Using Corpus Linguistics as a Tool in Legal Interpretation

Xiaoqiang Wu[1,a,*]

[1]*Suzhou University Wenzheng College, Jiangsu, Suzhou, 215100, China*
a. 2462494246@qq.com
*corresponding author

Abstract: Over the past decade, the application of corpus linguistics in legal interpretation has received attention from a group of scholars that engage in legal academics and linguistics. When the law takes into account the distribution of language usage, the application of corpus linguistics can help with addressing legal challenges. Before concluding that corpus analysis is optimally persuasive, the study addresses four difficulties that need to be addressed: The distribution of language use among a certain community must be the primary focus of the legal problem; Second, what constitutes a "ordinary" reading should be determined by the court; Thirdly, while searching a corpus to determine a term's usual meaning, one must predetermine what to look for; Fourth, there are a variety of reasons that a given meaning may turn up weak in a corpus search, and they should be recognized. Through analysis, this paper finds that though drawing on academic research and the experience of foreign judicial practice, Chinese scholars can enrich the research methods of the interdisciplinary field of law and linguistics, and provide inspiration for the nascent movement and the development of judicial practice.

Keywords: Corpus Linguistics, Legal Interpretation, Ordinary Meaning, Corpus

1.   Introduction

Over the past decade, the application of corpus linguistics in legal interpretation has received attention from a group of scholars that engage in legal academics and linguistics. Besides, the American legal community has treated it as an interdisciplinary field of law. And it has a promising prospect. In this paper, the author will argue that four issues should be addressed after introducing to corpus and corpus linguistics. And then, the definition of ordinary meaning will be given. First, corpus linguistics can clarify the actual use through analyzing the context of legal interpretation. This helps to reveal the practical application of legal language and helps to understand the meaning and application of legal contexts more accurately. Second, corpus linguistics can help study the variation and development of legal interpretation. Lastly, corpus linguistics can help analyze logical and reasoning relationships in legal texts. These four parts are especially important to language learners and the judicial domain.

2.   A Brief Introduction to Law and Corpus Linguistics and Ordinary Meaning

2.1.   The Definition of Law and Corpus Linguistics

Corpus linguistics has become a term referring to a wide range of activities and approaches [1]. Technological innovation has not only brought breakthroughs in data and methods in legal research,

but also brought opportunities and challenges. Law and Corpus Linguistics is a new interdisciplinary direction of law and linguistics. A corpus is simply a collection of texts stored according to specific criteria, which can be processed using specialist software [2]. Its main function is to rely on legal tests that are based on large-scale databases and a variety of corpus analysis tools to conduct corpus analysis. In other words, law and corpus linguistics use large-scale text collections as a tool and through concordance and collocation to conduct corpus analysis. Law and corpus linguistics is empirical and statistical. Thus, it can train AI models to improve language comprehension and also simply provide the necessary data base for AI. Though massive data computing training and optimization models are used to promote a more accurate and intelligent analysis of problems and decision-making in the smart justice system. When law and corpus linguistics becomes a methodology for practical legal interpretation, it applies to different contexts and adopts different theoretical preferences. At present, his research mainly focuses on two theoretical fields: dis ambiguity analysis and category determination in legal interpretation. What is clear is that methodological practice is the process of applying theory to practice.

First, in the field of legal interpretation, the most widely adopted research method is the corpus-based lexicography research paradigm. Its design concept is to understand the usage and meaning of words by integrating a large number of linguistic materials and analyzing and counting information such as the frequency, collocation, and meaning of words in different contexts. In this way, dictionaries no longer provide a fixed set of definitions, but meanings that can be flexibly adapted to the actual context. This research paradigm is widely used in the process of semantic disambiguation, which regards the lexicogrammatical pattern as a semantic unit rather than a single word, which goes beyond the limitations of traditional dictionaries.

Second, the corpus-based theory of cognitive categories can more effectively solve the problem of classification in legal interpretation, whether the things that are in dispute in the case are homogeneous to the categories specified in the relevant legal texts, and if a regulation is considered to distinguish between the tariffs of vegetables and fruits, should tomatoes be defined as fruits by botanists? From Crout's point of view, the ordinary meaning did not mean the most prevalent meaning of the terms [3].

Tomatoes, like cucumbers, squash, beans, and peas, are botanically considered fruits of the vine. However, all of these are vegetables grown in kitchen gardens, and like potatoes, carrots, parsnips, turnips, beets, cauliflower, cabbage, celery, and lettuce, they are known by the common language of the people, whether they are selling or buying provisions, and are typically served at dinner in, with, or after the soup, fish, or meats that make up the main course of the meal, rather than as dessert, like fruits are typically.

