1 What Has to be Mentioned Before

2 Data

The data for this study comprises two distinct corpora: a corpus with legal documents and a baseline corpus with comments from Reddit, the world's largest online-forum. The legal corpus contains court opinions from Court of Appeals for 1st to 11th circuit, based on open data provided by the Free Law Project (2020). For the baseline corpus, we gathered data using the API for the Pushshift Reddit Data Set provided by Baumgartner et al. (2020).

In our case, the corpus generation is an iterative process. We start off with a list of target adjectives we specified without information about the corpora. This initial list, L_1, includes 2x5 adjectives that often discussed by the TC-literature and we deem rather uncontroversial. The list contains epistemic concepts (XXX), thick (XXX) concepts (XXX) as well as legal concepts (XXX). However, some of these adjectives are rarely used in the legal context, while others may occur frequently, yet are most often part of legal phrases, which indicate a different semantic embedding. In order to avoid any sort of selection bias and exclude adjectives with predominantly phrasal use, we inductively select a second battery of adjectives. This inductive approach is based on an analysis of part of speech (PoS)-sequences in the legal corpus. PoS-tagging is an unsupervised method to annotate the syntactic structure of text data. For each of the subcorpora (1st to 11th Court of Appeals), we first draw a random sample of 2000 documents which are subsequently PoS-tagged using UDPipe (Straka and Straková, 2017, 2020). Based on these PoS-tags, we isolate all syntactic structures of the form $(M) * A(,) * C(M) * A$ (M = modifier, A = adjective, C = conjunction, (...) = optional part). The conjuncts are then pooled across all subcorpora and only AND-conjunctions are retained. Finally, all adjectives are ranked according to frequency as well lexical diversity in regards to the conjoined adjectives. We use Yule's K (Yule, 1944; Tweedie and Baayen, 1998) as a measure for lexical diversity. Based on this ranking, we manually select adjectives that match our concept classes and add antonyms to form list L_2. In a third step, we combine L_1 and L_2, and use them to retrieve documents containing suitable target structures both from the legal corpus and via the Pushshift API. In a last step, we determine the three most promising concepts per class, based on number of occurences and Yule's K, and reduce the respective corpora accordingly. The final list of adjectives is shown in ?? in the 5.

The full legal corpus has XXX entries, the baseline corpus XXX. In order to keep the computational resources low while keeping a high enough n, we reduce the legal subcorpora each randomly by -40%, resulting in XXX legal documents overall. Both corpora are subsequently cleaned, PoS-tagged, lemmatized and the conjoined adjectives are annotated with sentiment values from the SentiWords dictionary based on SENTIWORDNET (Esuli and Sebastiani, 2006; Baccianella et al., 2010; Guerini et al., 2013; Gatti et al., 2016).

3 Method

In the following, we will discuss the methods used in our two studies.

3.1 Study 1

In the first study, we assess the average context effects for both corpora. First, we estimate the average sentiment intensity of conjoined adjectives in each corpus. The basis for this analysis is a linear model with the absolute conjoined sentiment values as dependent and the corpus-dummy as independent variable. Based on this model, we compute the estimated marginal means for the corpora. This gives on overall estimate of differences in sentiment intensity between the corpora, irrespective of polarity.

The second model further discriminates between positive and negative target adjectives, but does not yet include the concept classes. Instead of using absolute sentiment values as dependent variable, this model uses the non-transformed values. The polarity-discriminator is added as an interaction term between the corpus-dummy and the polarity-dummy for the target adjectives. This allows us to assess the estimated marginal means of the conjoined sentiment values for each corpus by target polarity.

The third and final model estimates the differences between corpora for each concept class. Because we are mostly interested in differences in terms of intensity, the model once again uses absolute conjoined sentiment values. As estimator we use an interaction term between the corpus dummy and the concept class factor.

3.2 Study 2

In the second study, we focus on the legal corpus only. In order to be able to compare the results of the evaluative concepts classes to a baseline, we added corpus entries for the following descriptive target adjectives: XXX. Instead of comparing context effects (Study 1), we want to inquire whether the concept classes cluster differently within the legal context. We use a combination of K-Means Clustering and Principal Component Analysis (PCA), which is informed by the sentiment values of the conjoined adjectives and a measure for lexical diversity of the target adjective, i.e. Yule's K (Yule, 1944; Tweedie and Baayen, 1998). The cluster analysis further includes information based on the vector space of the corpus, which is created using the *Semantic Vectors* package (Widdows and Ferraro, 2008; Widdows and Cohen, 2010, 2016). For every pair of target adjective and conjoined adjective, we compute the cosine similarities. Cosine similarities are essentially high dimensional representations of co-occurence measures, and inform the cluster analysis about the semantic similarity of the conjuncts by taking the whole corpus into account. In the final ANOVA model, the cosine similarities are added as weights for the sentiment values of the conjoined adjectives.

