Examining Thick Concepts in the Legal Discourse Using Corpus Analysis

*Pascale Willemsen\*, Lucien Baumgartner, Severin Frohofer, Kevin Reuter  
(Department of Philosophy, University of Zurich, Zuerichbergstrasse 43, CH-8044 Zurich)  
\* corresponding author*

**Abstract**

This paper is super interesting. You will learn a lot. Read it. If you don’t like it, we don’t like you. And you’re a poopy-head.

# Introduction (Pascale)

In 1978, Ted Bundy stood trial for murder. To prove Bundy’s guilt, several pieces of evidence were provided, among others, a dental impression of his teeth. One of the victims showed a bite wound on her body, and it was argued that the particularities of Ted Bundy’s teeth would match it, thereby relating him to the crime. Forensic odontologists matched the wound to castings of Bundy’s teeth – a piece of evidence used for the first time in an American court. The defence attorney objected to the acceptability of this evidence, saying:

‘The evidence in this case presents many reasonable doubts. It is a sad day for your system of justice that can put a man’s life on the line because they say he has crooked teeth. How tragic it would be if a man’s life were to be taken from him because 12 people *thought* that he was probably guilty, but they were *not sure*.’

The evidence was yet accepted, and Bundy was convicted. In his statement, Judge Edward Cowart said:

‘The court finds that of both these killings were indeed heinous, atrocious, and cruel, and that they were extremely wicked, shockingly evil, vile, and the product of a design to inflict a high degree of pain and utter indifference to human life. This court independent of, but in agreement with, the advisory sentence rendered by the jury, does hereby impose the death penalty upon the defendant, Theodore Bundy.’

While both statements have been widely discussed, they do not appear to be particularly unusual in the way these parties put their widely differing standpoints. But should we not take issue with the rather strong evaluative views that are communicated here? After all, the legal and disa Shouldn’t then phrases like “sad day for our system of justice”, and “extremely wicked, shockingly evil” make us rather uncomfortable because they seem to violate the noble goal of reaching verdicts from as neutral a standpoint as possible?

Perhaps, you might say, we cherrypicked the Ted Bundy case in which Judge Cowart was a bit over the line. Perhaps, you might think, evaluative statements in extreme cases like the Ted Bundy trial are hardly avoidable but are less common in more mundane cases. Or perhaps, you say, what matters is not so much how a judge justifies a verdict, but rather that everything that happens up to the point at which a verdict is made, is expressed in a language that is as descriptive as possible.

If these points are correct, we should be able to find evidence that legal discourse is indeed more descriptive and less evaluative than public discourse. Surprisingly, little evidence has been collected in support of that view. Those studies that have investigated evaluativity in legal texts, have brought to light interesting findings. However, they lack an important aspect crucial for determining whether legal texts are indeed more descriptive: a direct comparison between legal and public discourse.

Thus, the main aim of this paper is to compare legal and public discourse in terms the evaluative extent and intensity of its contents. is eOur*public* We will report (Section 3) an initial comparative analysis of the overall sentiment values of both corpora. These results indicate that courtroom language is indeed more descriptive than public discourse. However, this result can be easily explained by the fact that legal experts debate many technicalities of cases. Thus, we propose that our initial corpus study does *not* show that when legal experts use evaluative language, they do so in a more professional and sober mood.

In order to investigate whether legal language is indeed more descriptive than everyday conversations, we suggest to focus more closely on the very use of evaluative terms (that legal scholars undoubtedly use a lot). When it comes to evaluative terms, we need to distinguish thin from thick terms. and

While thick concepts have been at the heart of various debates in philosophy, they have hardly been discussed in philosophy of law. However, we believe that investigating thick terms and concepts in legal discourse has the potential to inform questions about the evaluativity of legal language and to extent the debate about thick concepts. above

Because thick terms have both a descriptive and evaluative aspect, they are the perfect items to examine the evaluative intensity and extent of legal contexts. Our main corpus analysis suggests that thick terms are used more descriptively by legal experts than when ordinary people interact. At least, when we look at how thick terms function in legal contexts, they are more often used in conjunction with terms that are less evaluative. The empirical results we present are exploratory and the methods we used need to be adapted for more fine-grained and theory-driven research questions. Nonetheless, we believe they provide an interesting starting point for such a project.

Section 2 provides an overview of previous empirical studies focusing on evaluative language and the different means that are used to communicate evaluations and stance taking in the courtroom. The empirical literature on this topic is rather scarce and only two studies mention thick concepts directly. In Section 3, reports a study on the overall sentiment values of both corpora but argues that the results of these studies can be taken as direct evidence for the claim that legal corpora are less evaluative. In Section 4 we develop the main hypotheses and methods for comparing evaluative intensity in legal and public discourse. In Section 4 we present the results of our corpus study involving a wide range of moral, epistemic and legal thick terms. Section 6 describes some limitations of course study and the impact on both legal and philosophical discussions.

