The Specialized Vocabulary of Modern Patent Language:
Semantic Associations in Patent Lexis ∗

Darren Hsin-hung Lin [a] and Shelley Ching-yu Hsieh [b]

Department of Foreign Languages and Literature, National Cheng Kung University,

No. 1, University Road, Tainan City 701, Taiwan, R.O.C.

g364375@ms49.hinet.net [a], shelley@mail.ncku.edu.tw [b]

Abstract. This paper presents an analysis of the language of patents, as a contribution to the field of English for Specific Purposes (ESP). While there work appears to fill a niche in the ESP field (and particularly in the English for Occupational Legal Purposes), the present study insists that statistical approach is necessary for compiling patent technical word list for ESP. Since research studies on word associations of patent lexis have been relatively scarce, this paper reports the technique to select appropriate words from high frequency words is required for modern patent language. The research content and statistical investigations building up a patent technical word list which helps learners of modern patent language expand the vocabulary size for a better understanding of patent writing.

Keywords: English for Specific Purposes, intellectual property rights, corpus linguistics, lexical semantics, patent

1   Introduction

In the knowledge economy age, the intellectual property rights (IPR) become the important assets to human beings. Especially to the knowledge industry, the IPR is the key measure of a company competing with others.

The area of IPR includes patent law, copyright law, trademark law and trade secret law, together with some aspects of other branches of the law, such as licensing and unfair competition (American Bar Association, 2010). Besides, intellectual property lawyers are required interdisciplinary knowledge as new development in law generate needs for lawyers with specific backgrounds—patent law, technology law, business law, and economy law. The fact that even when the global markets have been affected by economic recession in the end of 2007, the demand for intellectual property lawyers remains unusually high (World Intellectual Property Organization, 2009) is worth mentioning. As long as novel inventions were created, there is a need for intellectual property law to be enforced to protect human rights and their invisible property for specific purposes.

As globalization has resulted in greater economic growth rapidly, inevitably the challenges of interdisciplinary communication that concerned with intellectual property and other significant sector encounters has increased. This recognition of the importance has brought intellectual property to the limelight. Resulting from such recognition, the recent emphasis that has been placed on using English as the lingua franca to apply patents on an international level and the

application of technical vocabulary for the writing of professional patents becomes an essential issue in applied linguistic research.

2   English for Specific Purposes (ESP)

ESP is now well established as an important and distinct part of English Teaching (Cheng, Sin, & Li, 2008). As English has acquired the status of lingua franca in almost any field of research, the teaching of ESP has generally been seen as a separate activity within English language teaching, and ESP research as an identifiable component of applied linguistic research (Dudley-Evans & St. John, 1998).

Basically, the origins of ESP can be traced back to the 1960s when there is a growing need for the technological and business industries (Swales, 2000). ESP, the prime realization of applied discourse analysis, was later evolved for every specialized area needs appropriate teaching materials. Recently, ESP is utilized as an umbrella term with multitudinous acronyms standing for the various sub-fields (Dudley-Evans & St. John, 1998).

Under ESP framework, there are two major sub-fields, English for Academic Purposes (EAP) and English for Occupational Purposes (EOP) which are distinguished by their research nature and pedagogical tradition (Robinson, 1991; Dudley-Evans & St. John, 1998). EAP concerning students' needs to learn academic language constitutes the majority of ESP, whereas EOP comprises of professional purposes in administration, medicine, law and business, and vocational purposes for non-professionals in work or pre-work situations (Dudley-Evans & St. John, 1998).

In the ESP domain, ELP is an important but comparatively uncultivated corner (Dudley-Evans & St. John, 1998:51). González and Vyushkina (2009) characterize English for Academic Legal Purposes (EALP) is to be used in university degree program, while English for Occupational Legal Purposes (EOLP) is to be used in training practical skills for workplace. Over the years there has been continuing interest in the research of EALP (Bhatia, 1993; Bowles, 1995; Harris, 1997; Feak et al, 2000; Candlin et al, 2002; Badger, 2003; Du, 2009). However, those were carried out were mostly concerned with material development, genre analysis, and curriculum design. Corpus-based studies on EOLP, in contrast, are relatively undeveloped. Badger (2003) once conducted a corpus-based study on law in the genre of newspaper law reports. He found that newspaper law reports serve the same function as law cases do which facilitates law school students to identify rational of the legal decision of the case. Basically, this corpus-based study is innovative, but that is EALP and solely for reading. To be specific, corpus-based applications on EOLP are comparatively unseen and the voice that professional writing gathers the workplace needs entail the directions for future research.

Accordingly, it is confirmed that while EALP is nowadays widely developed for law school students of academic purposes, there is an underlying purpose to build up EAOP, in particular, Patent English, for workplace needs.

