

A Comparative Corpus-linguistic Study of Thick Adjectives: Cases of the PRC Criminal Law and the Model Penal Code

Jinwen Yang^{1,a,*}

¹*School of International Studies, Zhejiang University, Yuhangtang Road, Hangzhou, China*

a. besty@zju.edu.cn

**corresponding author*

Abstract: With the promotion of law-based governance in China, laws and regulations play an ever more important role in our country whilst placing further demands on PRC law translation. Guided by previous research, this paper conducts a comparative study into the adjective phenomenon from the two selected corpus--the Model Penal Code and the PRC Criminal Law, respectively representing professional legal discourse and Chinese law translation. Concerning the research result, the author concludes the legal norms drawn from Model Penal Code and puts forward improvement proposals for the official Chinese-English law translations in terms of adjective distribution and category choices. The findings demonstrate that legal language tends to avoid adjectives and that thick concepts are more commonly used due to their evaluative intensity. By contributing to the promotion of future C-E legal translation, this paper aims to support China's legal system in its increasingly important role on the global stage.

Keywords: professional legal discourse, law translation, thick adjective

1. Introduction

The growing need for legal collaboration overseas has placed new demands on Chinese law, which now serves as an essential bridge for communication and exchange between common law and civil law systems. However, many of the official Chinese-English translation versions of key statutes remain crude and filled with serious traces of machine translation and inconsistent terminology. The PRC Criminal Law is one of the most representative samples of these shortcomings. To improve the quality of legal translation, it is essential to enhance modifier choices and adhere to legal language conventions in common law countries. This paper focuses on the misuse of adjectives in the PRC Criminal Law translation and conducts a comparative study of two corpora: the standard Model Penal Code, which employs normative legal language, and the flawed experimental PRC Criminal Law translation, which is the target of modification. The paper consists of three sections. First, it reviews current empirical legal discourse research and builds a part of speech algorithm model to compare the distribution distinctions of the two corpora. Second, it analyzes the statistical results concerning adjective frequencies and categories and concludes with a comparative conclusion and suggestions for optimizing the experimental corpus. This research aims to thoroughly study legal language and provide data-based references for future legal Chinese-English translations, as accurate and accessible

translations are essential for effective communication and understanding of the law, particularly in the context of China’ s increasing role in the global economy.

2. Previous Empirical Studies of Legal Translation and Thick Concepts

Early inquiries into improving the translation of Chinese law date back to the 1980s when Brenda Danet (1989) summarized the syntactic and discourse characteristics of legal English [1]. However, this early research did not sufficiently focus on morphology aspects, which remained unexplored until a dramatic advance brought by Professor Yue Zhongsheng (2020), who proposed that legal English contains a high degree of formality, obvious lexical features, and strong professionalism [2]. To further the legal discourse on the fuzzy phenomenon, Zhao Yongping (2020) presented translation tactics to avoid fuzziness from the perspective of legislative technical norms and normative measures [3]. The latest legal translation semantic research by Qu Wensheng (2022) illustrates that foreign translations of Chinese legislative texts should adhere to a system of rules of equal authenticity and accommodate common law legal language customs, which inspired this paper [4]. In addition, thick ethical concepts, which convey both attitudes and presentational content, are a significant phenomenon in law languages. Although thick concepts have been extensively discussed in the literature, much of the efforts have been paid to thick ethical concepts, such as quality adjectives like brutal, fiery, and diligent. In the past decade, a cohort of legal scholars has emphasized the importance of selecting thick adjectives in court languages. Heidi Li Feldman presented the dominance of qualitative adjectives and highlighted their significance in the legal field [5]. She also pointed out that forensic linguistics should acknowledge the nature and practices of quality concepts, especially when judges and lawyers apply, deploy, manipulate, exploit, and engineer thick concepts. Additionally, David Enoch and Kevin Toh strongly support Professor Feldman by claiming that several essential legal terms our legal judgments deploy are thick concepts [6]. Despite the numerous studies based on semantic and syntactic levels of legal translations and thick ethical concepts, few studies have been devoted to promoting thick adjectives in PRC law translation, and there was no comparative adjective analysis of authentic legal documents to reveal guiding principles [7]. Given the research deficiency, this paper constructs a corpus-based comparative model to bridge the gap.

3. Adjectives Distribution Control Discourse Analysis

Through the statistical scores and analysis from Sketch Engine, this section compares part of speech distribution in both two groups to screen the legal language usage patterns and the characteristics of adjectives in both codes. It establishes a measurement model to facilitate direct comparison, which marks a new step from former legal translation principles engaged only in idle theorizing.