2.2. The Definition of Ordinary Meaning Principle

Law is not a collection of "commands", but a collection of texts. The application of corpus linguistics in legal interpretation is inseparable from ordinary meaning. The ordinary meaning rule states that when a statutory term is not defined by the law, courts should assign that term its ordinary meaning [4]. Like the meaning of words commonly used by ordinary people in their daily lives. Thus, the meaning of the words, phrases, and sentences used in legal texts is the meaning that ordinary people usually understand. But this leaves us with a dilemma that we sometimes see that courts do not limit interpretation severely. The interpretation of criminal laws is limited by the rule of lenity, which remains to be deposed. For example, "Smith United States", a famous Supreme Court case [5]. Smith was charged with attempting to trade his unloaded machine gun for illegal drugs. The judge analyzed the practical semantics of "weapons", "vehicles", and "carrying". The judge determined that the defendant in this case "carried a firearm" in the car during the drug transaction, thereby violating federal law by carrying a firearm while in the drug transaction. The defendant emphasized that the

word "carrying" was searched according to the COCA corpus, and the subject in daily use was generally "people" rather than "objects". Smith fits within the typical definition of the word "use a firearm," according to the majority of the six Justices, using definitions of use from several dictionaries. And the majority was correct in most cases. In his vivid dissent, Justice Scalia contended that individuals would associate gun ownership with the use of the weapon rather than just with the possession of a valuable object. Scalia contended, citing vivid instances from daily life, that: Generally speaking, using an instrument implies using it for the intended function[5]. "Do you use a cane?" is not a question meant to probe if you walk with a cane or if you have your grandfather's walking stick with a silver handle on display in the hallway. In a similar vein, to talk about "using a firearm" is to talk about utilizing it for its intended use—that is, as a weapon. Though "one can use" a rifle "in a number of ways," such as an object of commerce, just as one can "use" a cane as a hall ornament, that is not the typical sense of "using" either weapon.

The ordinary meaning principle should have no disagreement when a term under particular circumstances is given in a context. Therefore, Smith used the gun to trade for the drugs, there is nothing odd sounding about saying that he used it to acquire the cocaine. There are more limitations in the concept ordinary meanings to the dissent. It comes closer to proto typical meaning or prevalent meaning. When we heard that Smith was using the gun as part of the drug deal, we would not have thought that he had traded his gun for the drugs. So, Smith using a firearm is the ordinary meaning for most of us.

In the United States, the reliance on dictionaries has increased dramatically since 1987, with as many as one-third of judicial decisions citing dictionary definitions. But being unambiguous and non-controversial is still difficult to do. In China, the phrases and sentences used in the drafting of laws are also supposed to consistent with the rules of language commonly used. For example, the law provides that the activities of the court are "written into the record", which shall be understood as a record in the form of notes on paper. The literal meaning of "writing" should not be simply understood as holding a pen in hand, and writing down verbatim; "in the record" should be understood as a vehicle for later reference, and not as any ordinary notebook. Thus, the usual meaning is neither simply equivalent to a direct literal meaning nor is it a specific meaning that is divorced from its specific context. It must have a universally accepted meaning. In judicial practice, courts not only need to confirm the legal meaning of specific terms involved in legal disputes and texts, but also often have to define the specific meaning of common words such as "use", "age", and "vehicle". Therefore, the experience of judges is no longer needed to judge currently, and it is necessary to use new technical means and data evidence from the corpus to do so.

3. Four Conditions We Should Take into Consideration Before Using Corpus Linguistic As a tool

3.1. Ordinary Meaning is not a Universal Standard

When the court determines the meaning of one word that is not the typical meaning of the word. It has a narrow or broader meaning. When the narrow interpretation happens, the legislature has only thought of some reasonable meaning under some specific circumstance. The same as a broader interpretation. Not every judge has the same standard to determine every single word. The central issue is the great extent to which the courts investigate the circumstances of the law-making, draw inferences about the intent of the law, and decide whether the "ordinary meaning" of language in the law is too narrow or too broad to achieve these objectives.

3.2. The Notion of Ordinary Meaning Should be Adopted clearly

Similar to this, the court is meant to determine what conventional meaning implies in cases when it predominates. Ordinary meanings are considered "ordinary"[6] because of two key ideas. Firstly, the term's usual meaning describes the situations in which it should be used regularly. According to its second definition, an ordinary meaning of a phrase is a description of the situations in which people who belong to a relevant speech group would feel comfortable using the term to convey those situations. This hypothesis suggests that a phrase may have many ordinary meanings.