3   Vocabulary in ESP

Writing for specific purposes requires familiarity with not only knowledge of the content but knowledge of the language. Unfamiliarity with vocabulary in writing is perceived to be a challenging task for language learners. As the importance of teaching vocabulary in ESP has gained recognition (Swales, 1983), Coxhead and Nation (2001) categorize vocabulary in ESP into four groups: high frequency words, academic vocabulary, technical vocabulary, and low frequency vocabulary.

Nation (2001) defines those words in the use of writing. High-frequency words refer to the most frequently used 2000 words of English that were used in all types of writing. Low-frequency words are the rarely used terms and covered only 5% of all words. Academic words, namely semi-technical or sub-technical vocabulary, were for academic purposes. This vocabulary is common to a wide range of academic fields but is not what is known as high frequency

vocabulary and is not technical in that it is not typically associated with just one field (Chung 2003). In contrast, technical words are the ones used in a specialized field, which are considerably different from subject to subject. As Chung (2003) point to, technical vocabulary is largely of interest and used to people working in a specialized field. In the genre of law, Mellinkoff (1963) suggests legal vocabulary were those of common words with uncommon meanings. For example, merger and acquisition bear the same literal meaning as 'combination' in general English. However, of economic and financial law, merger depicts the acquisition of one company by another. The combination into a single legal entity will increase the benefits to each other is semantically positive. As to acquisition, the combination often bears unequal treatments is often negative.

Since there is very little technical vocabulary in legal disciplines, Harris (1997) analyzed procedural vocabulary extracted from the area of English contract law. His research shows that technical words enhance not only legal reading but strengthen text analytic skills. Denton (2009) covered frequently used legal vocabulary in his teaching. Specific meaning of vocabulary, such as merger and acquisition of economy law, is viewed as concept for him to teach. His research concludes that the learning of terminology for Legal English is the priority that accounts for participants' motivation to foster when learning vocabulary conceptually. Better to say, learning legal vocabulary with concepts of the target context is essential in vocabulary development. Haberstroh (2009) develops the legal academic word list. His research enriches the well-established area of EALP at present day; however, the rapidly growing trend of EOLP remains comparatively undeveloped.

To sum up, a general conclusion can be drawn is that there is a need to prepare inter-disciplines for patent writing, but exploring technical vocabulary with corpus-based approach into such development has the high priority.

4   Compilation of the Patent Technical Word List

4.1   Technical Vocabulary of Intellectual Property

The relevant technical words in this study were mainly obtained from the USPTO (United States Patent and Trademark Office) Glossary[1]. Consider the proper coverage needed for a lexical study, the distribution of each IPR context is taken into consideration beforehand. Figure 1 shows the results of the coverage of technical words in intellectual property.
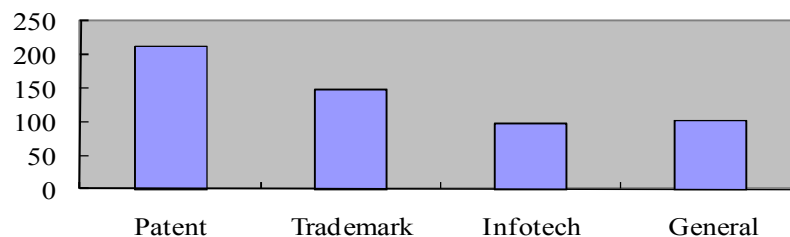
Figure 1: The coverage of technical vocabulary in intellectual property

The coverage was confined to the contexts. As USPTO glossary surveyed, four primary contexts were outlined—patent, trademark, infotech (information technology), and general context. Among the total 558 words in the glossary, 212 words are included in the patent on a context level,

---

[1] USPTO Glossary is available at http://www.uspto.gov/main/glossary

making up 38% of all. Better to say, the coverage of patent technical words was 38% which is much higher than the 18.3% of the 102 words used in a general context as shown in Figure 1. Compared with the coverage of trademark (26.6%) and infotech (17.4%) context, patent technical words are more widely covered in intellectual property. This suggests there is a growing need in the area of patent and is consistent with literature review of the present research which suggests that patent plays a significant role in the genre of intellectual property.

4.2   Frequency of Patent Technical Vocabulary

One of the major objectives in this section was to find out the most frequently used technical words in patents. This aim was achieved by calculating the frequencies of each patent word in Figure 1. The frequency of the patent technical words has been listed according to the frequency of their occurrence in the USPTO Patent Full-Text and Image Database (PatFT),[2] and the distribution is presented in Table 1.

