Lexical Primings of Licensing Agreements: A Preliminary Investigation

Wei-chen Chuang[1]

[1] Department of Applied Foreign Languages, Shih Chien University, Taipei, Taiwan

Correspondence: Wei-chen Chuang, Department of Applied Foreign Languages, Shih Chien University, Taipei, Taiwan. E-mail: weichenchuangatshihchien@gmail.com

Abstract

Legal texts such as licensing agreements have achieved a certain degree of notoriety because of their complexity. An analysis of English used in licensing agreements would provide teachers of English for specific purposes (ESP) valuable information for preparing appropriate course materials. Recent theories on the role of a lexicon in language and language acquisition have indicated the importance of the collocational patterns of the key lexicon of an ESP genre (Sinclair, 1991; Hoey, 2005). Based on Hoey's (2005) theory of lexical priming, the current study examined primary lexical priming patterns through identifying the essential key lexical priming patterns of licensing agreements in high-tech industries and exploring their semantic and pragmatic associations. For this study, 60 samples of licensing agreements were randomly selected according to Biber et al. (1990) for determining the linguistic variability required for a sample text to be included in a specialized corpus. The results showed that the lexical primings of the studied keywords substantially converged both semantically and pragmatically. These findings on lexical primings provide an approximate plot illustrating how the semantic and pragmatic associations of the keyword lexical primings combined with the lexical primings constitute the linguistic basics through which the purposes of granting a license are served. Implications for the pedagogical applications of the findings are discussed.

Keywords: licensing agreements, lexical priming, corpus

1. Introduction

A licensing agreement is a "written contract" between the owner/licensor of a patent, copyright, trademark, know-how, service mark, or other intellectual property, and a licensee to use, make, or sell copies of the original" (*Black's Law Dictionary*, "Licensing Agreement"). Because original design manufacturing (ODM) and original equipment manufacturing (OEM) have been the major business models of high-tech industries in many newly industrialized countries, it is not unusual that their manufacturers must acquire know-how and technologies that they do not possess from foreign suppliers. Thus, signing licensing agreements in English to obtain the rights required (such as the right to manufacture and obtain the related know-how) is a common practice in the high-tech industries of these countries. However, understanding and absorbing all of the information presented in a licensing agreement in English is not easy, not to mention the difficulty of writing a licensing agreement in English. The English used in licensing agreements is extremely different from the English familiar to business professionals. In fact, because of highly complex nominal expressions, heavy embedding and lengthy texts, legal texts such as licensing agreements have achieved a certain degree of notoriety. The complexities of their lexico-grammatical features affect the reading and understanding of legal texts (Bhatia, 1993). People may require years to sufficiently comprehend this particular genre of English without proper instruction. For those whose native languages are not English but who must be engaged in the licensing process, such as legal practitioners, high-ranking executives of high-tech corporations, and executive assistants, legal language renders a complex type of discourse that has been rarely taught at universities the colleges in which they have received their education. Thus, higher-educational programs of law, high technology, international trade, and applied foreign languages (whose graduates can work in high-tech industries) must include English for specific purposes (ESP) on licensing agreements in their English training. However, for most English teachers teaching English as a foreign language (EFL) at the college level, licensing agreements constitute a genre with which they are unfamiliar. Thus, to provide relevant language training on the English used in licensing agreements, the essential

linguistic features of licensing agreements must be analyzed, described, and explained systematically to develop instructional versions of the genre comprising of these essential features.

More than one aspect of the linguistic features of a genre may deserve study, and more than one approach can explore the linguistic features of a genre. Among the many issues meriting attention from ESP researchers, investigating genre-specific vocabularies has been a major interest of many studies on ESP and English for academic purposes (EAP) (e.g., Handford & Matous, 2011; Martinez, Beck, & Panza, 2009; Mudraya, 2006; Scott & Tribble, 2006). Some ESP researchers have called for limiting such studies to discipline-specific word lists within individual genres (Martinez, Beck, & Panza, 2009; Mudraya, 2006). EAP and ESP studies have begun to focus on generating a genre-specific or even discipline-specific vocabulary list for teachers and students engaged in teaching/learning ESP. However, the theories on the role of a lexicon in language and language acquisition developed by Sinclair (1991) and Hoey (2005) suggest that generating vocabulary lists may be insufficient for lexicon-based ESP studies; analyzing the collocational patterns of the essential lexicon of a genre is essential.

