

When Laws Get Common: Comparing the Use of Legal Terms in Two Corpora*

Benjamin K. T'sou and Oi Yee Kwong

City University of Hong Kong

For more than a century, English was the only official language for Hong Kong's legal system. Only very recently has Chinese been granted equal status. While legal concepts can often be unambiguously expressed in English, their expression in Chinese has not yet reached the same level of precision. This discrepancy raises interesting questions in Chinese lexical semantics and new challenges in the practice of law in Hong Kong. In this study, we compare the use of a set of semantically related and easily confused Chinese legal terms in two corpora, one of legal domain and the other of general domain, to see how polysemous these words are and if the legal senses of these words in the former are preserved in the latter. Our analysis has shown that Chinese legal words are quite polysemous compared to their English counterparts, and are used with considerable fuzziness in general texts. We also discuss the sense distinction of these legal words with respect to their English translation differences as well as dictionary definitions. In the future we will explore the automatic construction of some WordNet-like lexical resource for legal terminology and extend our analysis to cover legal-word uses in Chinese communities outside Hong Kong.

Key words: Chinese lexical semantics, legal terminology, corpus analysis, WordNet-like lexical resource, sense distinction

1. Introduction

For more than 150 years, the legal system in Hong Kong operated solely in English. This situation changed only very recently, following the implementation of legal

* An earlier version of this paper was presented at the Third Workshop on Chinese Lexical Semantics, Taipei, May 2002. The authors thank the Hong Kong Judiciary for providing the judgment data, and take sole responsibility for the views and findings expressed herein. Support for this work has come partly from the Hong Kong Judiciary under grants RCL0794 and RCL1508, and from the Research Grants Council of Hong Kong as a Competitive Earmarked Research Grant No. CityU1206/97H (1997-1999) and No. CityU1238/00H (2000-2003). Special thanks are extended to Tom Lai for useful suggestions and comments on earlier drafts of this paper. We also thank the anonymous reviewers for useful pointers for improving the paper.

bilingualism in the '90s. Hong Kong is the first community to follow the Common Law system and allow the use of both English and Chinese in court proceedings. This has profound implications on language use, especially in the synchronization of the balanced and proper use of both languages to the same effect in the legal domain. Since English has a long and established tradition in Common Law, many legal concepts have been quite precisely lexicalized. This is far from being true for their Chinese equivalents, which is evidenced from the lack of one-to-one correspondence of legal terms between English and Chinese.

The peculiar patterns of cross-lingual lexicalization of legal concepts between English and Chinese make it a very important topic in lexical semantics. The polysemy therein must be well understood and accurately resolved for the legal language to take on its expected precision, so as to guarantee the proper enforcement of justice. Given the long tradition of having been used in legal proceedings under the Common Law system, English has evolved to quite a mature stage, such that it is precise to the extent that the legal senses of terms are relatively unambiguous, at least to most legal practitioners. In English, for instance, a "contract" (合約) and an "agreement" (協議) are sufficiently distinguished, despite their semantic relatedness (as both refer to some obligatory relations between two parties). However, when legal terms are expressed in, or more often translated into, Chinese, this preciseness is somehow weakened, which is problematic in the legal scenario because court proceedings should not be unnecessarily complicated as a consequence of imprecise language use.

Meanwhile, the problem gets worse when legal terms are used in more general, and less law-abiding texts, e.g., in news reports. According to Ahmad (2001), for example, the terminology difference is apparent between a domain-specific corpus and a general one. We observed that even the same legal terms might be used in different senses in different corpora. Thus closely related legal concepts might already be easily confusable in the legal domain, and may be further confused in general usage. In the case of English, code switching or overt code mixing by means of Latin, French, or Latinate words provides a means to mark the social setting and thus to define the register (e.g., *alibi*, *modus operandi*). This means is usually not available in the case of Chinese, which must rely on its own classical and literary tradition, and which involves no comparable code switching similar to that of English. In Hong Kong, the use of English terms within parentheses in official or legal documentation provides a special, perhaps transitional, context reflecting the traditional dominance of English in the official domain. This comparison invites the suggestion that the situation with English could be less problematical than that with Chinese.

The situation regarding the use of Chinese in the legal domain hence calls for at least two things to be done in Chinese lexical semantics: (1) to distinguish the senses of

Chinese legal terms, especially for semantically related concepts, perhaps via some WordNet-like constructions (Miller et al. 1990), and (2) to study the usage of semantically related legal terms in both legal texts and more general ones and see how the precise legal senses in the former are preserved in the latter. The current study makes a preliminary attempt in both regards.

In Section 2, we will first briefly describe the polysemy problem of legal terms between English and Chinese; and then in Section 3, we will outline the approach taken in this study for classifying the senses of Chinese legal terms. Section 4 reports the results of our corpus analysis, with respect to the senses found for a set of closely related Chinese legal terms in a legal-domain corpus, and how they compare to those found for the same terms used in a general-domain corpus. The results are further discussed in Section 5, with the sense distinction illustrated via English translation differences. Future work is suggested alongside a conclusion in Section 6.

2. Polysemy of legal terms

The multiple-rendition phenomenon mentioned at the beginning of this paper suggests that upon translation from English to Chinese, on the one hand, many different English legal terms are found to be expressed by the same Chinese legal term (e.g., “裁決” has been identified as the translation equivalent for “decision”, “verdict”, and “award” in a bilingual corpus of court judgments). In these cases, some of the subtleties, including the positive/negative connotations as well as the extensions of the various English words, may not be distinctly reflected and accurately captured in the Chinese terms. On the other hand, the same English legal term may give many different Chinese counterparts (e.g., “decision” has been rendered as “裁定”, “決定”, “裁決”, and “判決”).