3.1. Introduction to the Measurement Models

Three measurement models are applied to the algorithm, namely DOCF, ARF, and ALDF [8]. Altogether they will be taken into the Keyness Score account with a frequency degree.

Table 1: Keyness Score Model.

DOCF	ARF	ALDF	Keyness
Document	Average Reduced	Average Logarithmic	Score
Frequency	Frequency	Distance Frequency	

Document frequency (DOCF) refers to the number of documents in which a word or phrase appears. It can be a better comparison metric in cases where the corpus contains a small number of documents with a high frequency of specific words. On the other hand, average reduced frequency

(ARF) is a modified frequency that prevents results from being excessively influenced by a high concentration of a token in only one or more small parts of the corpus. When the token is evenly distributed in the corpus, ARF and absolute frequency have similar or identical values. Meanwhile, average logarithmic distance frequency (ALDF) is a corrected frequency that can be applied to word lists, n-grams, and keywords. The more ALDF is similar to absolute frequency, the more evenly distributed the token is. ALDF is based on distances between tokens and differs from ARF in its calculation. The Keyness score is obtained by comprehensively calculating the above formula and represents a thorough result after identifying keywords, terms, key n-grams, and keyword sketch collocations in Sketch Engine. The frequency measurement employed in the study is K-Score, and the higher the K-Score, the more dominant a word becomes in the text.

3.2. Part of Speech Distribution Results Analysis

The three statistical models generated a Keyness Score ranking for each non-phrasal term in the two corpora. The author filtered the scores with algorithm and data classification and presented the results in a table format with all adjectives italicized. An analysis of the statistical results will be provided, along with specific modification suggestions, including recommended adjective substitutions for the current translation version.

Table 2: Top 20 keyness scores in the PRC Criminal Law and the Model Penal Code.

RK	<i>CRIMINAL</i>	P	K-SCORE	<i>PENAL</i>	P	K-SCORE
1	<i>more</i>	<i>j</i>	4167.774	currentness	n	1304.29
2	<i>less</i>	<i>j</i>	1665.868	culpability	n	485.072
3	incarceration	n	1492.43	offense	n	289.093
4	<i>limited</i>	<i>j</i>	525.367	misdemeanor	n	270.905
5	<i>fixed</i>	<i>j</i>	500.177	felony	n	264.364
6	imprisonment	n	464.078	recklessness	n	264.225
7	sentence	v	385.235	crime	n	243.466
8	<i>serious</i>	<i>j</i>	307.922	imprisonment	n	239.825
9	circumstance	n	299.2	<i>justifiable</i>	<i>j</i>	235.216
10	crime	n	294.436	recklessly	a	235.136
11	commit	v	227.783	works	v	228.253
12	<i>preceding</i>	<i>j</i>	216.163	presentence	n	224.524
13	fine	n	207.613	intoxication	n	202.195
14	prison	n	200.876	laws	n	175.751
15	especially	a	198.334	<i>inchoate</i>	<i>j</i>	169.071
16	punish	v	189.856	purposely	a	157.582
17	directly	a	182.292	negligently	a	157.032
18	<i>responsible</i>	<i>j</i>	167.6	parole	v	155.978

Table 2: (continued).

19	probation	n	151.468	knowingly	a	146.103
20	other	j	148.639	duress	n	143.205

On the whole, the top 119 rankings of K-Scores in the total frequency of the four relative combined average frequency, up to around 20 adjectives emerged in both corpora, accounting for 17-18% of the total. The frequency seems approximate though, the K-Scores bear extreme variations. After complex calculations of DOCF, ARF, and ALDF, it turns out the Average K-Scores of the translation corpus is 2.15 times the professional one. The higher the score soars, the more dominant the word stands. Among the top 20 weighted K-Scores in the PRC Criminal Law, the frequency of adjectives jumps as high as 8/20, and 5 over 10 (50%) keywords with weighted total scores are adjectives; compared with the top 20 in the Model Penal Code, positing only 2 adjectives, among which the highest only ranked ninth. The statistical results demonstrate that among the approximately 120 keywords in each of the two codes screened in this paper, the Keynes Scores of the PRC Criminal law are significantly higher than the Model Penal Code on average, by a factor of nearly 2 or well over twice the number among top 500. The current translation of the PRC Criminal Law relies too heavily on adjectives, which leads to some obvious differences from the features of authentic legal texts. Since the use of adjectives is less formal than that of nouns and verbs, thin adjectives such as more or less will reduce the seriousness and accuracy of the legal text itself. The low frequency of adjectives in the legal texts of English-speaking countries also proves that they are more cautious in using adjectives and do not choose them to express their meaning easily.