1.1 Theoretical Background of Analyzing Collocational Patterns

Sinclair was one of the pioneers who recognized the value of collocations in describing English as a language. He contrasted two models that account for how meaning arises from language text: the open-choice and idiom principles.

1.1.1 Open-Choice Principle

The open-choice principle considers a text the result of a set of complex choices and is a "slot-and-filler" model with the nodes on a tree structure, which illustrates the configuration of a sentence, as choice points. According to the principle, to complete a grammatical unit (a node on the tree structure), a person must choose from a large range of choices, with grammatical rules as the only restraints. Because a text is composed of sentences, a sentence may consist of several phrases and/or clauses. To complete a text, therefore, requires making a series of complex choices at each choice point to complete each grammatical unit of a text, which consists of various grammatical units belonging to various grammatical levels (Sinclair, 1991).

1.1.2 The Idiom Principle

Sinclair claimed that the open-choice principle cannot account for the actual processing of a text. He indicated that words do not occur randomly in a text and that the open-choice principle does not provide enough restraints on consecutive choices. Both the nature of the world around us and the registers within which a text occurs impose restraints on the choices made. Even after considering the nature of the world and register, a language user still has too many lexical items to choose from at each choice point. Sinclair proposes the idiom principle to account for the restraints that are not captured by the open-choice principle. The idiom principle is that a language user has a high number of semi-preconstructed phrases stored in his/her mind, which constitute single choices in language processing. These semi-preconstructed phrases may appear analyzable into segments, but they function together as a whole with certain semantic meanings. Sinclair illustrated the existence of semi-preconstructed phrases by offering the example of the simultaneous choice of the two words *of course*. This phrase functions as a single word, and it is argued that the phrase should be treated as a compound. In addition to phrases such as *of course,* there are words that appear to be chosen in pairs or groups but are not necessarily adjacent. These are called "collocations," and they also illustrate the idiom principle. Sinclair insightfully indicated that usually it is combinations of words rather than single words that are the basic units of meaning.

Based on Sinclair, Louw (1993) and Stubbs (1995, 1996) also provided critical findings on collocations. However, Hoey's *Lexical Priming* (2005), a comprehensive theory built on lexical priming, offers an explanation of collocation that prioritizes the role of a lexicon in language and language acquisition.

According to generative grammar, grammar has been argued to be generated first, with words filling in the syntactical slots to make sentences. Similarly, semantics is generated first, and the lexis simply actualizes it. A lexis, therefore, is at a level in a language system that is widely separated from that of grammar. These views of language are familiar to most linguistic students.

By contrast, Hoey maintains an extremely different view on language, particularly regarding the role of a lexicon in a language; he argues that these generally accepted theories on grammar and semantics account for what is possible in a language rather than what sounds natural in a language. Following Sinclair on collocation, Hoey claims that an essential factor in naturalness is collocation. The phenomenon of collocation is pervasive in a language and is how a language learner encounters most of the words he/she learns. In other words, a person encounters a word with its lexical co-occurrence and may encounter it with the same lexical co-occurrence

repeatedly, which introduces the concept of "priming." Priming has been a psychological concept discussed by psycholinguists as the semantic association referring to a word that provokes a particular target word (e.g., Anderson, 1983; Neely, 1977, 1991). Psycholinguists have focused on the relationship between the prime and the target. Hoey, in applying the psycholinguistic concept of priming, shifted the focus of discussion to the priming item per se. Hoey suggests that collocation can be accounted for only when every word is assumed to be mentally primed for collocational use. He explains that a word is encountered repeatedly co-occurring with certain other words, certain semantic sets, particular pragmatic functions, and many other collocational patterns in certain types of context. Such co-occurrences become part of a person's knowledge of a word. Thus, as a speaker/writer makes an output in a certain context, led by priming, he/she tends to use appropriate collocational patterns stored in his/her mind.