If we take English as the reference point and assume the various English terms all have a distinct legal sense (disregarding their non-domain-specific senses, e.g., “award” can refer to “a prize in a competition”), the multiple Chinese renditions for each English term would suggest that as far as the legal senses are concerned, the Chinese words are much more polysemous than their English counterparts. It is one of our objectives in this paper to lay out the semantic relation between a set of semantically related and easily confused legal terms; and understanding the polysemy patterns would be a prerequisite for doing so.

The morphemic structure of the above Chinese words may also help identify the specific meanings the words convey. For example, “裁定” (“裁”+“定”) may as a result be different from “裁決” (“裁”+“決”) in the sense that the former has more to do with a “conclusion” (定論) whereas the latter is more about a “decision” (決定).

Nevertheless, these words may be used on different occasions in the Chinese context by convention, so it would be more reliable to look directly at the contexts of use, which might sufficiently dictate how the words are used and what senses they carry. We have therefore taken a corpus-based approach in this work to study the polysemy of Chinese legal terms.

3. Sense distinction for Chinese legal terms

Word sense distinction constitutes a notorious problem in lexicography, lexical semantics, as well as word sense disambiguation. “How to define a certain sense” and “how fine-grained the senses should be” have been some of the debating issues. In this section, we explain the approaches we have adopted for this study regarding these issues.

On the definition side, we follow the philosophy of the Collins COBUILD English Dictionary (Sinclair 1987) which defines words by illustrating their uses via corpus linguistics, using authentic examples from naturally occurring texts to “explain” word uses, and such “explanations” often include pragmatic meanings like the positive/negative connotation in the use of a word. We based the definitions solely on what the corpus suggests because existing lexical resources and reference materials are not sufficiently detailed for the purpose. There are many comprehensive English dictionaries of law (e.g., Garner 1999), but standard references for legal Chinese in Hong Kong have not matured to the same level. Most of them exist in the form of a glossary, with only very crude definitions, if any (e.g., Department of Justice 1998; Department of Justice 1999; Li & Poon 2000). So we have to rely on authentic usage to help distinguish the subtleties among the word senses. Nevertheless, we refer to existing resources considerably when making our own judgement.

For sense granularity, as pointed out by Wilks (1998), the results of sense discrimination tests often only show whether the judges are sense “lumpers” or “splitters”. Hence the final sense set might not be very reliable if we only take human intuition into account. Rather, in addition to human intuition and perception of the linguistic contexts, we also consider more objective criteria. On the one hand, we judged by verb argument structures and subcategorization patterns where appropriate. On the other hand, Resnik and Yarowsky (1999) suggested, for the evaluation of word sense disambiguation systems, that senses be distinguished and defined on the basis of multi-lingual data, that is, a word is considered to have two different senses if the same word is translated into two different words of different meanings in a second language (or still others). Hence we also look for translation difference in our bilingual legal-domain corpus when judging the sense distinction of the Chinese legal terms.

4. Corpora analysis

4.1 The analysis task

In this study, we start by selecting a set of semantically closely related and often confusable legal terms to work with. The set consists of “裁定” (hold, convicted), “裁決” (determine, verdict), “判決” (judgment, conviction), “裁斷” (find, finding), and “裁判” (Magistracy)¹. Sentential contexts for these words were extracted from the following corpora:

(1) *Corpus of bilingual Hong Kong court judgments*

The sample corpus contains English and Chinese court judgments from legal proceedings in Hong Kong, with a total of about 1M Chinese characters for the Chinese half.

(2) *LIVAC corpus (texts collected from Hong Kong in 1997-98)*

This is the synchronous corpus² developed by the Language Information Sciences Research Centre of the City University of Hong Kong. The corpus contains mostly newspaper articles collected synchronously and regularly from six Chinese speech communities. For the current study, we only make use of about 1.8M Chinese characters from the Hong Kong section for a one-year duration out of a total of 100M characters.

It may be noted that in this study the corpus size is much bigger for the general corpus than the domain-specific one. Given that we are looking at legal terms, we need a bigger general corpus to extract a comparable amount of examples from the two corpora. We randomly selected 30 samples³ for each word in our set from each corpus, analyzed the contexts embedding the words, and tried to define the senses assumed by the words in these examples. This is essentially a sense-tagging task, except that we derive and define the senses simultaneously as we tag them for each word occurrence in the corpora.

As far as sense distinction is concerned, apart from the meanings suggested by contexts, the following additional criteria were taken into consideration during the analysis:

¹ The English terms are the more common translations of the corresponding Chinese terms as observed from the bilingual corpus of Hong Kong court judgments. They are included here for reference purposes only, not necessarily as the absolute or correct translation suggestions *per se*.

² <http://www.rcl.cityu.edu.hk/livac> (鄒嘉彥 1998)

³ Except for “裁判”, which did not have enough samples.

(1) Part of speech

The selected words were first classified by the part of speech they took on in the sample sentences. Verb uses were distinguished from noun uses.

(2) Translation equivalence

With the court judgment corpus, we also compared the English version of the sample sentences as translation difference could be an indicator of sense difference. For example, the first two verb senses for “裁定” shown below (Section 4.2) correspond to “convict” and “hold” respectively, which suggests a difference in their meanings.