Table 1: Distribution of patent technical words in PatFT

Times of Occurrence	Number of Words	Percentage	Accumulative Percentage
$\cong$ 1,000,000	7	53.20	53.20
1,000,000 ~ 999,999	23	42.56	95.76
10,000 ~ 99,999	21	3.62	99.38
1,000 ~ 9,999	39	0.53	99.91
100 ~ 999	40	0.07	99.98
1 ~ 99	53	0.02	100.00
0	29	0	100.00
Total	212	100	100.00

Based on USPTO Glossary, we retrieved 212 patent technical word families from PatFT, we found 90 word families (99.91%) occurred more than 1000 times, which were considered as frequently used patent technical words. There are only 53 word families (0.02%) appeared less than 100 times and 29 word families did not appear at all, both of which were viewed as not frequently used technical words in patent. The other 40 word families occurred less than 1000 times but more than 100 times in PatFT.

In the present research, there are 7 word families occurred more than one million times. Among them, the most frequently used technical word was a verb 'comprising' which appears 3,785,213 times in PatFT. Other technical word items such as scope, patent, group and element, consisting of, and drawing occurred frequently over one million times. The high-frequency of these words reflects the important role of technical vocabulary in patent texts.

4.3   Word Associations of Patent Technical Vocabulary

With regard to word associations, Nattinger (1988) suggests that grouping of the words according to their meanings enhances vocabulary learning. He once mentioned word grouping can be presented in the form of topic (situational sets) with a library such as book, shelf, borrow, loan, and so on can be taught together for teaching and learning. In order to get a clearer picture of the patent technical words in a better use, the researcher made a detailed analysis into the word associations based on topic.

---

[2] USPTO Patent Full-Text and Image Database (PatFT) is available at http://patft.uspto.gov/netahtml/PTO/search-adv.htm

The 212 patent technical words are considered to be statistically unusually frequent in their occurrence, but it was then noted that they seemed to fall into a limited number of recurring topic sets; six sections were therefore proposed based on words in the same semantic network or field that share similar meanings or semantic features in PatFT: 'patent activity (99)', 'patent aid (25)', 'patent community (23)', 'patent claim (17)', 'patent description (30)' and 'people of the patent community (18).' This was not only made on an intuitive basis, but on the criteria of the produced data. The following illustrates the criteria the researcher sets up for each section.

(1) What do patent-specific activities usually comprise? (patent activity)
(2) What tools can be applied in a patent-specific context? (patent aid)
(3) Where are patent-specific places in United States? (patent community)
(4) What entities do patent applicants need for specific requests? (patent claim)
(5) What specific entities can usually be found in US patents? (patent description)
(6) Who are in the patent-specific contexts? (people of the patent community)

Table 2 presents the top 10 frequently used technical word in each of the six sections, which are arrayed according to their frequency of occurrence in descending order in PatFT.

Table 2: Top 10 technical word items of six topic-based sections

Rank	Activity	Aid	Community	Claim	Description	People
1	patent	concept	Group	comprising	specification	representative
2	disclosure	doctrine of equivalents	Pubs	scope	sequence listing	person
3	application	file wrapper	TC	element	filing date	assignee
4	patent application	ADS	Technology Center	consisting of	serial number	applicant
5	continuation	mask work	ISA	drawing	application number	inventor
6	interference	EFS	IB	dependent claim	PLT	practitioner
7	demand	PAIR	RO	composed of	Control No.	attorney
8	restriction	OG	IPEA	independent claim	publication number	disclaimer
9	designation	PSIPS	GAU	benefit claim	issue date	CSR
10	divisional application	PALM	Group Art Unit	priority claim	patent number	lawyer

The keyword analysis made on a large number of words in this thesis was not intended solely to keep interdisciplinary learners informed of the frequency of some word items, but also to awaken the learners to the influence of intellectual property and patent on lexical units, which might vary in accordance with the different topics as the research drawn.

4.4   Frequency of Word Associations in Patent Lexis

In addition to the top 10 word items, the researcher calculates the total frequency and total word of each section. Table 3 shows the total frequency and total words of each topic section.

Table 3: Frequency of the patent technical word list

Topic	Technical Words	Total Frequency	Percentage	Rank
Patent Activity	99	6,622,873	28	2
Patent Claim	17	12,695,484	54	1
Patent Community	23	1,455,693	6	3
People of the Patent	18	1,468,215	6	3
Patent Description	30	1,060,782	4.5	5
Patent Aid	25	342,988	1.5	6
Total	212	23,646,035	100	

In the patent technical word list, patent claim account for 54%, followed by patent activity, making up 28%, patent community and people of the patent come next at 6%, and finally come patent description represent 4.5%. Patent aid constitutes only 1.5% of all.

As patent law 35U.S.C.§112 paragraph 1 reads, "patent claim" is viewed as the specification contain a written description of the invention, and of manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention. That is to say, patent claims of a published patent inform the public the scope of rights that distinguished the invention. As it is technically dealt with specific terms used, it allows the users to familiarize with the invention an applicant owns.