Hoey listed 10 priming hypotheses:

1)   Every word is primed to occur with certain words called "collocates."

2)   A word is primed to occur not only with particular other words but also with particular semantic sets, which combine to form the "semantic associations" of a word.

3)   In addition to collocates, semantic associations, and pragmatic associations, a word also has its "colligations," which are certain grammatical positions or grammatical functions (e.g., subject, object, etc.) in which a word is primed to occur or avoid to occur.

4)   Although cohyponyms and synonyms have closely related or even similar references, they differ with respect to their collocations, semantic associations, and colligations.

5)   For a word that is polysemous, each connotation of the word has its own collocations, semantic associations, and colligations that are different from those of the other connotations of the word.

6)   If a word has multiple meanings, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses.

7)   Every word is primed for use in one or more grammatical roles (e.g., noun or verb), which are defined as the grammatical categories of a word.

8)   Every word is primed to (or avoid to) participate in particular cohesive chains, which constitute its "textual collocation."

9)   Every lexical item is primed to occur (or avoid to occur) as part of a specific type of semantic relation (e.g., contrast, comparison, and time sequence), which is defined as its textual semantic association.

10) Every word is primed to (or avoid to) occur at the beginning or end of "chunks" of text, which are its textual colligations.

One characteristic of lexical priming deserving the attention of ESP scholars and practitioners is its sensitivity to contexts; that is, when lexical primings are studied in different genres, they are likely to be described differently in different genres. Thus, studying genre-specific lexical primings is crucial for ESP scholars.

In summary, according to the theories of Sinclair and Hoey, the basic lexical unit that carries meaning is more frequently a semi-preconstructed lexicobundle than a single word. Thus, a list of the key lexicobundles of a genre and the lexical knowledge regarding them, including knowledge on any aspect of the lexical primings of the keywords of a genre, should have greater pedagogical implications than a simple, must-learn vocabulary list. Considering the limited scope of a preliminary study on the collocational patterns of the studied genre, the researcher focused on the major collocations of specific keywords and their semantic and pragmatic associations.

2. Method

2.1 Corpus-Based Study of Collocational Patterns

Numerous recent studies have used corpora in confirming collocational patterns and established the lexical primings of various genres used in professional settings, information that is valuable for ESP teachers (e.g., Fan, 2009; Handford & Matous, 2011; Martinez et al., 2009; Mudraya, 2006; Nelson, 2006; Peacock, 2012; Walker, 2011). A specialized corpus was compiled for the current study, and AntConc, a freeware concordance software program, was adopted for analyzing the collocational patterns of the key lexis.

2.2 A Specialized Small Corpus

For this study, a specialized corpus was compiled. Generally, the greater the variation in a studied corpus text is, the more samples are required to ensure "representativeness" (Meyer, 2002). A specialized corpus tends to have

less internal variation than a general one; accordingly, the size of a specialized corpus with acceptable representativeness can be smaller than a general corpus (Sinclair, 2005).

Bhatia et al. (2004) provided two critical points on studying legal genres that are relevant to designing corpus-based studies on legal genres: legal genres should be considered as a continuum, and corpora of different sizes should be used to study different types of legal genre. Legal discourse, as indicated by Bhatia et al., can be placed on a continuum; on one end are the most conservative forms of discourse such as legislation, and at the other end are literary discourses such as the problem-question genre written by law students in an academic setting. Legislative genres such as rules and regulations, contracts, and agreements are written in the manner to restrict textual interpretations. Thus, the form-function correlations in legislative genres are somewhat fixed. For example, as described, complex prepositions with similar syntactic forms such as *in accordance with*, *in pursuance of*, and *by virtue of* are primed to occur with specific pragmatic associations in legislative texts. By contrast, liberally interpreted genres occur at the other end of the continuum; hence, the form-function correlations of these genres are relatively less fixed, with a range of variation existing in the collocational patterns of a word. For instance, the verb-noun collocational patterns of several verbs in law cases, a genre near literary discourse on the continuum, show greater variance in form-function correlations than those in legislative texts (Bhatia, Langton, & Lung, 2004).