# Evaluative Language and the Legal Discourse (Severin)

Legal professionals need to be objective in many respects. For instance, each defendant has a constitutional right to be given a fair trial, independent of any personal liking or disliking the legal professionals involved might have for the defendant. This involves an objective, unbiased treatment of the available evidence, and, especially on the side of the defence, a fair representation of the defendant. (In cases like Ted Bundy’s this might be tough.) In addition, legal professionals must only follow the law and cannot allow their own norms and ideals to affect their legal judgment. Whether or not legal professionals object to anyone’s personal lifestyle and decisions must not affect their judgment as long as these issues are not in violation of the law.

Because of this need for objectivity, one might suspect that the legal discourse is devoid of verbally expressed evaluations[[1]](#footnote-1). The legal system is there to reveal the truth (e.g., that Ted Bundy in fact killed these women) and, relatedly, legal processes seem to be characterized by a strictly regulated, objective, impersonal and unbiased adjudication in order to not distort the quested facts. Accordingly, so the assumption, this must also be reflected in the language.

However, it is part of especially the adversarial system that the quested facts are not found “by conducting disinterested research, but rather through the vigorous presentation of evidence slanted toward different positions” (Solan 2010, 395). Legal professionals, particularly in court, not only present impartial information in their texts but also evaluate claims of their opponents’ and contrast them with their own. Thus, it is not too far-fetched to conclude that evaluative language plays a pivotal role in judicial discourse.

Especially in the Anglo-American adversarial system, court cases are largely about two parties (prosecution and defence) presenting their conflicting versions of events and trying to convince the jury that *their* “constructed version of reality is more plausible than another” (Cotterill 2003). Consequently, the dialectical debate about how the evidence is to be assessed is at the centre of an adversarial trial. What is more, not only prosecution and defence have to take a stance in court, but also the judges themselves will eventually be required to state the reasons for the sentence she passes.

Stance-taking on certain events or agents is inherent both in the persuasive process and in decision making, which are integral parts of legal discourse. One might even go so far as to say that what legal professionals *mainly* do is evaluate, and, thus, communication in the legal domain is permeated by evaluative expressions. But how are these evaluations made? How do legal professionals express their stance, especially taking into consideration the “need to conform to the evidentiary rules which prevent explicit construal of judgement” (Heffer 2007, 145) and the demand on the judges’ attitude to be neutral?

There are several empirical studies investigating evaluative language in the courtroom by focusing on different linguistic devices by which legal professionals express their evaluative views. Davide Mazzi (2010) for example investigates evaluative lexis within a corpus of 67 US Supreme Court judgments and studies the judges’ various strategies to take a stance. On the one hand, he examines the most striking linguistic tools underlying judges’ evaluative statements, including straightforwardly evaluative verbal and adjectival items (like ‘disagree’ and ‘incorrect’), and he analyses the more finely-grained pattern ‘this/these/that/those + labelling noun’ on the other hand. When quantitatively studying this pattern, Mazzi finds that this pattern ‘is characterised by the occurrence of inherently evaluative elements as labelling nouns’ (e.g., ‘distortion’, ‘misapplication’, ‘omission’, ‘nonsense banner’), while the negative semantic prosody is predominant.[[2]](#footnote-2)

Much like Mazzi (2010), Goźdź-Roszkowski & Pontrandolfo (2012) explore the role of evaluative phraseology in the discourse of the highest courts in the US and the Italian criminal justice systems, focusing in particular on the pattern ‘noun + that’ (e.g. ‘fact that’). They look at the most frequent individual nouns in the ‘N + that’ pattern that belong to the semantically defined category of ‘argumentation’ and examined their functions in co-texts. That way, they find that certain nouns tend to entail negative polarity in their collocational environment (e.g. ‘fact that’), while others are used primarily with a positive polarity (e.g. ‘view that’). They conclude that the ‘N + that’ pattern is a widely used device to evaluate arguments in judicial reasoning.

Like Mazzi’s findings, this result crucially depends on the nouns themselves that occur in the pattern and the linguistically discrete items in the co-text, as the authors themselves point out: The pattern of ‘la conclusione che’ (the conclusion that), for example, “takes on a positive semantic polarity by virtue of its collocations with the adjectives ‘logica’ (logical), ‘corretta’ (correct), ‘giusta’ (right), ‘legittima’ (legitimate), etc.” (Goźdź-Roszkowski & Pontrandolfo 2012, 53). One could thus argue that the evaluative power that ostensibly lies in the ‘N + that’ pattern does not mainly come from this very structure but for instance from the adjectives (or verbs) involved.

In his corpus study, Edward Finegan (2010) examined adverbial expressions of attitudinal stance and emphasis as they were used in decisions rendered by the Supreme Court of the United States and the Supreme Court of California. He found that, among stance adverbials, the words “properly”, “improperly”, “appropriately”, and “correctly” occur particularly more frequently than in the more general *Brown Corpus of Written American English* and *British National Corpus*. This suggests that, “contrary to much public belief and considerable legal posturing, both state and federal supreme court opinions are far from lacking in expressions of attitude” (Finegan 2010, 73) but that legal professionals make use of subtly evaluating expressions remarkably often.