As shown in Table 3, "patent claim" which has high priority is valuable for a corpus-based research. Besides, to build up a small-scale corpus for the present research, the researcher analyzed 'patent claim' based on verb and noun word for further investigation. Table 4 shows the results.

Table 4: Distribution of patent claim in the patent technical word list

Group	Patent Technical Word	Total Frequency	Percentage
Noun	scope	2,459,656	55.24
	element	1,245,265	
	drawing	1,015,261	
	dependent claim	625,886	
	independent claim	587,926	
	benefit claim	437,599	
	priority claim	381,352	
	withdrawn claim	227,433	
	canceled claim	32,306	
	multiple dependent claim	494	
	rejoinder	80	
	claims	6	
Verb	comprising	3,785,213	44.76
	consisting of	1,165,427	
	composed of	617,353	
	consisting essentially of	114,211	
	having	16	
Total		12,695,484	100

As can be seen, noun words outperform verb words making up a 55.24%. As noun words provide a comprehensive view of patent claim texts, it is noted that more efforts have to be made to explore the possibilities. Based on this, the researcher suggests future studies research on noun words of patent claim to compile a patent technical word corpus.

5. Conclusion

There has been little investigation probed into the modern patent language in applied linguistic research. The present research therefore fills the gap by compiling a patent technical wordlist regarding frequently used word items of six primary patent areas in Table 2. Such word list is significant in that it can help learners expand the vocabulary size by displaying which words they should learn. Since patent claim denotes the precise legal definition of the invention and identifies the specific elements of the invention for which the inventor is claiming rights and seeking protection, the compilation of patent technical word corpus of patent claim texts entails the direction for future research.

References

American Bar Association. 2010. *ABA section of intellectual property law: Introduction*. Retrieved January 7, 2010 from http://www.abanet.org/intelprop/intro.html

Badger, R. 2003. Legal and general: Toward a genre analysis of newspaper law reports. *English for Specific Purposes*, 22, 249-263.

Bhatia, V. K. 1993. Analyzing genre: Language use in professional settings. New York: Longman.

Bowles, H. 1995. Why are newspaper law reports so hard to understand? *English for Specific Purposes*, 14, 201–222.

Candlin, C. N., Bhatia, V. K. and Jensen, C. H. 2002. Developing legal writing materials for English second language learners: Problems and perspectives. *English for Specific Purposes*, 21, 299-320.

Cheng, L., Sin, K. K. and Li, J. 2008. A discursive approach to legal texts: Court judgement as an example. *The Asian ESP Journal*, 4(1), 14-28.

Chung, T. M. 2003. Technical vocabulary in specialized texts. *Reading in a Foreign Language*, 15(2), 103-116.

Coxhead, A. and Nation, P. 2001. The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for Academic Purposes*, pp. 252-267. Cambridge: Cambridge University Press.

Denton, J. 2009. Content vs. concept: Two different focuses in the teaching of Legal English. *Proceedings of ESP Seminar: English for Legal Purposes*, pp. 4-11. National University of Kaohsiung, Taiwan.

Du, J. 2009. Content and language integration in tertiary education in China: A case study in Wuhan Law College. *The Asian ESP Journal*, 5(1), 61-77.

Dudley-Evans, T. and St. John, M. 1998. *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.

Feak, C. B., Reinhart, S. M. and Shinshimer, A. 2000. A preliminary analysis of law review notes. *English for Specific Purposes*, 19, 197-220.

González, M. and Vyushkina, E. G. 2009. International cooperation in designing effective methods to prepare non-native EFL teachers for training and assessing Legal English skills. *Georgetown Law Global Legal Skills Conference IV*, Washington D.C., United States. June 4-6.

Haberstroh, J. 2009. The LAW List (The Legal Academic Word List). *Georgetown Law Global Legal Skills Conference IV*, Washington D.C., United States. June 4-6.

Harris, S. 1997. Procedural vocabulary in law case reports. *English for Specific Purposes*, 16(4), 289-308.

Mellinkoff, D. 1963. *The language of the law*. Boston: Little, Brown and Co.

Nation, I. S. P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Robinson, P. C. 1991. *ESP today: A Practitioner's guide*. New York: Prentice Hall.

Swales, J. M. 1983. Vocabulary work in LSP: A case of neglect. *Bulletin CILA*, 37, 21-34.

Swales, J. M. 2000. Language for specific purposes. *Annual Review of Applied Linguistics*, 20, 59-76.

World Intellectual Property Organization. 2009. *WIPO intellectual property handbook: Policy, law and use*. Geneva: WIPO Publication.