Because of the varying fixedness of form-function correlations between the conservative and liberal forms of discourse, a small corpus may be satisfactory for analyzing linguistic features such as collocational patterns and their pragmatic associations of the conservative forms of discourse, whereas a large corpus is required for analyzing the liberal forms of discourse. .

Bhatia et al. claims that an analysis based on a low number of texts may generate valid findings on the characteristics of legislative genres. The findings of several studies investigating the same legislative genres and using variously sized corpora support his point that the findings of studies using a large corpus of legislative genres and those of studies using a small corpus show a high degree of convergence (Bhatia et al., 2004). For a study focusing on licensing agreements, a genre near the most conservatively interpreted end of the legal discourse continuum, a small corpus may be adequate.

Although the size of a specialized corpus can be smaller than a general one, exactly how much smaller a specialized corpus can be without affecting representativeness remains a question. Biber (1993, p. 243) defined representativeness as the "extent to which a sample includes the full range of variability in a population" and determined that *situational* variability and *linguistic* variability are critical when considering representativeness. According to Biber, situational variability means the range of genres and the speech situation in a target population (1993). Even within a specified genre, some variability may occur within it, thereby ensuring that possible variations are included in a corpus is crucial. Research articles in applied physics can serve as an example. Although articles published both in the *Journal of Applied Physics and in Applied Psychics* Letters are based on research in the field of applied physics, variances in organization and language choices between the two are expected because the former publishes full research articles whereas the latter prefers concise reports. If a corpus of research articles in applied physics included research articles from the *Journal of Applied Physics* but failed to include articles published in *Applied Physics Letters*, the representativeness of the corpus would be in question. A compiled corpus should reflect the criteria of both types of variability to ensure the representativeness of the corpus. However, situational variability should be considered first because no method exists for establishing a corpus that is linguistically representative without situational representativeness.

Linguistic variability is the range of linguistic distributions found in a population (Biber, 1993). Linguistic variability can be achieved by having a sufficient number of samples per register or genre with an adequate number of words per text sample. Based on the results of numerous statistical tests, Biber (1990) suggests that studies use text samples containing a minimum of 1,000 words because the most common linguistic features (e.g., personal pronouns, contractions, past and present tenses, and prepositions) become stable in their frequencies across 1,000-word samples and that studies have a minimum of 10 (and in some situations only five) text samples for a register or a genre in a corpus. Biber indicates that having hundreds of texts with millions of words in a corpus is unnecessary for reliable results.

Although reasonable guidelines for compiling a representative corpus have been provided, there is no method that guarantees the "representativeness" of a corpus, nor is there a method for conducting an entirely objective evaluation (Sinclair, 1991; Tognini Bonelli, 2001). Some researchers have even suggested that representativeness must be accepted on faith (Leech, 1991). In fact, the awareness of a corpus not being representative may be post hoc if "the results are skewed in some way" (Koester, 2010, p. 69).

Based on the basic principles and limitations of corpus compliation specialized corpus of licensing agreements exclusively from high-tech industries was compiled for the the current study. Because this study focused on a specified subgenre of contracts (i.e., licensing agreements) near the most conservative form of the legal discourse continuum, the form-function correlations were somewhat fixed; that is, relatively little linguistic variability occurred in this specific subgenre. In addition, because the studied licensing agreements were limited to those used in high-tech industries, the speech situations were somewhat fixed and without wide variation. Thus, a small corpus satisfying Biber's standard for small corpora was an appropriate size for the current study.