Table 3: Summary statistics of the adjective distribution.

Algorithm	Top Proportion	Total Proportion	Top 20	> 5 Frequency	Average K- Score
<i>Penal Code</i>	20/119	16.80%	2/20	83	279.56
<i>Criminal Law</i>	22/119	18.48%	8/20	270	599.66

To name practical instances, when the term *recklessness* (n.) is set to account for “overconfident fault”, the *Model Penal Code* would not apply the adjective form of *reckless* (adj.) but rather the noun form; when expressing “able to perform”, the term *work* (n.) would be chosen rather than *workable* (adj.). More precisely, in § 3.09 of the *Model Penal Code*: “Mistake of Law as to Unlawfulness of Force or Legality of Arrest”, as is in the case of *recklessness*, the exact article is presented as “those Sections with unavailability in a prosecution for an offense for which recklessness or negligence, as the case may be, suffices to establish culpability”. In determining negligence, the terms chosen in the *Model Penal Code* are the nouns *unlawfulness* and *prosecution*, whilst the translation of the *Criminal Law*, quite the opposite, incorporated 5 adjectives in the translation version: “A **negligent** crime committed by two or **more** persons jointly is not to be punished as a **joint** crime; those who should bear **criminal** responsibility are **punishable** and are to be punished separately according to their crimes.” However, not a single adjective appears in

the formulation of negligent offenders in *the Model Penal Code*, and phrasal terms appear through “*suffices to establish culpability*” in *the Model Penal Code*. Hence, adjectives of PRC law translations are to be largely substituted by nouns and verbs, therefore the translation discourse will be more appropriate to legal conventions. The author revised the above translation and a modified version is presented hereinafter: ***Mistakes of Negligence committed by two or more persons are not to be jointly prosecuted; those who suffice to establish culpability are to be punished separately according to their crimes.***

The results are compatible with legal discourse conventions as judges and lawyers have no choice but to conform to the law in case any unrelated evidence influences legal justice. Legal professionals are obligated to maintain impartiality in their decision-making process, irrespective of their personal beliefs regarding lifestyle choices. Provided that such choices do not contravene the law, they must not be allowed to influence any judgments. Objectivity is a fundamental requirement in the field of law, as evidenced by the relatively lower distribution of adjectives in English legal discourse compared to nouns and verbs since adjectives are typically evaluative [9]. Moreover, legal language eschews verbally expressed evaluations. To ensure that legal processes maintain the integrity and do not distort the facts, they must be characterized by strict regulation, objectivity, impersonality, and impartiality. The language employed by legal professionals reflects this necessity for objectivity, with a greater emphasis on nouns and verbs and a correspondingly reduced use of adjectives. Not as an optimized language pattern in legal contexts though, adjectives somehow appear in law languages in the form of thick adjectives, which serves as the focus of the following paper [10][11].

4. Thick Adjectives Distribution Comparative Analysis

4.1. Thick Adjectives as a Linguistic Phenomenon

According to Green (2013), the two research corpora within the criminal justice system have the potential to effect changes in moral perspectives, as past ignored qualities become weighing more values by legal scholars [12]. While not all considerations are law-oriented, there is an obvious overlap in the terminology employed. This paper seeks to investigate legal language through the lens of “thick adjectives” as a linguistic phenomenon and examines how legal professionals and non-native translators apply these terms in practice. To this end, the paper begins by quantifying its frequency and concludes with a comprehensive result of the principles of professional law language through mass analysis [13]. The author specifically focuses on thick adjectives as a widespread, evaluative, and multifaceted linguistic phenomenon, to arrive at an integrated and adamant result regarding the distribution and usage of legal discourse. Thin terms typically convey single-dimensioned meanings of approval or disapproval and are therefore inadequate for measuring the evaluative nature of legal texts. In contrast, quality legal discourse incorporates both quality and presentational content and forensic linguistics frequently eschew plain terms like big and small, instead preferring more subtle and nuanced evaluative terms [14].