Stanisław Goźdź-Roszkowski (2018) shows what judges’ use of evaluative language could reveal, namely that legal argumentation relies upon (ethical) values. Looking at two landmark civil rights cases concerning same-sex marriage given by the USSC, he finds that ‘liberty’ and ‘dignity’ (among others) are keywords in the respective majority and dissenting opinions and concludes that ‘the judicial argumentation in these cases is to a large extent framed in terms of basic fairness and human dignity‘. Goźdź-Roszkowski sees these keywords as manifestations of a superordinate (ethical) value – i.e. respect –, towards which the judges’ argumentation is orientated, and highlights the ‘central importance’ of the related evaluative language for legal argumentation. Although the terms ‘liberty’, ‘dignity’, and (presumably) ‘fairness’ are usually considered thick concepts in the philosophical literature, no attention is drawn to this particular class of concepts, however.

Chris Heffer (2007) examines the evaluation of witnesses and defendants by trial lawyers and judges through lexis using corpus analysis. The corpus consists of several sub-corpora corresponding to various trial genres (e.g., counsel’s opening and closing statements, judge’s sentencing) and includes speech by 103 barristers, 72 judges from 150 ordinary trials of various Crown Courts in England. Following Martin and White’s (2005) appraisal framework, Heffer states that, when written into the text, judgement is conveyed by “the use of attitudinal lexis, particularly adjectival epithets (*normal*, *capable*, *reliable*), but also through attitudinal nouns (*liar*, *thief*, *saint*) and verbs (*lie*, *steal*)” (Heffer 2007, 154).[[3]](#footnote-3) After constructing a lexicon of such candidate judgement items (including not only words but also phrases), Heffer thus retrieved the items automatically from each of the sub-corpora in the form of concordances and then checked the concordance lines for judgmental relevance. He finds that “the figures on the whole are comparatively low, as would be expected of genres where explicit construal of judgement is proscribed” (Heffer 2007, 159). Interestingly enough, he also finds that the intensity of the inscription of judgement in sentencing seems to match the severity of the respective crime.

These studies strongly suggest that legal speech, especially in court, is indeed evaluative (although mostly subtly) and that there are numerous linguistic devices by which evaluation can be achieved. Two features of these studies stand out. First, the empirical evidence is generated by analysing linguistic corpora consisting of legal communication and by applying of corpus-linguistic tools to detect evaluations. Given the invaluable insights this method has already provided, we aim to extend this method even further. Second, these studies already recognise the evaluative function of many terms which are considered thick concepts by philosophers, such as ‘fairness’, ‘dignitiy’, ‘liar’, etc. However, the explicit connection to the philosophical debate is not made. Even though there are quantitative studies that focus on lexis, there is no systematic examination of the use of thick concepts. And this is despite the fact that they are perfectly suited for a subtle or blatant expression of evaluation, since they have a prominent descriptive dimension but also reveal the speaker’s attitude towards the person (or event) in question.

Only a few papers address thick legal concepts directly. Heidi Li Feldman (1997), for example, in discussing Bernard Williams (1995), simply presupposes the existence of thick legal concepts and stresses their significance for common-law reasoning. She claims that not only the philosophical debate about thick concepts can shed light on the mechanisms of legal language but, conversely, philosophy itself can “learn more about the nature and workings of thick concepts”, in analysing how “judges and lawyers apply, deploy, manipulate, exploit, and engineer” them.

Like Feldman, also David Enoch and Kevin Toh (2013) assume that “many of the crucial legal concepts that our legal judgments deploy are thick concepts”. However, they do not investigate which of them really are either but focus on the notion of legality itself. They conclude that ‘legal’ is indeed a thick term and use this finding to tackle four influential themes in twentieth-century analytical jurisprudence that were put forward by Hart, Dworkin, Raz, and Kelsen respectively.

As illuminating as these results are, the question still remains how prevalent thick concepts in the legal sphere are. Purely theoretical investigations into this matter cannot suffice, considering that what we are dealing with here is a linguistic phenomenon and an empirical question of how legal professional use such terms. Thus, if we agree that the distinction between descriptive, thin and thick concepts is necessary in order to understand how, and to what extent, evaluative language is deployed in a seemingly objective and purely fact-based discourse, we need reliable empirical data. The aim of this paper is to fill this lacuna.

# What Thick Concepts Are Likely to Be Used? (Severin & Pascale (erster Teil) UND Lucien und Kevin (zweiter Teil))

Philosophers usually distinguish between different kinds of thick concepts. The probably most intensively discussed kind are thick *ethical* concepts. Typical example of thick ethical concepts are virtue concepts, such as *rude*, *friendly*, *cruel*, *compassionate*, and, as Williams (1985, p. 144) claims “*treachery* and *promise* and *brutality* and *courage*”. Thick ethical concepts, so it is argued, combine descriptive and non-descriptive, often called “evaluative” or “normative” content. Depending on one’s metaethical views, this non-descriptive or evaluative content can be an expression of a person’s approval or disapproval (reference), or liking of disliking (reference), a command on how others ought to act (reference), an emotional stance (reference) or a statement about moral facts (reference).