Because the licensing agreements in high-tech industries are drafted by lawyers at different law firms, slight variations occur because of the different writing styles of the various writers and because of differing needs across the high-tech industries. Ideally, collecting data from as many law firms and as many high-tech industries as possible would ensure the situational variability more effectively. However, the confidentiality agreements of law firms render it difficult to collect a high number of samples of licensing agreements directly from law firms. An alternative method for collecting a substantial number of samples from diverse sources is to collect the samples from an open source such as websites for legal professionals. Two major websites providing a wide variety of real contracts that include licensing agreements are *FindLaw* and *Onecle*, each of which has a collection of hundreds to thousands of contracts for different purposes provided by various law firms in the United States. (Onecle has 1468 samples on licensing agreements, whereas *FindLaw* has approximately 410 samples)

A specialized corpus of licensing agreements from high-tech industries was compiled using samples collected from *FindLaw* and *Onecle*, which are the largest open sources of contracts currently available. The licensing agreements from high-tech industries contained in the two websites were categorized based on the industries specified in the agreements' provisions. Four major industries were identified: semiconductor, communication, information, and biomedical technologies. According to the requirement of situational variability, this study selected samples from all four industries in compiling the corpus. Sixty samples of licensing agreements (15 from each industry) were randomly selected.

A licensing agreement contains several sections, each of which has its theme and purposes. To reduce the complexity of linguistic variations resulting from the differing sectional themes and purposes, the collected corpus texts were restricted to the licensing granting and sublicensing sections. The selected sample texts were examined to ensure that each sample text met the requirements of Biber et al. (1990), who determined the linguistic variability required for a sample text to be included in a specialized corpus. When a text with less than 1,000 words was removed from the corpus, another text was randomly selected and added if the word count exceeded 1,000. The words of the corpus, complied according to the discussed principles, totaled 60,503.

2.3 The Analytical Approach and the Tool Used in This Study

Based on Hoey's (2005) theory of lexical priming, the lexical features of the keywords were explored through the collocational patterns of the keywords and their semantic and pragmatic associations. AntConc, a freeware concordance software program, was used in this study. Two analytical AntConc functions were applied in this study: the "keyword listing" and "cluster" functions. Keywords refer to words that appear substantially more in one corpus than in another. In this study, the Brown corpus, a general English corpus, was used as the reference corpus. The words that appeared substantially more frequently in the licensing agreement corpus than in the Brown corpus were identified as keywords for this study. Keywords can be a useful starting point for investigating a specialized corpus. Because AntConc can generate a long list of keywords, analyzing each listed word would not be practical and might not be necessary. Hence, the words at the top of the list sorted by frequency were analyzed first. The top 100 words on the keywords list were divided into several tiers based on "keyness", which indicates the statistical significance of the difference between the frequency of a word in the target corpus and its frequency in the reference corpus. Only the words in the top tier, those with the highest keyness, were studied further.

After the top tier keywords were identified, the clusters of each of them were explored. The AntConc "cluster" function, which can generate a list of clusters based on the probability of the first word in the cluster preceding the remaining words, was applied. Each keyword thus had a list of clusters ranked by probability. These clusters were the combinations of a keyword and its possible collocates; the probability of each cluster showed the degree of tendency at which the word sequence occurred. Clusters were selected for further examination for these generated lists.

The final step was to select some of the clusters with high probability and significance for further qualitative study. Selections were based on the quantitative evidence of the probability values and on qualitative evaluations

of the generated cluster lists. The concordance lines and corresponding original texts of the selected clusters were examined carefully to determine the semantic associations of the clusters and describe their pragmatic functions. Through these steps, a description of the preferred clusters (i.e., collocations of a keyword) of each keyword and the semantic and pragmatic associations of the preferred keyword clusters were provided.

3. Results

The keyness results showed that 79 of the top 100 keywords had keyness values lower than 1,000. By contrast, only 15 words showed keyness values between 1,000 and 2,000. Furthermore, only five words showed keyness values higher than 2,000, the top tier. The five words clearly showed considerable lexical significance in the corpus and were selected to be analyzed in this study: *license*, *licensed*, *agreement*, *any*, and *shall*. Table 1 shows the frequency and keyness of the keywords.