4 Results

Before we move to the results of our studies, there will be a brief discussion of the summary statistics of the text data.

4.1 Descriptive Statistics

The summary statistics are split by corpus. Table 1 shows the measures of sentiment dispersion and lexical diversity by class and polarity for the legal corpus (LC). Table 2 contains the same for the baseline corpus (BC).

The sentiment dispersion in Table 1 indicates more extreme values for negative target adjectives overall. According to Yule's *K*, their conjuncts are also more lexically diverse than the ones of the positive target adjectives. The average observed sentiment is consistent with our assumption that AND-conjunctions pair adjectives of the same polarity. The descriptive concept have a very neutral average (0.02), which is also consistent with our expectations. However, there seems to be an overlap between descriptive concepts several concept classes, which might prove to be significant later on.

| Class | Polarity | Sentiment Quantiles | | | | Lex. Diversity | | |
		25%	50%	75%	Avg.	TTR	CTTR	K
Descriptive	neutral	-0.18	0.00	0.22	0.02	0.15	6.17	168.36
Epistemic	negative	-0.44	-0.33	-0.10	-0.26	0.16	6.09	124.78
Epistemic	positive	0.04	0.14	0.30	0.17	0.05	3.68	342.11
Legal	negative	-0.40	-0.29	0.00	-0.19	0.13	4.69	249.80
Legal	positive	0.07	0.22	0.22	0.17	0.04	2.93	1530.47
TC	negative	-0.44	-0.35	-0.20	-0.30	0.11	4.81	507.77
TC	positive	0.06	0.16	0.33	0.21	0.06	3.55	640.33

Table 1: Summary Statistics LC

Table 2 indicates a far more polar sentiment dispersion in BC compared to LC, which is also reflected by the more polar averages. Lexical diversity is a lot higher in BC than LC, which was to be expected, due to the more codified nature of legal language use. The difference in homogeneity is most acute for legal concepts and TC, and less so for epistemic concepts.

| Class | Polarity | Sentiment Quantiles | | | | Lex. Diversity | | |
		25%	50%	75%	Avg.	TTR	CTTR	K
Epistemic	negative	-0.52	-0.42	-0.18	-0.34	0.10	7.65	73.19
Epistemic	positive	0.16	0.29	0.50	0.30	0.08	7.26	94.69
Legal	negative	-0.55	-0.44	-0.24	-0.38	0.13	6.65	198.62
Legal	positive	0.00	0.22	0.45	0.21	0.09	6.43	132.81
TC	negative	-0.55	-0.45	-0.29	-0.39	0.09	6.93	80.10
TC	positive	0.16	0.32	0.52	0.31	0.09	7.06	125.27

Table 2: Summary Statistics BC

The sentiment dispersion on the level of the target adjectives are shown in Figure 1. As we can see, the polarity of the target adjective is a good indicator for the polarity of the conjoined adjective, and vice versa. With the exception of *lawful* and *unlawful*, the sentiment spread is mostly limited to either the positive or the negative region of the scale and does not include the midpoint. The differences between the corpora also become apparent on the level of the target adjectives: LC has lower averages (i.e. dots) than BC across the board, with the exception of *dishonest* and *improper*. This becomes even more apparent on when aggregated by class (i.e. horizontal lines).

4.2 Study 1

The first model of study 1 is presented in Table 3 and shows the effect of context (LC v. BC) on the absolute sentiment of conjoined adjectives. According to the linear model, LC has an average effect of $\beta = -0.1262$ compared to BC, t-value: -99.63, $Pr(>|t|) =< 2e^{-16}$, all other things equal. The estimated marginal mean for BC is 0.3622, the one for LC is 0.2360. Hence, the sentiment values of the conjoined adjectives are indeed less intense for LC than for BC.

Corpus	Est. Mean	SE	df	lower CL	upper CL
BC	0.3622	0.0008	109920	0.3607	0.3637
LC	0.2360	0.0010	109920	0.2341	0.2380

Results are given on the absolute scale. Confidence level used: 0.95

Table 3: Absolute Estimated Mean Sentiment Difference Between Corpora

The second model shows similar results, even if we take polarity into account. The effect of positive polarity compared to negative polarity is $\beta_1 = 0.6456$, t-value:324.43, whereas the effect of LC compared to BC drops slightly to $\beta_2 = 0.1085$, t-value:35.31. The interaction of positive polarity and LC compared to the intercept has an effect of $\beta_3 = -0.2130$, t-value: -58.68. All effects are highly significant on a 0.05 alpha-level ($Pr(>|t|) =< 2e^{-16}$). Table 4 contains the estimated marginal means for this model. In summary, LC has significantly more neutral sentiment values than BC on both sides of the sentiment scale.