Secondly, philosophers have discussed thick epistemic concepts, such as “*intellectual* *curiosity*, *gullibility*, and the concepts expressed by certain occurrences of ‘conscientious,’ ‘careless,’ ‘lucky,’ and ‘trustworthy.’” (Värynen 2008, p. 392). There is some disagreement about the non-descriptive content expressed by thick epistemic concepts. Some philosophers argue that some very general attitudes and evaluations are expressed, such as liking or disliking, and that this non-evaluative content is the same as in ethical thick concepts. Others have objected that the non-descriptive content is more domain-specific, with thick epistemic concepts expressing some unique, non-descriptive epistemic content. [[4]](#footnote-4)

It seems likely that the legal system makes use of both thick ethical and epistemic concepts. Thick ethical concepts might play a significant role in especially the criminal law. The criminal law has it its core the criminalisation of certain types of action as legally wrongful. A criminal act is something which ought not to be done (see Hart 1994, 27; Gardner 2007, 239; Tadros 2016, 91). As a consequence, the criminal law necessarily uses deontic language known from ethics and, arguably, thick ethical concepts as well. In addition, many legal scholars have stressed the intimate relationship between the criminal law and morality. The topics at issue on both fields of discourse are similar, and they concern various forms of harm (including physical, psychological, financial, and others forms of harm) and under what circumstances a harmful act is wrongful (for a discussion, see Edwards 2019)). Some people have argued that the source of the criminal system are the moral convictions of the people within the criminal-legal system (REFERENCES). Others have stressed that not only is the criminal system influenced by morality, the criminal system also has the ability to change our moral views, “such that neglected values come to be taken seriously by community members” (Green 2013a) and people become aware of the ethical reasons that make an act wrong. While certainly not all moral matters are also legally relevant, we should expect at least a significant overlap in the terms that are used.

Thick epistemic concepts are also likely to be used. One of the central features of the law is to determine whether a crime has been committed. This involves determining whether the defendant did whatever harmful or law-violating act he is accused of, but also whether doing so was wrongful.[[5]](#footnote-5) *\*Not Finished Yet\**

*\*\*Ich dachte, wir könnten hier nahtlos in eine Beschreibung übergehen, wie wir unsere Wortlisten erstellt haben. Haltet ihr das für sinnvoll?\*\**

# Corpus Analysis (Kevin und Lucien)

*Hier würde ich eine kleine Einführung in die Methode generell machen, ungefähr so, wie wir es in dem Methods Paper auch machen. Wir sollten kurz sagen, wie die Korpora zusammengestellt wurden, was Sentiment Values sind… Hintergrund-Informationen, die jemand, der die Methode nicht kennt, braucht.*

## Data Sources

To investigate thick concepts in the legal discourse, we had to create our own linguistic corpus for analysis. We decided to choose two main sources. First, we created a corpus with legal documents (henceforth: legal corpus or LC), based the Free Law Project 2020 which provides open data from court opinions of the American Courts of Appeals for the 1st to 11th regional circuit (without DC and the federal court). The courts of appeals are the intermediate appellate courts of the federal judiciary of the United States. They are divided into 13 circuits with each hearing appeals from the district courts within its borders and from other designated federal courts and administrative agencies. Its task is to determine whether the law was applied correctly in the actual trial court. The courts of appeals sit below the Supreme Court which is the last judicial instance to be consulted. Appeals courts consist of three judges and no jury. The courts of appeals are considered among the most powerful and influential courts in the United States, as they often set legal precedent which guide subsequent legal rulings. In addition, in over 98% of federal cases, these courts of appeals constitute the final legal instance. The court opinions – our text data – announce the decision after the case is tried. They usually include a summary of facts, the applicable law and how it relates to the facts, the rationale for the decision, and a judgment. Second, we created a corpus which includes non-legal language, based on comments on the world’s largest online forum Reddit (henceforth: Reddit Corpus or RC)[[6]](#footnote-6). The idea behind this corpus is to have a comparative baseline for the legal use of thick concepts in LC. With the two separate corpora, LC and RC, we are able to measure how evaluative the use of a certain adjective (e.g. *dishonest*) is in the legal context compared to a more colloquial use on social media. Reddit has dedicated legal subforums (so-called “subreddits”, like r/legal or r/legaladvice), where legal professionals are often among the users. It is very likely that these subreddits contain a more technical jargon, especially compared to the vocabulary used to react to cat pictures or memes. Hence, we excluded data from legal subreddits to avoid legal speech in RC as much as possible.

## Corpus Generation and Adjective Selection

Both the Pushhift Reddit Data Set [@Baumgartner2020] as well as the court opinions from the Free Law Project 2020 contain a wealth of information. To measure evaluative differences between concept classes, we first need to define the adjectives that are representative for these concept classes. The selection of suitable adjectives is based on the thick concepts literature,[[7]](#footnote-7) but also takes into account empirical considerations like frequency of occurrence or lexical diversity of the textual embedding. There is no point in selecting an adjective such as *courageous,* for instance, which is a typical example of a thick concept, but only occurs very infrequently our court opinions. In the following we detail the specifics of the corpus generation.