(3) Argument structure and subcategorization pattern

Sense 2 and sense 4 for “裁定” in Section 4.2 might be conceptually similar, but in fact they differ not only in the meaning (sense 2 refers to the action of providing the conclusion, while sense 4 refers to the action of judging on some issue), but also possibly in their subcategorization.

(4) Focus

The core aspect of the meaning is another indicator for us to split senses as this could have different implications in court cases and could thus lead to different interpretations and outcomes. For instance, although the two noun senses in Section 4.2 for “裁定” apparently have similar English counterparts, they refer to different kinds of conclusions. One is targeted at the issues of dispute in a case, while the other pinpoints the final outcome of the case itself. Obviously the former sets the premises leading to the conclusion covered by the latter, and they should be treated as different senses, although the difference might only be subtle.

The analysis for each word is reported below, and their comparison is discussed in Section 5. In the following tables, the “POS” column refers to the part of speech for the word in the particular contexts of the corresponding senses. “Legal (%)” shows the relative distribution of the senses in the judgment corpus, whereas “Gen (%)” shows the distribution in the LIVAC corpus. The definitions in Chinese and English are phrased by the authors, and are based on the authors’ understanding of the linguistic contexts. The (L) or (G) at the end of each example indicates the source of the example, i.e., from the legal-domain corpus or the general corpus, respectively. The English translations, where available, are also included for reference.

4.2 裁定

Sense	POS	Definition and Examples	Legal (%)	Gen (%)
1	v	<p>法庭對訴訟的結果作出定論 (the court decides on the outcome of a case)</p> <p>於 1997 年 12 月 19 日，陪審團裁定申請人三項控罪均罪名成立。(L) On 19 December 1997 the jury convicted the applicant on all three counts.</p> <p>在八三年七月，法庭裁定<u>荷爾</u>誤殺罪名成立。(G)</p>	43.33	43.33
2	v	<p>法庭對案中的爭論點作出定論 (the court resolves an issue in a case)</p> <p>原審法官裁定所提出的要求沒有得到滿意答覆。(L) She [the judge] <u>held</u> that they had not been satisfactorily answered.</p> <p>裁判官裁定僱員的行為不足以構成行為不當。(G)</p>	36.67	20.00
3	v	<p>法庭作出指令 (the court gives an order)</p> <p>本院又裁定上訴人可獲得他在向本院和上訴法庭提出上訴時所招致的訟費，除非有人在本判案書發下的十四天內提出書面陳述。(L) The Court <u>awards</u> costs in the appellant's favor, incurred in this Court and in the Court of Appeal, unless written representations are made within 14 days of the handing down of this judgment.</p> <p>法官裁定港府要即時釋放他們。(G)</p>	3.33	26.67
4	v	<p>法庭就某爭論點判斷以解決爭端 (to judge on some issue to resolve dispute)</p> <p>陳文敏解釋，《基本法》已訂明特區永久性居民的子女享有居留權，但臨立會卻通過法例，規定偷渡來港的小人蛇必須遣返，實是有違《基本法》，因此法庭屆時將需裁定臨立會的合法性。(G)</p>	0.00	6.67
5	n	<p>對案中爭論點所下的定論 (the resolution of an issue of dispute)</p> <p>原審法官和上訴法院先後裁定賣方就第一項要求的回覆，不能令人滿意。本席認為上述裁定正確無誤。(L) I therefore consider that the judge and the Court of Appeal were right in holding that the first requisition had not been satisfactorily answered.</p>	10.00	0.00
6	n	<p>對案件本身所下的定論 (the decision on the outcome of a case)</p> <p>故此我認為暫委法官的裁定是正確的，除了因時間過去而需要作出修改外，我會復還她的命令。(L) I therefore think that the Deputy Judge was right and would restore her order, subject to certain amendments which have been made necessary by the passage of time.</p>	6.67	3.33

4.3 裁決

Sense	POS	Definition and Examples	Legal (%)	Gen (%)
1	v	<p>法庭考慮證供及法律原則後作出決定 (the court makes a decision based on evidence and law)</p> <p>代表原告人的大律師所提爭論點並不新奇：相信大家會清楚知道，事實上類似的論據在以前的訴訟案中亦有提出過，包括在最近裁決的兩宗案件中，但並不成功。(L)</p> <p>The points raised by counsel for the Plaintiff are not novel: as will become apparent, similar arguments have in fact been deployed, unsuccessfully, in earlier cases including two recent <u>decisions</u>.</p> <p>若有法律觀點分歧，最終交由法庭裁決。(G)</p>	16.67	6.67
2	v	<p>法庭決定訴訟結果，包括刑罰和指令 (the court decides on the outcome, sentence, and orders of a case)</p> <p>最高法院法官艾布拉姆斯將要裁決：把伍德沃德的謀殺罪減為誤殺罪和只判處她二百七十九天監禁；抑或判佐貝爾越權。(G)</p>	0.00	3.33
3	n	<p>法庭對訴訟結果的決定（如勝訴/敗訴、罪成/無罪、死因等） (the court's decision on the outcome of a case)</p> <p>陪審團有充份的證據讓他們達致誤殺的裁決。(L)</p> <p>There was ample evidence for the jury to arrive at the <u>verdict</u> of manslaughter.</p> <p>醫院管理局將成立專責小組調查今次事件，而家屬則表示要等待死因庭的裁決後，才決定是否向法院方追討賠償。(G)</p>	30.00	6.67
4	n	<p>法庭的決定，多指牽涉仲裁庭對金錢賠償的判決 (the court's (arbitration tribunal's) decision on monetary compensation)</p> <p>在以公共政策為理由而拒絕執行公約裁決的案件和文本中，都有提述“國際公共政策”。(L)</p> <p>In regard to the refusal of enforcement of Convention <u>awards</u> on public policy grounds, there are references in the cases and texts to what has been called “international public policy”.</p>	23.33	0.00
5	n	<p>法庭對訴訟結果的決定，以及相應的判決和指令 (the court's decision on a case, and relevant orders)</p> <p>樞密院在 Mayson v. Clouet ([1924]AC980) 一案中對 Howe v. Smith ([1884]27Ch.D.89) 一案的裁決表示贊同。(L)</p> <p>In Mayson v. Clouet [1924] AC 980 the Privy Council approved the decision in Howe v. Smith (1884) 27 Ch.D.89 where ...</p> <p>港府昨日發表聲明，表示會就高等法院上月底裁定港府要釋放十名越南人的裁決上訴。(G)</p>	16.67	53.33

