning and its role in legal interpretation. In: Coulthard, Malcolm, May, Alison, Sousa-Silva, Rui (Eds.), The Routledge Handbook of Forensic Linguistics, 2nd ed. Routledge, New York & London, pp. 79–92.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning with Applications in R, 2nd ed. Springer, Berlin & New York.
Labov, W., 1975. Empirical foundations of linguistic theory. In: Austerlitz, Robert (Ed.), *The scope of American Linguistics. The Peter de Ridder Press, Lisse, pp. 77–133.
Lakoff, G., 1987. Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. University of Chicago Press, Chicago, IL.
Lee, T.R., Mouritsen, S.C., 2018. Judging ordinary meaning. The Yale Law J. 127, 788–879.
Lee, T.R., Mouritsen, S.C., 2021. The corpus and the critics. Univer. Chicago Law Rev. 88 (2), 275–366.
MacWhinney, B., 2005. A unified model of language acquisition. In: Kroll, Judith F., Annette, M.N.de Groot (Eds.), Handbook of Bilingualism: Psycholinguistic Approaches. Oxford University Press, Oxford, pp. 49–67.
Mouritsen, S.C., 2010. The dictionary is not a fortress: Definitional fallacies and a corpus-based approach to plain meaning. BYU Law Rev. 2010 (5), 1915–1978.
Mouritsen, S.C., 2017. Corpus linguistics in legal interpretation. Int. J. Lang. Law 6, 67–89.
Pennington, J., Socher, R., Manning, C.D., 2014. GloVe: global vectors for word representation. In: Proceedings of Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543.
Peters, M.E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, & L. Zettlemoyer. 2018. Deep contextualized word representations. arXiv:1802.05365v2, <https://arxiv.org/abs/1802.05365>.
Core Team, R, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org.
Rosch, E., 1978. Principles of categorization. In: Eleanor, Rosch, Barbara, Lloyd (Eds.), Cognition and Categorization. Lawrence Erlbaum, Hillsdale, NJ, pp. 27–48.
Scalia, A., Garner, B.A., 2012. Reading Law: The Interpretation of Legal Texts. Thomson/West, St. Paul, MN.
Schmidt, B., Li, J., 2021. wordVectors: Tools for creating and analyzing vector-space models of texts.http://github.com/bmschmidt/wordVectors.
Schütze, C.T., 1993. The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology. The University of Chicago Press, Chicago, IL.
Slocum, B.G., 2015. Ordinary Meaning: A theory of the Most Fundamental Principle of Legal Interpretation. The University of Chicago Press, Chicago, IL.
Solan, L.M., 2006. Brooklyn Law School Legal Studies Research Papers 30.
Solan, L.M., 2020. Corpus linguistics as a method of legal interpretation: some progress, some questions. Int. J. Semiot. Law 33 (2), 283–298.
Taylor, J.R., 2004. Linguistic Categorization: Prototypes in Linguistic Theory, 3rd ed. Clarendon Press, Oxford.
Taylor, J.R., 2011. Prototype theory. In: Maienborn, Claudia, Heusinger, Klaus von, Portner, Paul (Eds.), Semantics: An International Handbook of Natural Language Meaning. De Gruyter Mouton, Berlin & Boston, pp. 643–664. Vol. 1.
Textor, M., 2011. Frege on) Sense and reference. In: Maienborn, Claudia, Heusinger, Klaus von, Portner, Paul (Eds.), Semantics: An international handbook of natural language meaning. De Gruyter Mouton, Berlin & Boston, pp. 25–49. Vol. 1.
Tobia, K., 2021. The corpus and the courts. Univ. Chicago Law Rev. Online 88.
Tobia, K., D. Walters, B.G. Slocum. in press. Major questions, common sense? *South. Calif. Law Rev.* 97.