4.2. Corpus Sources and Thick Adjective Frequency Contrast

The author presents the findings of a corpus study that examined assessable law contexts. To obtain a comprehensive picture of evaluativeness, the author compared the translation corpus and professional corpus from the perspective of thick adjective frequency. Although a range of linguistic phenomena offers insights into evaluativeness, the use of thick adjectives is particularly relevant in the legal domain [15]. The analysis shows that legal professionals tend to use wide terms with less evaluativeness, which suggests that legal texts have less evaluativeness than plain English contexts. The results support the jurisprudence consensus that thick concepts can vary in their evaluative

intensity. The present analysis reveals that legal professionals exhibit a tendency to employ wide terms with a comparatively lower degree of evaluativeness, thereby indicating that legal texts display less evaluativeness than their plain English counterparts. The findings lend support to the prevailing jurisprudential consensus that thick concepts are capable of varying in intensity of implications.

Table 4: Summary statistics of thick adjective frequencies.

CRIMINAL	FREQ	SCORE	PENAL	FREQ	SCORE
more	656	10404.6	mental	54	717.3
less	540	8564.8	bodily	49	650.9
<i>serious</i>	518	8215.8	<i>extended</i>	47	624.3
other	489	7755.9	<i>necessary</i>	40	531.3
<i>limited</i>	380	6027.1	<i>deadly</i>	34	451.6
preceding	210	3330.7	<i>serious</i>	34	451.6
criminal	147	2331.5	<i>justifiable</i>	32	425.1
public	101	1601.9	former	26	345.4
large	98	1554.3	<i>guilty</i>	25	332.1
responsible	97	1538.5	<i>affirmative</i>	23	305.5
huge	59	935.8	<i>petty</i>	21	279.0
political	54	856.5	<i>indefinite</i>	21	279.0
<i>liable</i>	51	808.9	<i>specific</i>	20	265.7
illegal	48	761.3	<i>substantial</i>	17	225.8
heavy	48	761.3	<i>sufficient</i>	16	212.5
military	46	729.6	<i>authorized</i>	16	212.5
national	40	634.4	legislative	15	199.3
FALSE	37	586.8	<i>psychiatric</i>	14	186.0
few	36	571.0	<i>actual</i>	14	186.0

This paper conducts an experiment based on the control corpus, focusing on the main representations of adjectives in the legal languages of the translated countries and comparing them with the current translations to identify areas for improvement. The italicized adjectives are thick ones, the rest are thin. Yet according to the statistics above, the top 20 most frequent adjectives in *the Model Penal Code* share similar features, with 16 out of 20 adjectives being *thick* adjectives. The results reveal that authoritative legal language prefers thick adjectives, which mainly provide restrictions for generalized cases to meet the requirements of the principle of legality in criminal practice. For example, the 36th ranked *indefinite* is reflected in the Code as “An offender sentenced to an indefinite term of imprisonment in excess of one year,” where *indefinite* serves as the restrictive definite term of the legal bail period, and the use of a thick adjective reduces evaluativeness. The frequent use of the adjective *indefinite* reflects the precision of legal language and is intended to enhance the ambiguity of the law. Without thick adjectives, the text of the actual law would become rigid, which is not conducive to subsequent judicial interpretation and revision. *Extreme* and *reasonable* are examples of thick adjectives that allow the court to determine their scope according to the actual situation. If there are excusable reasons in the actual situation, the

extended interpretation of extreme and reasonable can facilitate the defendant's release, which is the essence of the legal language of English-speaking countries.

In contrast, the overuse of thin adjectives such as *more*, *less* and other single-dimensioned verbal adjectives is a prominent problem in the translation of *the PRC Criminal Law*. The absolute and relative frequencies of *more* and *less* in the translation of *the PRC Criminal Law* are significantly higher, with some 656 and 540 times respectively, and the expressions *more or less* are prominently applied when expressing contrasts. The author checked the BNS corpus and revealed that the sentences accompanied by *more* and *less* are colloquial daily conversations or within informal written texts, while rarely appearing in contracts, documents, or even more serious legal texts and law language. Thin adjectives, when applied in current translations, tend to appear in the form of "less than three years" but the legal discourse selects the formal equivalents of *excessive*, *exceeding*, and *not exceeding* as a replacement for the informal discourse. Based on the study of the legal language of *the Model Penal Code*, the author modified the above translation to be more appropriate to the legislative text of English-language countries: A crime syndicate **herein refers** to a **relatively fixed** crime organization **exceeding** three **offenders** for the **prosecution of joint crimes**.