6	n	<p>對爭論點所下的結論 (the resolution of an issue)</p> <p>不過，不論舉行聆訊與否，法定程序體制也預期關於證據接納性的<u>裁決</u>，應在聆訊控罪當日之前作出，以便審訊得以有效率地進行。(L)</p> <p>But whether there be a hearing or not, the scheme contemplates that the determination of admissibility would be made some time before the date fixed for hearing the charge. This facilitates the efficient conduct of the trial.</p> <p>政府在考慮過程當中，都考慮得很周到，所以不需要到人大才可得到一個<u>裁決</u>。(G)</p>	13.33	20.00
7	n	<p>宗教、派系、團體的指令 (religious orders, etc.)</p> <p>情報官員指出，雖然未曉得發出追殺令的回教教士的確實身分，然而回教激進分子向來都非常認真聽命於這類宗教<u>裁決</u>。(G)</p>	0.00	10.00

4.4 判決

Sense	POS	Definition and Examples	Legal (%)	Gen (%)
1	v	<p>法庭決定訴訟的結果/成敗 (the court decides on the outcome of a case)</p> <p>本席<u>判決</u>上訴得直。(L) This appeal is allowed.</p> <p>倫敦南部薩瑟克刑事法院於周五<u>判決</u>一宗偷竊案件。</p>	13.33	20.00
2	v	<p>法庭判斷和解決爭論點 (the court resolves an issue in a case)</p> <p>上訴法院是否<u>判決</u>在 1993 年 1 月 28 日的一封信件之前，受託人本人（與破產人分開）有權根據《公司條例》第 179 條提出呈請。(L) Whether the Court of Appeal <u>ruled</u> that prior to the letter of 28 January 1993, the trustee himself (separately from the bankrupt) had the right to petition under s179 of the Companies Ordinance.</p> <p>一位菲律賓法官上星期<u>判決</u>，這些中國漁民並未非法進入菲律賓水域。(G)</p>	3.33	6.67
3	v	<p>法庭作出指令或判刑 (the court gives an order or a sentence)</p> <p>國際法庭未被授權<u>判決</u>罪犯死刑。(G)</p>	0.00	6.67
4	n	<p>法庭對案中一切事項的決定，及有關指令 (the decisions made by the court, and related orders)</p> <p>但這些理由至低限度可以合理地將當地法院<u>判決</u>或公約裁決作廢。(L) But the reasons must go beyond the minimum which would justify setting aside a domestic <u>judgment</u> or award.</p> <p>本港勞工法例中，唯一保障在職婦女不會因為懷孕而遭解僱的條例，可能會因高等法院昨日一宗上訴案的<u>判決</u>，變成形同虛設。(G)</p>	56.67	66.67
5	n	<p>定罪的結果 (conviction, the judgment of being guilty)</p> <p>以上是二項罪名定罪<u>判決</u>的立場。因此，第一項罪名定罪<u>判決</u>的“撞擊”影響問題亦消失了。(L) The foregoing being the position in regard to the <u>conviction</u> on the 2nd count, the question of a “knock on” effect on the conviction on the 1st count falls away.</p>	26.67	0.00

4.5 裁斷

Sense	POS	Definition and Examples	Legal (%)	Gen (%)
1	v	<p>法庭推斷或認定事實真相 (the court infers and confirms a fact)</p> <p>1998 年 5 月 17 日，阮雲道法官<u>裁斷</u>上訴人及第二被告人兩人皆有疏忽。(L)</p> <p>On 17th May 1998, Nguyen, J. <u>found</u> negligence on the part of both the appellant and the 2nd defendant.</p>	46.67	0.00
2	v	<p>法庭決定訴訟的結果 (the court decides on the outcome of a case)</p> <p>1995 年 11 月 20 日，警務處處長通知上訴人，他對控罪 B 被<u>裁斷</u>有罪所提出的上訴被駁回，並且按照《警察（紀律）規例》第 27(1) 條，處長會向總督呈遞一份該案的報告，因為當處長覺得應該判處革職或迫令退休的懲罰時，他須按這條規例的規定去做。(L)</p> <p>On 20 November 1995, the Commissioner of Police informed the appellant that his appeal against the <u>finding</u> of guilt in relation to Charge B had been dismissed and that a report of the case would be forwarded to the Governor in accordance with Regulation 27(1) of the Police (Discipline) Regulations which provide for this course where it appears to the Commissioner that dismissal or compulsory retirement is merited.</p>	3.33	0.00
3	n	<p>法庭對事實問題的推斷 (the court's finding of a fact)</p> <p>從這些條文中完全不會得出見習騎師是馬會之僱員的結論，而上訴法庭也沒有作出這樣的<u>裁斷</u>。(L)</p> <p>None of these provisions remotely point to a conclusion that the apprentices are employees of the Club and the Court of Appeal did not so <u>find</u>.</p>	46.67	0.00
4	n	<p>法庭對訴訟結果的決定 (the court's decision on a case)</p> <p>1995 年 7 月 17 日，警司 De Oliveira 宣布他的<u>裁斷</u>：裁定上訴人控罪 A 罪名不成立，但裁定他控罪 B 罪名成立。⁴ (L)</p> <p>On 17 July 1995, Superintendent De Oliveira announced his <u>finding</u> acquitting the appellant of Charge A but convicting him of Charge B.</p>	3.33	0.00