The same situation the author have talked about in 2.2. There is no uniform definition of vegetable and fruit. So ordinary means the understanding of ordinary folk. Justice Frankfurther stated: "If a statute is written for ordinary folk, it would be arbitrary not to assume that Congress intended its words to be read with the minds of ordinary men [7]. Judges reading them with the expertise of the specialists in mind are required if they are directed to specialists."

On the contrary, the second meaning tells us that every single word has more than one meaning under a specific circumstance.

3.3. Conducting the right search on the right corpus

In United States v. Costello, Judge Ponser criticized the drawbacks of relying on dictionaries [8]. He used Google News to determine that defendant in the case did not "harbor" her foreign boyfriend who was a drug trafficker. The federal district court judge found that the defendant had attempted to harbor the offender by knowing that the man was an illegal immigrant and had driven to the station to take him to his residence. She had attempted to harbor the criminal. Therefore, the court sentenced her to a suspended sentence of two years and a fine of $200. Subsequently, the defendant appealed, and the case was brought to the U.S. Court of Appeals VII. Posner concluded:

These data make it clear that the term "harboring," as it is actually used, has a connotation that includes "sheltering" and, more generally, "giving a person a place to stay." However, it does not imply purposefully defending members of a particular group against the authorities, either by physical protection, concealment, or moving to a safe location. This interpretation makes it possible to understand that providing emergency care to an alien even if he spends the night in the emergency room may not constitute "harboring"; similarly, offering an alien a lift to a gas station in order to help them get a flat tire may not constitute harboring; driving an alien to the local DHS office to apply for a status adjustment to that of a lawful resident may not constitute harboring; inviting an alien for a "one night stand" may not constitute an attempt at harboring; putting an illegal alien in a school may not constitute harboring; and, finally, letting your boyfriend live with you may not constitute harboring, even if you know he shouldn't be in the United States [9].

In the course of hearing this case, Posner did not rely on traditional dictionaries, but used the comments on the Internet as evidence to solve the problem of ordinary meaning in legal texts.

3.4. Being aware of drawing inferences from the absence of particular usages in a corpus

There is no doubt that a term is used by a speaker or author whose language has been documented in a corpus instantiation when the usage was used frequently. On the contrary, when a term is not used frequently or has never been used, no concept or interpretation will be found. Therefore, people are not supposed to rely too much on corpus today. With the development of AI, people can use the legal corpus as a tool to quantitatively analyze linguistic phenomena and draw conclusions or findings through the comparison of data descriptions and numerical values.

4. Conclusion

Understanding and interpreting the meaning of a text is an important basis for determining judicial interpretation. Chinese scholars can learn from the development experience of foreign legal corpus linguistics to improve their legal research and application capabilities. The legal corpus linguistics of the United States provides a rich and convenient resource of legal texts for legal scholars and practitioners, which is helpful for further analysis and understanding of legal texts. China can strengthen the analysis of legal texts by constructing its own legal corpus and applying relevant linguistic corpus, so as to promote adjudication, education research, and legislation. The application of corpus linguistics as a methodology in law has only happened in the past ten years, its core lies in treating corpus linguistics as an expansion method to innovate strategy and optimize the path of law. At the same time, corpus linguistics has a broad application prospect in the field of Chinese law. Through the use of corpus linguistics, legal texts can be processed and analyzed more accurately and objectively, which can provide strong support for legal theory research and practical work, and can also help lawyers and judges better understand and interpret legal language, and provide a scientific basis for judicial decision-making.

References

[1] Hunston, Susan. 2011. Corpus approaches to evaluation. Phraseology and evaluative language. New York/London: Routledge

[2] Zanettin, Federico. 2012. Translation-driven corpora. Corpus resources for descriptive and applied translation studies. Manchester: St. Jerome Publishing.

[3] Nix v. Hedden. 1893. 149 U.S. 304

[4] Slocum, Brian. 2015. Ordinary Meaning: A Theory of the most Fundamental Principle of Legal Interpretation. Chicago: University of Chicago Press.

[5] Smith v. United States. 508 US 2231993.

[6] Solan, Lawrence M., and Tammy Gales. 2017. Corpus linguistics as a tool in legal interpretation. BYU Law Review 2017: 1311–1357.

[7] Frankfurter, Felix. 1947. Some reflections on the reading of statute. Columbia Law Review 47: 527–546, United States v. Costello, 666 F.3d 1040 (7th Cir. 2012)

[8] United States v. Costello, 666 F.3d 1040 (7th Cir. 2012)

[9] United States v. Costello, 666 F.3d 1044–45 (7th Cir. 2012)