We start with a list of target adjectives which we specified based on the literature. This initial list contains adjectives in two categories, namely descriptive concepts and concepts which have a high potential to count as thick concepts. Among the thick concepts, we created sub-groups which differ in what non-descriptive information might be conveyed. First, there are *ethical thick concepts* whose non-descriptive content is ethical in nature. Examples were selected based on the vast literature on thick ethical concepts that are of special interest in the legal domain, as offences in the criminal law are not merely legal offences, but transgress moral norms too. Good candidates seemed to be terms related to physical harm, violations of someone else’s property or dignity, and terms with which we may describe the defendant’s character.[[8]](#footnote-9) Second, the legal system operates within a set of epistemic norms – norms of what we should believe and may conclude from a given set of premises. We therefore created a group of *thick epistemic concepts* which is inspired by the philosophical literature. Finally, it is plausible to believe that some thick concepts are used exclusively or predominantly in the legal context, such as the term “legal” itself. These concept we will refer to as *legal thick concepts*.

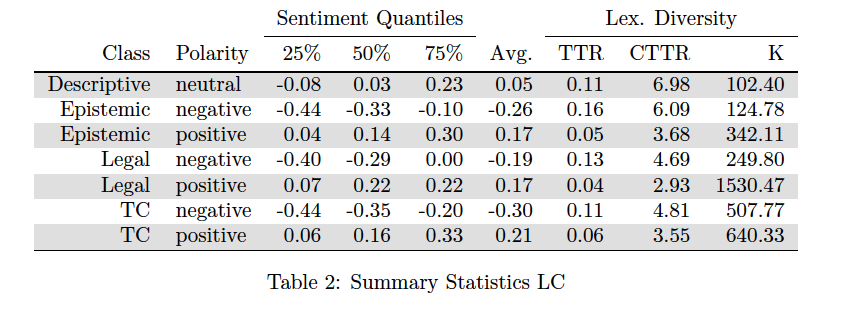
In a second step, we examined the validity of our selection of terms based on our legal corpus. But why only LC, and not RC? The reason is simple: RC contains billions of comments, whereas LC is much smaller and has most likely a more restricted and more codified vocabulary. Hence, for a term to be analyzable, we need a sufficiently high number of occurrences within LC; the number for RC is by experience always higher. In addition, if a term hardly ever occurs in LC, we need to question the assumption that the term is of legal interest at all. As it turns out, some of the adjectives on our initial list are rarely used in the legal context of the Court of Appeals, such as ‘brutal’ or ‘cruel’. We therefore had to adapt our list. Other terms may occur frequently, yet are most often part of legal phrases, which indicate a different semantic embedding, as is the case with ‘constitutional’ or ‘unconstitutional’. Infrequently occurring adjectives and the ones with a predominantly phrasal use were subsequently dropped. In order to avoid selection bias and compensate for the drops, we inductively selected a second battery or adjectives.

This inductive selection process is based on an analysis of part of speech tags (PoS-tags) in the legal corpus. PoS-tagging is an unsupervised method to annotate the syntactic structure of text data. For each of LC’s subcorpora (1st to 11th court circuits), we first draw a random sample of 2000 documents which are subsequently PoS-tagged using UDPipe [@Straka2017; @Straka2020]. Based on these PoS-tags, we isolate all adjective-pairs in ‘and’-conjunctions. Subsequently, we compute the co-occurrences of all adjectives and rank each adjective based on frequency and lexical diversity. As a measure for lexical diversity, we use Yule’s K [@Yule1944; @Tweedie1998]. We then select promising adjectives that also match our pre-defined concept classes. The full list of 24 target adjectives – including both the literature-based and the inductive selection – is shown in Table [tab:ADJlist] below. Lastly, we reduce LC to conjunctions that contain our target adjectives, and query the Reddit API to collect analogous data. All observations in the respective corpora are ‘and’-conjunctions of two adjectives, of which one is a pre-defined *target adjective*, whereas the other adjective is simply referred to as *conjoined adjective*. The corpus also includes information about additional adverbial modifiers or intensifiers for either of the adjectives. The full LC contains 49’199 entries, whereas RC has 69’211. Both corpora are cleaned, PoS-tagged, lemmatized and the conjoined adjectives are annotated with sentiment values from the SentiWords dictionary based on SENTIWORDNET [@Esuli2006; @Baccianella2010; @Guerini2013; @Gatti2016].

## Summary Statistics

In the following, we present the summary statistics for the key variables in each corpus: the sentiment values of conjoined adjectives (on a (-1,1] interval-scale), the sentiment polarity of the target adjective (pos/neg/neutral), and the concept classes of the target adjectives (Descriptive/Epistemic/Ethical/Legal). Table 1 shows the measures of sentiment dispersion and lexical diversity by class and polarity for LC. Table 2 contains the same for RC.

There are three main takeaways from Table 1: first, in regards to sentiment polarity, we have more extreme sentiment values for negative target adjectives overall. Their conjuncts have a higher diversity than ones of the positive target adjectives (the lower K, the more diverse). Secondly, irrespective of the concept class, positive target adjectives have a positive average conjoined sentiment, negative ones a distinctly negative average, and the descriptive ones have a more neutral average. So, at first glance, the average observed sentiment seems consistent with our assumption that ‘and’-conjunctions pair adjectives of the same polarity. Thirdly, the sentiment dispersion (quantiles) of the concept classes overlap only marginally, which already indicates that sentiment values are good discriminator.