⁴ In this particular context, the superintendent (警司) was the chair of a special tribunal for disciplinary proceedings within the police force.

4.6 裁判

Sense	POS	Definition and Examples	Legal (%)	Gen (%)
1	v	指法院級別 (a court level) 申請人於東區 <u>裁判</u> 法院接受審訊，在 1996 年 7 月 9 日被裁定有罪。(L) The applicant was tried at the Eastern <u>Magistracy</u> and was convicted on 9 July 1996.	100.00	37.50
2	v	法庭對犯人判處刑罰 (the court sentences the convict) e.g., 司徒敬在 <u>裁判</u> 被告未來一年做二百四十小時義工時，亦要求他接受心理治療，在感化官指定的地方居住，並充公他所有子彈、槍械零件、氣槍，只歸還電腦。(G)	0.00	12.50
3	n	法庭對案件結果的決定 (the court's decision on a case) e.g., 上訴庭三名法官昨日一致推翻這 <u>裁判</u> 。(G)	0.00	37.50
4	n	賽事（如球賽）的公證人 (the umpire in a match) e.g., 此次世界杯賽共有來自十三國家的一百零五名運動員參加，其中日本代表隊有八名女運動員和七名包括教練、領隊、 <u>裁判</u> 在內的隨團人員。(G)	0.00	12.50

5. Discussion

5.1 Senses and morphemic patterns

In Section 2, we suggested that the morphemic patterns of the legal terms may be indicative of their subtle sense differences. This could have been a result of the differences in word formation for English and Chinese. While many legal concepts are uniquely lexicalized in English, in Chinese they might be expressed by near-synonyms which share a similar or identical base morpheme (e.g., “裁”) to distinguish the subtle differences among those concepts. This is evident from the corpora analysis as shown by the most frequent sense of each word. For instance, “裁定”, with its morpheme “定”, is mostly used when “a conclusion is made for something”. On the other hand, “裁決”, as “決” would suggest, has more to do with “decisions”. Meanwhile, “判決” emphasizes “判”, which involves “judgment leading to conviction and orders”. Finally, for “裁斷”, the morpheme “斷” refers to “logical inference on factual matters”, and hence the primary sense of the word. Table 1 shows further illustrative examples.

Table 1: Morphemic Patterns as Sense Indicators

裁	定	Conclusion	<ul style="list-style-type: none"> • 第二被告人獲<u>裁定</u>無罪，而申請人則被<u>裁定</u>交替性的誤殺罪行有罪，被判處 4 年 8 個月監禁。 (The second defendant was acquitted while the applicant was found guilty of the alternative offence of manslaughter. He was sentenced to 4 years and 8 months imprisonment.) • 基於蕭錫森所言屬實，賴馨德法官乃<u>裁定</u>申請人的三項罪名，全部成立。 (On the basis that Peter SIU was telling the truth, he [the judge] convicted the applicant of all three charges.)
裁	決	Decision	<ul style="list-style-type: none"> • 陪審團有充分的證據讓他們達致誤殺的<u>裁決</u>。 (There was ample evidence for the jury to arrive at the verdict of manslaughter.) • 關於(iii)這一點，第 79(2)條規則規定，凡任何人向樞密院申請上訴許可，但有關申請未有<u>裁決</u>，則該申請人可向終審法院申請發出指示。(As to (iii), rule 79(2) provides that where a person has applied for leave to the Privy Council but the application has not been determined, the applicant may apply to the Court for directions.)
判	決	Judgment	<ul style="list-style-type: none"> • 上訴法庭於 1996 年 12 月聆訊該宗上訴案，並於 1997 年 4 月 29 日作出<u>判決</u>，駁回上訴，並確認這四項控罪的定罪。 (The appeal was heard by the Court of Appeal in December 1996 and judgment was given on 29 April 1997 dismissing the appeal and confirming the convictions on the four charges.) • 買方不服上訴法庭的<u>判決</u>向本院提出上訴，請求恢復范達理法官的<u>判決</u>。(The purchaser being dissatisfied appealed to this Court, seeking a restoration of Findlay J's judgment.)
裁	斷	Inference	<ul style="list-style-type: none"> • 原審法官<u>裁斷</u>蕭錫森此言難以使人入信，且損害其可信性，但他又認為這並不是說蕭錫森的供詞全屬虛言。 (The Judge found Peter SIU's evidence that ... , hard to believe and that this obviously damaged Peter SIU's credibility. But he found that it did not follow that other parts of Peter SIU's evidence were untrue.) • 上訴法庭<u>裁斷</u>祇是違反《仲裁規則》第 32 條和《仲裁法》第 45 條，除此之外沒有違反其他條文。(The Court of Appeal found that there was no breach of these provisions, except for Article 32 of the Rules and Article 45 of the Law.)