Corpus	Est. Mean	SE	df	lower CL	upper CL
Polarity = negative					
BC	-0.3659	0.0015	109918	-0.3689	-0.3629
LC	-0.2574	0.0027	109918	-0.2627	-0.2522
Polarity = positive					
BC	0.2797	0.0013	109918	0.2772	0.2822
LC	0.1752	0.0015	109918	0.1723	0.1780

Confidence level used: 0.95

Table 4: Estimated Mean Difference Between Corpora by Target Polarity

Table 5 shows the contrasts between the absolute estimates for each concept class and corpus. The differences indicate higher estimated values for BC compared to LC across the board, which is consistent with the findings of the previous models. Interestingly, legal concepts show the smallest difference of all concept classes. All contrasts are significant on a 0.05 alpha-level. The effect sizes for all estimators can be found in Table 7 in the Appendix. This shows that the differences between legal and everyday use of concepts are robust across concept classes.

4

Contrast	Estimate	SE	df	t-ratio	p-value
Class = Epistemic					
BC - LC	0.1426	0.0020	109916	71.640	<.0001
Class = Legal					
BC - LC	0.0985	0.0023	109916	42.669	<.0001
Class = TC					
BC - LC	0.1153	0.0024	109916	48.231	<.0001

Note: contrasts are still on the abs scale

Table 5: Planned Absolute Contrasts by Concept Class

4.3 Study 2

In study 2, we focus on LC and add descriptive concepts as a baseline. The aim of this study is to assess the possibility of distinguishing concept classes within the legal discourse, based on sentiment values. Table 6 presents the absolute contrasts between the concept classes in LC. As we can see, the newly introduced baseline does not have a significantly different sentiment intensity compared to legal concepts ($\text{diff}(\bar{x}) = -0.0027$, p-value: 0.8548), on a 0.05 alpha-level. All the other contrasts are significant. Hence, in regards to our cluster analysis, we can expect the different target adjectives from the same concept class to cluster together, with the exception of descriptive concepts, which might overlap with epistemic concepts.

Contrast	Estimate	SE	df	t-ratio	p-value
Desc. - Epistemic	0.0101	0.0034	44175	3.013	0.0138
Desc. - Legal	-0.0027	0.0034	44175	-0.799	0.8548
Desc. - TC	-0.0433	0.0035	44175	-12.355	<.0001
Epistemic - Legal	-0.0128	0.0020	44175	-6.253	<.0001
Epistemic - TC	-0.0534	0.0023	44175	-23.510	<.0001
Legal - TC	-0.0406	0.0023	44175	-17.752	<.0001

Note: contrasts are still on the absolute scale. Confidence level used: 0.95

P value adjustment: tukey method for comparing a family of 4 estimates

Table 6: Planned Absolute Contrasts by Class

References

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 2200–2204.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J., and Io, P. (2020). The Pushshift Reddit Dataset. Technical report.

Esuli, A. and Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 417–422.

Free Law Project (2020). Bulk Data CourtListener.com.

Gatti, L., Guerini, M., and Turchi, M. (2016). SentiWords: Deriving a High Precision and High Coverage Lexicon for Sentiment Analysis. *IEEE Transactions on Affective Computing*, 7(4):409–421.

Guerini, M., Gatti, L., and Turchi, M. (2013). Sentiment Analysis: How to Derive Prior Polarities from SentiWordNet. *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pages 1259–1269.

Straka, M. and Straková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*.

Straka, M. and Straková, J. (2020). UDPipe at EvaLatin 2020: Contextualized Embeddings and Treebank Embeddings. In *Proceedings of Language Resources and Evaluation*. arXiv.

Tweedie, F. J. and Baayen, R. H. (1998). How Variable May a Constant be? Measures of Lexical Richness in Perspective. Technical report.

Widdows, D. and Cohen, T. (2010). The semantic vectors package: New algorithms and public tools for distributional semantics. In *Proceedings - 2010 IEEE 4th International Conference on Semantic Computing, ICSC 2010*, pages 9–15.

Widdows, D. and Cohen, T. (2016). Graded semantic vectors: An approach to representing graded quantities in generalized quantum models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9535, pages 231–244. Springer Verlag.