Compared to Table 1 (LC), Table [2](#tab:RCstats) (RC) shows a far more polar sentiment dispersion, which indicates that overall laypeople use more evaluative language than legal professionals. Lexical diversity is also a lot higher in RC than LC – this was to be expected, due to the more codified nature of legal language use. The difference is most acute for legal and ethical thick concepts, but less so for thick epistemic concepts. Otherwise, RC exhibits the same patterns we noted above for LC.

Table

Description automatically generated

Figure 1 shows the sentiment dispersion on the level of the target adjectives shared by LC and RC (excl. descriptive concepts). We expect a generally negative distribution for negative target adjectives, and a generally positive distribution for positive target adjectives. As we can see, the polarity of the target adjective indeed looks like a good indicator for the polarity of the conjoined adjective, and vice versa. The sentiment spread (i.e. the whiskers) is mostly limited to either the positive or the negative region of the scale – respective to the polarity of the target adjective – and does not include the midpoint, with the exception of *lawful* and *unlawful*. In addition, the differences between the corpora we noted above are also present on the level of the target adjectives: LC has lower averages (i.e. dots) than RC across the board, with the exception of *dishonest* and *improper*.

# *Chart Description automatically generated*Study 1 (Kevin und Lucien)

## Methods

In Study 1, we assess the average context effects for both corpora. First, we are interested in whether there is a *difference in intensity* of evaluative language between legal professionals and laypeople. In this first step, we do not look at the polarity of the the evaluation (pos/neg), and thus use *absolute* sentiment values, rather than the ones on the initial scale (i.e. (-1,1]). To test H1, we use a linear model with the absolute sentiment values as DV and the corpus-dummy (LC/RC) as IV. Based on this model, we compute the estimated marginal means (EMMs) for the corpora. This gives us an overall estimate of differences in sentiment intensity between the corpora (LC/RC), irrespective of sentiment polarity (pos/neg).

The model for H2 assesses the difference in intensity, like the first model, but further discriminates by the *polarity* of the target adjectives.. Accordingly, the second model simply uses non-transformed sentiment values as DP, instead of the absolute values used in the previous model. Since we want to measure the effect of sentiment polarity in relation to legal and everyday language, we add the polarity-discriminator (pos/neg) as part of an interaction term (IV) with the corpus-dummy (LC/RC). This allows us to perform pairwise contrasts between the EMMs of the sentiment values for each corpus by target polarity.

Lastly, we inquire whether the *concept classes* behave significantly different in their legal use compared to their everyday use. Because we are mostly interested in differences in terms of intensity, the model for H3 once again uses absolute sentiment values as DV. As IV, we use an interaction term between the corpus dummy (LC/RC) and the concept class factor (Ethical/Epistemic/Legal). Subsequently, we use pairwise contrasts between LC and RC for the EMMs of each concept class.

## Results

Table 3 presents the EMMs based on the first model of study 1. The EMM for RC is 0.3622, the one for LC is 0.2360, on the absolute sentiment scale. The difference is significant on 0.05 alpha-level.. According to the linear model, LC has an average context-effect of β=-0.1262 compared to RC, t"-value:"-99.63, Pr(>|t|)=<2e^(-16), all other things equal. Hence, the sentiment values of the conjoined adjectives are indeed less intense for LC than for RC.



Graphical user interface, table

Description automatically generated

The second model shows similar results, which means that the difference between the corpora persists when we take polarity into account. The effect of positive polarity compared to negative polarity is , , whereas the effect of LC compared to RC drops slightly to , (note that the sign change is due to the different scale). The interaction of positive polarity and LC compared to the intercept has an effect of , . All effects are highly significant on a 0.05 alpha-level (). Table [4](#tab:s1m2) contains the EMMs by sentiment polarity for this model. The pairwise contrasts are all significant, which supports that LC has more neutral values than RC on both sides of the sentiment scale.

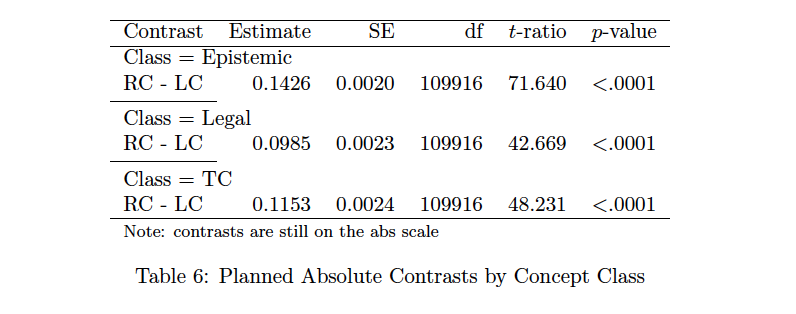


Table

Description automatically generated

Table [5](#tab:s1m3) shows the pairwise contrasts between the absolute estimates for each concept class and corpus, on the absolute scale. The differences indicate higher estimated values for RC compared to LC across the board, which is consistent with the findings of the previous models. All contrasts are significant on a 0.05 alpha-level. This shows that the differences between legal and everyday use of concepts are robust across concept classes. Interestingly, legal concepts show the smallest difference in EMMs of all concept classes. The effect sizes for all estimators can be found in Table [9](#tab:s1m3lm) in the Appendix.