5.2 Degree of polysemy

From the analysis, we see that although each word in the set has one or two major senses, there is still considerable polysemy for each word. On the one hand, all of them demonstrate categorial ambiguity, such that they can be used as a verb or a noun, and, what is not shown in the tables above, where there is corresponding English translation, the parts of speech of the translation equivalents are not always identical. On the other hand, the intra-POS polysemy is noteworthy. Some of the sense distinction is legitimate, while others may be a consequence of misuse, as suggested by their low individual relative frequency. It is apparent that “裁定” is more often used as a verb, whereas “裁決” and “判決” more often as a noun; and more interestingly, “裁斷” has about equal chance of being a noun or a verb.

5.3 Semantic relatedness

We chose the set of legal words in this study because they are semantically close and related, and they can be easily confused, especially when used by laymen. From the analysis, we can in fact probe how they relate to one another and to what extent they are confused in general usage.

In terms of semantic relatedness, we can identify sets of near synonyms among the observed senses. Manually we have arranged the different senses into some WordNet-like structure as shown in Figures 1 and 2 respectively. Words enclosed in curly barckets { } are near-synonyms, and the number after each word refers to the sense of the word as we described in Section 4.2 to 4.6. Where there is more than one word in a set, we boldfaced the one with highest relative frequency with respect to the legal-domain corpus. The arrows mark the hierarchical structure, pointing to a hyponym in the noun hierarchy and a troponym in the verb hierarchy.

A fragment of the noun hierarchy from WordNet 1.7.1 corresponding to similar legal concepts is shown in Figure 3. The nodes in the hierarchy are sets of near-synonyms (synsets) in WordNet. Compared with Figure 2, we see that the hierarchy of Chinese legal words derived from our corpus analysis does not differ much from the way their English counterparts are organized and related in WordNet.

Figure 1: A Verb Hierarchy

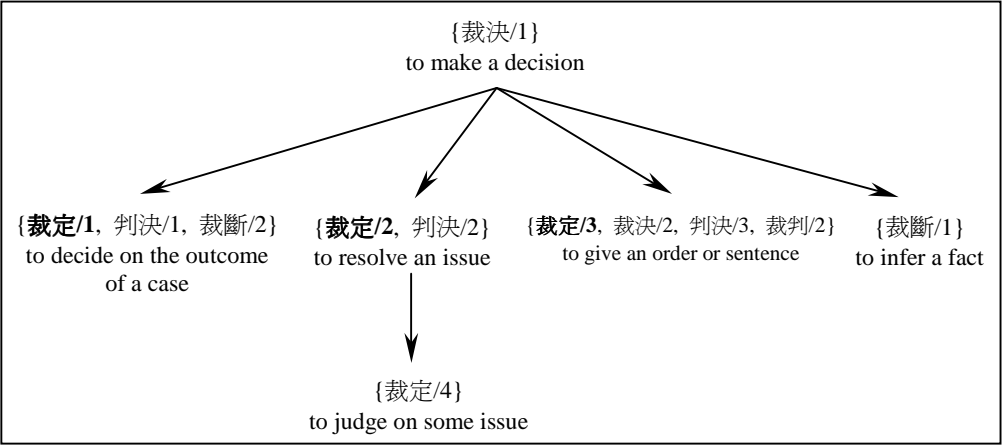


Figure 2: A Noun Hierarchy

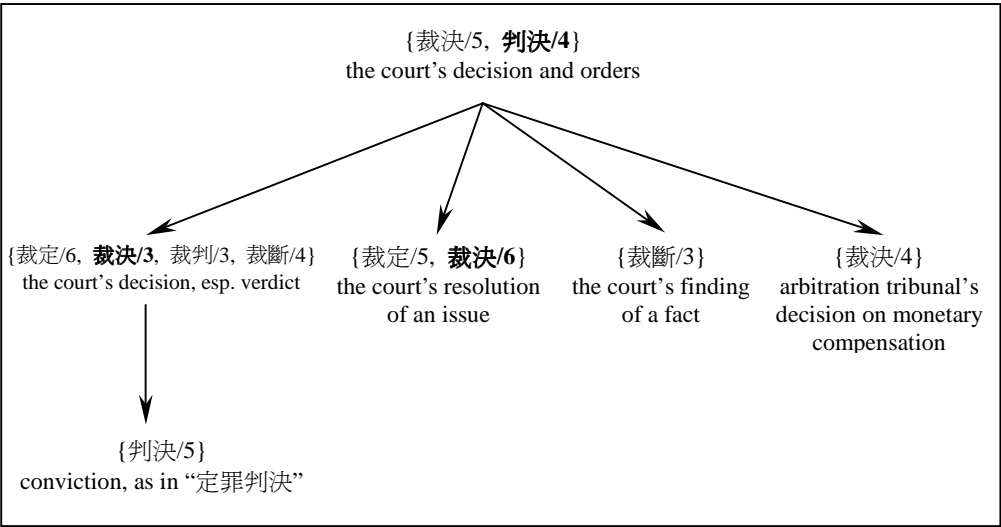
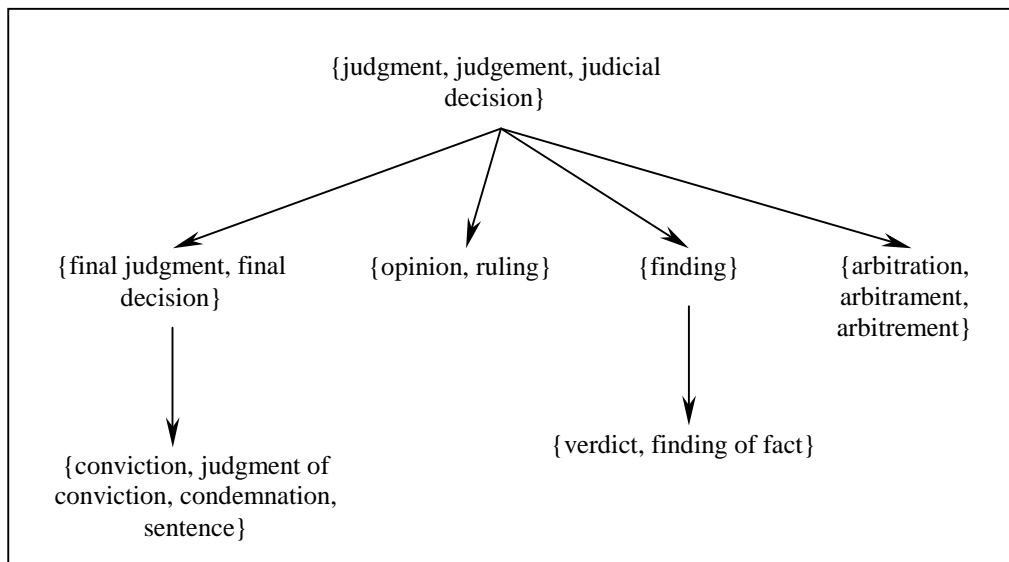