5. Conclusion

Translating the PRC Criminal Law is a complex and challenging task due to the unique legal concepts, principles, and terminology presented in China's legal system. However, accurate translation is essential for effective communication and understanding, particularly given China's increasing role in the global economy. Therefore, it is important to make concerted efforts to ensure that translations of Chinese criminal law are accurate and accessible to people from different countries and legal systems. This paper presents a novel approach to investigating the use of adjectives in legal language through corpus linguistics. The author created an authentic legal corpus and a non-professional legal translation corpus to analyze the use of parts of speech and thick adjectives, serving as common linguistic tools in legal communication. The analysis proves that legal professionals prefer using nouns and verbs over adjectives, and thick adjectives are used more frequently with descriptive terms, indicating that legal language is less intensive than everyday conversations. By comparing the frequencies and viscous adjective distributions in both corpora, the paper identifies the shortcomings of current translations and provides practical references for improving the quality of legal translation in terms of part of speech and category choices. As China's law-based governance and globalization initiatives continue to expand, PRC laws are expected to receive increasing attention from other countries' judicial and political circles [16][17]. Therefore, legal translators should not rely solely on machine translation and haphazardly copy and paste legal documents. Instead, they should strive for utmost accuracy in translating legal materials, carefully choose terminology, and study customary wording and expressions in the target language to produce precise, concise, succinct, and formal legal translations. Legal translations can become more accurate and effective by reducing the number of adjectives, improving word choice rigor, and pursuing greater compatibility with authentic legal language in common law countries.

References

- [1] Lawrence, P. (1990). *Book Reviews: Brenda Danet: Pulling Strings: Biculturalism in Israeli Bureaucracy 1989*, Albany, N.Y.: State University of New York Press. 374 pages. *Organization Studies*, 11(3), 457–459. <https://doi.org/10.1177/017084069001100310>
- [2] Zhongsheng Y. (2020). *Properness of Translation Community Translating Normative Laws and Regulations of China*[J]. *Shanghai Journal of Translators*, 4, 30-34.
- [3] Falian, Z., Yongping, Z. (2020). *On the Singular and Plural Phenomena in Legal English and Their Translation Strategie*. *Shanghai Journal of Translators*, 5, 64-69.

- [4] Wensheng, Q. (2022). *The System of Principles of Foreign Translation of Chinese Legislative Texts: Based on the English Translation Practice of Chinese Civil Laws*. *Foreign Languages in China*, 19(1), 10-20.
- [5] Feldman, H. L. (1997). 'Blending Fields: Tort Law, Philosophy, and Legal Theory,' *South Carolina, Law Review*, 49(1), 167-185.
- [6] 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. doi: 10.1093/acprof:oso/9780199675517.001.0001
- [7] Marin, M. J., Rea, C. (2014). Researching legal terminology: a corpus-based proposal for the analysis of sub-technical legal terms. *Asp. La revue du GERAS*, 66, 61-82.
- [8] Webster, M. (2009). *Merriam-Webster Collegiate Dictionary*.
- [9] 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.
- [10] Chuang, L. (2021). Processing for Sentences without Subjects in C-E Translation of Communication Patent Documents. *Chinese Science & Technology Translators Journal*, 34(3), 5-7, 46.
- [11] Falian, Z. (2021). The Construction of a legal translation quality control System from the perspective of English Translation of the Civil Code. *Chinese Translators Journal*, 42(5), 121-130.
- [12] Green, L. (2013). Should Law Improve Morality?. *Criminal Law and Philosophy*, 7: 473-494. doi: <https://doi.org/10.1007/s11572-013-9248-3>.
- [13] Kexing, L. (2018). The Chinese-English Translation of legal provisions -- Take the translation of a judicial interpretation of "Two High" as an example. *Chinese Translators Journal*, 39(5), 94-101.
- [14] Ken, C., Xiang, L. (2021). On Expert Li Changshuan's Papers in the Field of Legal English Translation. *Chinese Science & Technology Translators Journal*, 34(2), 41-43, 37.
- [15] Finegan, E. (2010). Corpus linguistic approaches to "legal language": Adverbial expression of attitude and emphasis in Supreme Court opinions, in M. Coulthard and A. Johnson (eds.) *The Routledge Handbook of Forensic Linguistics*, 65-77.
- [16] Changshuan, L. (2016). English legal interpretation methods and their implications for translation. *East Journal of Translation*, 2, 76-80.
- [17] Wang, S., & Yin, J. (2020). A Corpus-Based Study of Keywords in Legislative Chinese and General Chinese. In *Workshop on Chinese Lexical Semantics*, Springer, Cham. 639-653.