## Discussion

Study 1 supports our intuition that everyday language is more evaluative than legal language. All three concept classes (ethical, epistemic and legal thick concepts) are used in conjunction with distinctly more extreme adjectives. These results should nonetheless be taken with a grain of salt. It might be objected that legal language is highly codified, and there probably are adjectives which signal the severity of an accusation, even though it is perfectly neutral out of context. In that respect legal jargon is similar to phrases used in reference letters or testimonials. The prior polarities from the sentiWords dictionary are not able to capture such context-dependent form of evaluativeness. While we limited the phrasal use of adjectives, less frequent phrases might nonetheless attenuate our results for LC. Having said that, the high number of observations and diversity certainly relativizes the weight of such phrases for the analysis.

# Study 2 (Kevin und Lucien)

## Methods

In the Study 2, we focus on LC only. Its aim is to measure whether our concept classes are distinguishable in legal language, i.e. they form distinct strata along the evaluative sentiment continuum. More specifically, we want to inquire whether they are relatively distinct from each other *and* are used more evaluatively than descriptive concepts. For our neutral baseline, we use the following descriptive adjectives: *active*, *ambiguous*, *complex*, *explicit*, *limited*, and *practical.*  Study 2 thus has four concept classes (Descriptive/Ethical/Epistemic/Legal), rather than the three in Study 1. Accordingly, the linear model for H4 simply includes absolute sentiment as DP and the concept classes as IV, followed by pairwise contrasts between the EMMs for the concept classes. While it would be interesting to perform planned contrasts by polarity, this is not possible, because the descriptive concepts only have a neutral polarity, which leads to empty interaction levels and contrasts. Hence, our model only allows us to assess differences in sentiment *intensity* between our concept classes.

While it is not possible to perform polarity contrasts as long as descriptive concepts are involved, we are still able to control that the remaining classes are distinct for negative and positive concepts respectively. For that, we simply focus on a subset of data without descriptive concepts. The linear model for H5 contains untransformed sentiment values as DP, and an interaction of polarity (pos/neg) and concept class (Ethical/Epistemic/Legal) as IV, and is followed up by pairwise contrast between concept classes of the same polarity. This allows us to measure whether *positive* ethical, epistemic and legal thick concepts have significantly different sentiment averages, and the *negative* ones do so as well among each other.

## Results

Table 7 presents the pairwise contrasts between the concept classes in LC (on absolute scale). As we can see, all concept classes have significantly different sentiment intensities (on a 0.05 alpha-level). The smallest differences are between epistemic and legal concepts (Δy =-0.0128), descriptive and epistemic concepts (Δy =-0.0140), as well as descriptive and legal concepts (Δy ‾=-0.0268). The contrasts involving ethical thick concepts, on the other hand, have a much wider spread, indicating that they are a much more distinct concept class. This is an interesting finding, because it raises the question whether sentiment polarity is actually a meaningful discriminator for concept classes or not, especially regarding future classification tasks.











Table

Description automatically generated

Table 8 shows the polar contrast for all concept classes, except descriptive concepts. Among negative concepts, all classes have significantly distinct sentiment averages (on 0.05-alpha level). For positive concepts, however, the difference between epistemic concepts (*consistent*, *logical*, *reasonable*) and legal concepts (*lawful*, *legal*, *legitimate*) is *not* significant.

Table

Description automatically generated

## Discussion

Study 2 shows that concept classes can be distinguished in legal language. This supports our hypothesis that legal professionals use certain concept classes in a more evaluative manner than others. On average, ethical concepts are much thicker (i.e. more evaluative) than epistemic concepts, and epistemic concepts are thicker than legal concepts. Positive epistemic and legal concepts are the only exceptions, since they are not distinct enough. In summary, legal professionals do not neutralize the evaluative component of concepts, they clearly use certain concepts evaluatively. Together, Study 1 and 2 establish a precedent for a sentiment-based analysis of concepts’ thickness. Finally, we want to stress that our results are an approximation of thickness in the form of aggregates (averages), and they do not account for additional factors on the level of the phrase or any other text unit which might have an effect on the thickness of an adjective.

# General Discussion (alle)

## The Limitation of our Empirical Approach

Our empirical studies show that sentiment dispersion within ‘and’-conjunctions is a reliable indicator to distinguish descriptive concepts as well as ethical, epistemic, and legal thick concepts. There are, however, serious limitations to our approach. First, our results only indicate *measurable differences* between concept classes. We do not claim that we can *classify* concepts solely based on their sentiment distribution in a corpus. In other words: we have shown pattern in overserved data, but not that these patterns can be used to classify new observations. Secondly, even if sentiment was a good basis for classification, our specific methods relies on a *known* sentiment distribution with a high enough *n*. In a classification setting, for every new concept, we would need to know its sentiment distribution in ‘and’-conjunctions in order to classify it. This information would have to be provided along with the classification task, which adds another layer of complexity and dependency on a potential sentiment-based classifier. Thirdly, and in contrary to our approach, conjunctions are not isolated features of sentences. [XXX have shown that contextual parameter can change how adjectives behave in conjunctions. Example] Thus, to develop more sensitive and adequate measurement (or classification), we have to take context into account. Finally, our idea, to use sentiment aggregates to measure thickness, assumes that a concept, e.g. *honest*, always belongs to the same concept class. This presupposes that *honest* always conveys an evaluation which is stable within a specific sentiment-bandwidth. Put differently, we essentially measured the thickness of the *lexical* representation of *honest*. This raises the question of how variable non-lexical thickness is, and how we would go about measuring it.