Figure 3: A Fragment of the WordNet (1.7.1) Noun Hierarchy



The relative frequency of usage of individual words in a particular sense may suggest the usage conventions and even the possibility of misuse. For instance, the following points of interest are observed:

- “裁判” is only used to refer to a court level in judgments, as in “裁判法院” (Magistracy), but more variably used in news reports, whereas those other senses, although represented by “裁定”, “裁决”, etc., in legal texts, are not as frequent. We suspect that the same concept may possibly be represented by “判” or “判處” (to sentence), which were not included in our analysis.
- “裁斷”, on the other hand, was not found at all in the LIVAC corpus. As the term has most to do with findings of facts in a case, but news articles are often more interested in reporting the final outcome of a case, it is therefore hardly used outside legal texts.
- “裁决” and “判決”, in the {裁决/5, 判決/4} sense, are most frequent among our LIVAC sentence samples, which suggests that general texts tend to use the less specific senses of these words.
- In LIVAC, we also observed some non-legal senses of the words, e.g., “裁判” as the umpire, and “裁决” as religious orders. Expectedly these senses rarely occur in the legal-domain corpus.

- Finally, the relative distribution of senses for each word is not at all similar between the two corpora, showing that the proper use of the words is not as strictly observed outside the legal arena, and their legal implications not as seriously considered.

5.4 Translation difference and polysemy

The sense distinctions and semantic relations discussed above have been derived entirely from the contexts of use in the actual corpus, with some consideration of translation difference where available. The plausibility of the manually derived semantic hierarchies in Figures 1 and 2 thus depends on (1) the extent to which we have been able to accurately capture the meaning of the legal terms as found in their actual contexts of use, and (2) the extent to which the translators (by whom the Chinese version of the judgments were produced) have been able to accurately capture the meaning of the legal terms used in the original English version of the judgments. To see how these two factors might have affected our analysis, we look up Black's Law Dictionary for definitions of the noun senses found for the Chinese terms in our samples:

Decision: A judicial determination after consideration of the facts and the law; esp., a ruling, order, or judgment pronounced by a court when considering or disposing of a case.

Judgment: A court's final determination of the rights and obligations of the parties in a case. The term judgment includes a decree and any order from which an appeal lies.

Verdict: 1. A jury's finding or decision on the factual issues of a case. 2. Loosely, in a nonjury trial, a judge's resolution of the issues of a case.

Finding (of fact): A determination by a judge, jury, or administrative agency of a fact supported by the evidence in the record, usu. presented at the trial or hearing.

Conviction: The judgment (as by a jury verdict) that a person is guilty of a crime.

Holding: 1. A court's determination of a matter of law pivotal to its decision; a principle drawn from such a decision. 2. A ruling on evidence or other questions presented at trial.

Determination: A final decision by a court or administrative agency.

Award: A final judgment or decision, esp. one by an arbitrator or by a jury assessing damages.

As we can see, “decision” and “determination” are more or less synonymous as they are cyclically used to define each other, referring to the final decision of the court in the broadest sense. “Judgment” is defined as a kind of determination concerning the rights and obligations of the parties in a case. On the other hand, “verdict”, “finding”, “holding”, and “award” are each a specific kind of the court’s “judgment”, such that “verdict” has to do with factual issues, “finding” with facts supported by evidence, “holding” with matters of law, and “award” with damages assessment. Moreover, “conviction” is some judgment on guiltiness, expressed via a verdict, which would be even more specific. Hence, while we might be skeptical to whether we should lump all the uses of “裁決”, for example, in Figure 2, which sometimes might not be distinguished enough from the contexts, into one single sense of “the court’s decision”, evidence from translation difference helps justify their splitting as terms like “裁決” are in fact used in very polysemous ways.