Widdows, D. and Ferraro, K. (2008). Semantic Vectors: A Scalable Open Source Package and Online Technology Management Application. In *In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*.

Yule, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge University Press.

5 Appendix

	Dependent variable:
	abs(sentiment)
corpusLC	−0.143***
	(0.002)
classLegal	−0.031***
	(0.002)
classTC	0.026***
	(0.002)
corpusLC:classLegal	0.044***
	(0.003)
corpusLC:classTC	0.027***
	(0.003)
Constant	0.361***
	(0.001)
Observations	109,922
R^2	0.093
Adjusted R^2	0.093
Residual Std. Error	0.202 (df = 109916)
F Statistic	2,249.365*** (df = 5; 109916)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Linear Model for Planned Absolute Contrasts by Concept Class
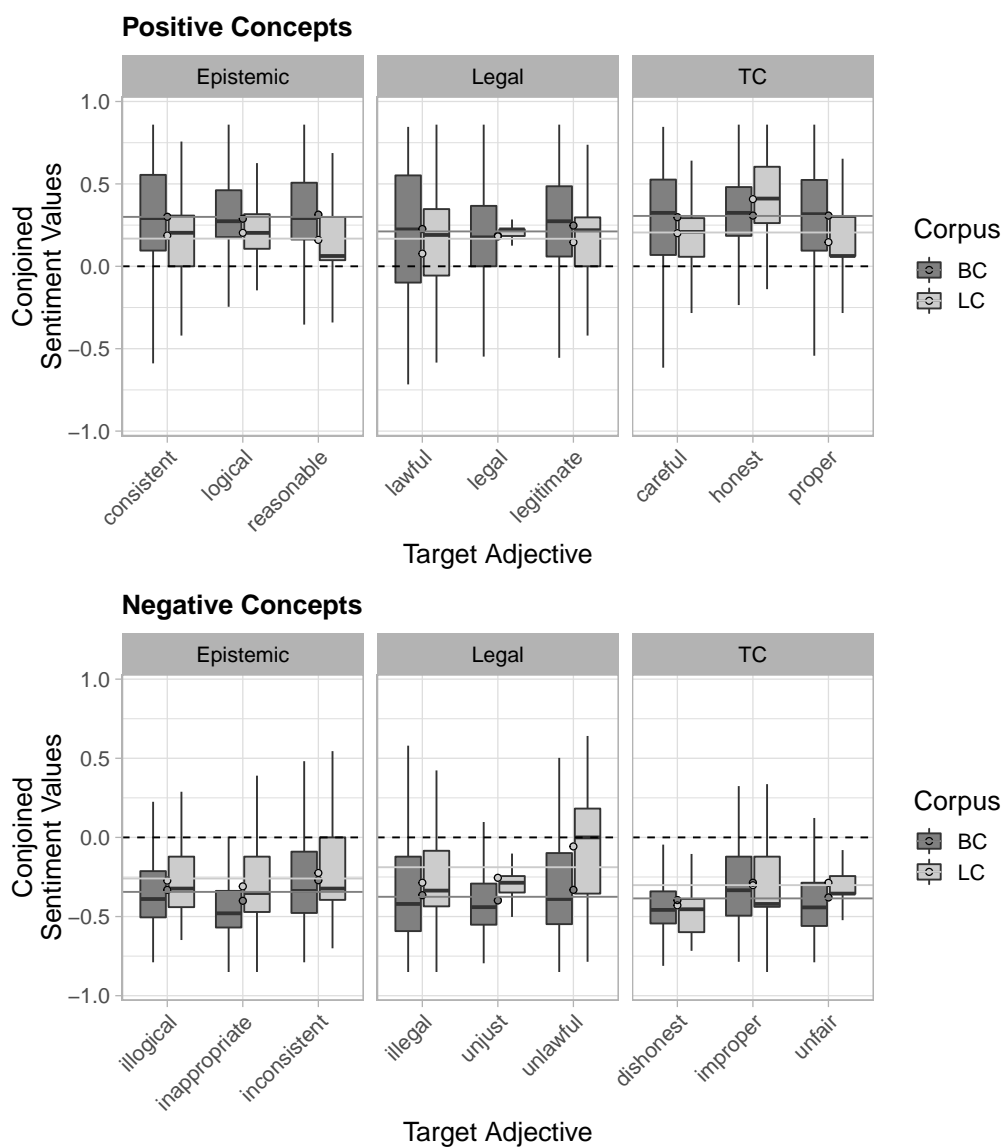
Positive Concepts

Negative Concepts

Figure 1: Sentiment Dispersion by Target Adjective