# Referenzen

Cotterill, J. (2003). *Language and Power in Court. A Linguistic Analysis of the O.J. Simpson Trial*. London: Palgrave Macmillan.

Enoch, D. and Toh, K. (2013). ‘*Legal* as a Thick Concept,’ in W. Waluchow & S. Sciaraffa (eds.), *Philosophical Foundations of The Nature of Law*. Oxford: Oxford University Press, 257–278.

Feldman, H. L. (1997). ‘Blending Fields: Tort Law, Philosophy, and Legal Theory,’ *South Carolina Law Review*, 49(1), 167–185.

Goźdź-Roszkowski, S. (2018). ‘Values and Valuations in Judicial Discourse. A Corpus-Assisted Study of (Dis)Respect in US Supreme Court Decisions on Same-Sex Marriage,’ *Studies in Logic, Grammar and Rhetoric*, 53:1 (66), 61–79.

Goźdź-Roszkowski, S., and Hunston, S. (2016). ‘Corpora and Beyond – Investigating Evaluation in Discourse: Introduction to the Special Issue on Corpus Approaches to Evaluation,’ *Corpora*, 11(2): 131–141.

# Goźdź-Roszkowski, S. and Pontrandolfo, G. (2012). ‘Evaluative Patterns in Judicial Discourse: A Corpus-based phraseological perspective on American and Italian criminal judgments,’ *International Journal of Law, Language & Discourse*, 3 (2), 9–69.

Heffer, C. (2007). ‘Judgment in Court: Evaluating Participants in Courtroom Discourse,’ in K. Kredens and S. Goźdź-Roszkowski (eds.), *Language and the Law: International Outlooks*. Frankfurt am Main: Peter Lang, 145–179.

Hunston, S. (2011). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Routledge.

Martin, J.R. and White, P. (2005). *The Language of Evaluation: Appraisal in English*. London: Palgrave Macmillan.

Mazzi, D. (2010). ‘‘This Argument Fails for Two Reasons …’: A Linguistic Analysis of Judicial Evaluation Strategies in US Supreme Court Judgments,’ *International Journal for the Semiotics of Law*, 23, 373–385.

Williams, B. (1995). ‘What Has Philosophy to Learn from Tort Law?’ in D. G. Owen (ed.), *Philosophical Foundations of Tort Law*. Oxford: Oxford University Press, 487–497.

1. The question of whether and how protagonists in the legal discourse use language not only to factually describe but also to assess people, events etc. is tackled by different approaches. Hence, a variety of expressions is used in the literature to describe this phenomenon, e.g., “evaluation”, “stance”, “stance-taking”, “attitude”, or “appraisal”. As Hunston (2011) puts it, evaluative language “expresses an attitude towards a person, situation or other entity and is both subjective and located within a societal value-system” (see also Goźdź-Roszkowski and Hunston 2016). We shall thus use “evaluative” and “normative” interchangeably here. [↑](#footnote-ref-1)
2. By stating that the encapsulating pattern ‘this/these/that/those + labelling noun’ is ‘a widely spread device by which judges express their evaluative views’ – which depends heavily on what philosophers might want to call ‘thick concepts’ –, Mazzi offers a good reason to further investigate thick concepts in order to fully grasp the way legal professionals take a stance in judicial discourse [↑](#footnote-ref-2)
3. Note again that many of these terms are considered thick concepts. [↑](#footnote-ref-3)
4. In addition to thick ethical and epistemic concepts, aesthetic concepts are discussed. For an overview, see XXX. In the rest of this paper, we will omit thick aesthetic concepts, as they play no immediately obvious role in the legal system [↑](#footnote-ref-4)
5. WE SHOULD HAVE A FOOTNOTE WITH AN EXAMPLE. Maybe something like: Killing another person is a harmful act that is impermissible under most circumstances. However, the law knows expectations in which the killing of another person can be excused. Exculpating circumstances include self-defence or duress. [↑](#footnote-ref-5)
6. For the Reddit corpus (RC), we gathered data using the API for the Pushshift Reddit Data Set provided by [@Baumgartner2020]. [↑](#footnote-ref-6)
7. For instance, to determine which thick concepts we examine, we selected those which are frequently discussed in the philosophical literature as prototypical, agreed-upon thick concepts. We further required all thick concepts to be actually used in ordinary conversations. For instance, the terms ‘lewd’ or ‘chaste’ belong to some of the paradigmatic examples of thick terms. However, we did not expect these terms to belong to the vocabulary of the average person. [↑](#footnote-ref-7)
8. Again, lewdness and chastity might play no or at least a negligible role in legal discourses. OTHER EXAMPLES?! [↑](#footnote-ref-9)