6. Future work and conclusion

The analysis in this study, as a preliminary step, has enabled us to arrive at a clearer sense distinction of some easily confusable, and semantically related Chinese legal words in the legal domain *per se* as well as in general usage. A lexical resource for legal terminology detailed at the sense level, when scaled up, will be useful in many areas, including:

- natural language processing
- translation, both legal and general
- lexicography
- reference for terminology standardization in the Hong Kong legal system

As mentioned in Section 1, our objectives are (1) to distinguish the senses of Chinese legal terms, especially for semantically related concepts via some WordNet-like constructions, and (2) to study the usage of these terms in more general texts and see how the precise legal senses are preserved. Our next step is to scale up the analysis. For the first objective, we will explore ways to automate the process of rendering the senses according to their semantic relatedness and to construct a bilingual WordNet for legal terms, perhaps with reference to the definitions of the corresponding English legal terms in law dictionaries and the contextual similarities of the Chinese terms. For the second objective, we shall widen the scope of our analysis. On the one hand, we will study other sets of closely related words, e.g., {剔除, 撤消, 駁回, etc.} (strike out, dismiss, etc.), {合約, 合同, 協議, 契約, etc.} (contract, agreement, covenant, etc.). On the

other hand, we will also look at how such words are used in the newspapers of Chinese communities other than Hong Kong, e.g., Beijing, Taiwan, Singapore, etc., to explore any local differences in the expression of legal concepts in Chinese because of differences in the social structures or legal systems.

We have thus taken a first step in comparing the use of a set of semantically related legal terms in Chinese, in a legal-domain corpus of court judgments and a general corpus of news articles. We found that Chinese legal terms were considerably polysemous and their uses not as clear-cut as their English counterparts. Some might in fact be due to misuse or innovation, as suggested by their extremely low relative frequency. This might be a reflection of the longer history of English being used in the Common Law system, whereas the use of Chinese in this domain has been only recently established, which is not mature and standardized enough. Our next step is to scale up the analysis with other sets of closely related words, covering Chinese communities outside Hong Kong, and to automate the classification of the senses to form some WordNet-like lexical resource. In conclusion, the issues discussed thus far are centered at a point where at least three sub-fields of linguistics, including jurilinguistics, corpus linguistics, and lexical semantics, converge. It is also our hope that the modest beginning we have made in this paper could serve to stimulate more research of both theoretical and practical interests.

References

- Ahmad, K. 2001. The role of specialist terminology in artificial intelligence and knowledge acquisition. *Handbook of Terminology Management*, vol.2, ed. by S. E. Wright and G. Budin. Amsterdam and Philadelphia: John Benjamins.
- Department of Justice. 1998. *The English-Chinese Glossary of Legal Terms*. Hong Kong: Hong Kong Special Administrative Region.
- Department of Justice. 1999. *The Chinese-English Glossary of Legal Terms*. Hong Kong: Hong Kong Special Administrative Region.
- Garner, B. A. 1999. *Black's Law Dictionary*. MN: West Group.
- Li, Z. E., and E. Poon. 2000. *English-Chinese Dictionary of Law*. Hong Kong: Commercial Press.
- Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography* 3.4: 235-244.
- Resnik, P., and D. Yarowsky. 1999. Distinguishing systems and distinguishing senses: New evaluation methods for Word Sense Disambiguation. *Natural Language Engineering* 5.2:113-133.
- Sinclair, J. 1987. *Collins COBUILD English Language Dictionary*. London: HarperCollins.
- Wilks, Y. 1998. Senses and texts. *Computational Linguistics and Chinese Language Processing* 3.2:1-16.
- 鄒嘉彥. 1998. 《中文各地區共時語料庫研究報告》。香港：香港城市大學語言資訊科學研究中心。

[Received 5 September 2002; revised 1 May 2003; accepted 26 May 2003]

Benjamin K. T'sou
Language Information Sciences Research Centre
City University of Hong Kong
83 Tat Chee Avenue
Kowloon, Hong Kong
rlbtsou@uxmail.cityu.edu.hk

Oi Yee Kwong
Language Information Sciences Research Centre
City University of Hong Kong
83 Tat Chee Avenue
Kowloon, Hong Kong
rlolivia@cityu.edu.hk

普通法的普通用法： 法律詞匯於法律文本和一般文本的語義差別比較

鄒嘉彥 鄺藹兒
香港城市大學

英語在過去曾是香港法律體系中唯一的官方語文，直至香港回歸，漢語成為另一官方語文，才有「雙語法律」的實行和應用。然而由於大部分法律概念及詞匯源起英文（甚至拉丁文），直接翻譯成中文未必就能精確地、毫無歧義地表達原有的法律理念。這種差異不僅是漢語詞匯語義學中一個重要而有趣的課題，對香港如何落實雙語法律亦是一項挑戰。我們選取了一系列語義相關而又容易混淆的中文法律詞匯，並比較它們在不同範疇中的用法（包括法律語料及一般語料），從而探討中文法律詞匯的多義性，與有關法律意義在兩種文本中的差別。語料分析結果顯示中文法律詞匯的多義性比英文繁複，且在一般語料中用得比較含混。本文亦討論到以英文翻譯的差異以及詞典中定義來判別中文法律詞匯的義項。我們下一步將研究以自動方式建立一個「詞網」式的法律詞匯詞庫，以及把比較範圍擴展至香港地區以外的語料。

關鍵詞：漢語詞匯語義，法律詞匯，語料分析，「詞網」式詞庫，義項判別