Table 1. Frequency and keyness of the keywords

Rank	Frequency	Keyness	Keywords
1	670	3817.713	license
2	452	2701.702	licensed
3	498	2495.551	agreement
4	841	2360.904	any
5	524	2217.602	shall

Among the keywords, *shall* is an auxiliary verb of considerable controversy. Whether it should be used in legal texts has been debated by law professionals, as well as linguists (Elliott, 1989; Lauchman, 2002; Cheek, 2003; Maine Legislative Drafting Manual, 2003). Several law professionals have called for eliminating *shall* and suggest using *must* instead (e.g., Australian Office of Parliamentary Counsel or Plain Language Association). Although many scholars have called for abandoning the use of *shall* in legal texts, Willims (2005) argues that the varied adoption of *shall* in prescriptive texts for various situations can be complicated but may not be completely unnecessary. Because of the complexity of the use of *shall*, using this auxiliary verb in legal texts warrants its own study. Consequently, it was not included in the current study.

3.1 License

The first word on the keyword list was *license*. That *license* occurred substantially more in a corpus of licensing agreements than in a corpus of general English was expected. The AntConc cluster function was applied to see the collocational patterns of *license*.

The cluster results showed that *license* was often collocated with *exclusive* (frequency: 121) on its left and *to* (frequency: 288) and *under* (frequency: 87) on its right, mainly to form (a [*adjectives*]) *exclusive license,* (*to grant to [the licensee] a*) *license to [do something]*, and (*to grant to [the licensee] a*) *license under….* Among these combinations of *license* occurring with high frequency in the collected texts, the semantic and pragmatic associations of *to grant to [the licensee] a license under…* were explored further because of its rarity in general English. The concordance results showed that the preposition *under* in the cluster (*to grant to [the licensee] a*) *license under…* introduced the *Copyrights*, *Patent Rights*, *Technology*, *Intellectual Property*, and the *Know-how* owned by the licenser specified in the definition section of the agreements; they were specific intellectual creations or knowledge, or the ownership of them that was being licensed in the agreement. Thus, this prepositional phrase had the pragmatic function of introducing the scope of what was being licensed in the licensing granting provision.

In summary, the results showed that the pragmatic function of the patterns with *license* was a "declaration" of the granting of license-related rights, which was mainly accomplished using the verb-noun collocation *to grant* + a *license*, whereas restrictions of the rights were accomplished by the prepositional phrase beginning with "under."

3.2 Licensed

The cluster results showed that *licensed* was used as a participle modifier in the corpus. The nouns that it modified included *Products, Technology*, and *Know-how*; the specific referents of which were described in the definition section of a licensing agreement. Concordance lines with the noun phrase *licensed + Noun* usually showed a nominalization of a verb, which often regarded manufacture, trade, or use occurring before the noun phrase to form an extended noun phrase such as "the manufacture of the licensed Product." Furthermore, the phrases usually appeared in licensing granting provisions, as in the following example:

> Subject to the terms of this Agreement, Gilead hereby grants to P&U an exclusive license to use the Gilead Technology only for the manufacture, importation, use, sale and offer for sale of the Licensed Product in the P&U Territory, with a right to sublicense its Affiliates. (FindLaw, 2014).

The pragmatic function of such a provision is to pronounce the act of licensing. However, the phrase *Licensed + Noun/Nouns* was primed to occur with a pragmatic function of restriction within a declaration of licensing.

3.3 Agreement

The clustering results of *agreement* showed that *this Agreement* had the highest frequency of all clusters with *agreement*. 66% of occurrences of *agreement* had *this* preceding it. Because most instances of agreements in the corpus appeared as a capitalized *Agreement* (referring to the licensing instrument itself), the collocate of *Agreement* that occurred more frequently than other words within a one-word span was this.

However, in general English, *the + Noun* generally has a higher frequency than *this + Noun* has. For example, in the Corpus of Contemporary American English (COCA), the frequency of *the document* is 2,337; by contrast, the frequency of *this document* is much lower, with only 667 occurrences. Different from the tendency in general English, the occurrence of *this + Agreement* was 327 in the investigated corpus, whereas that of *the + Agreement* was only 23.

Despite their considerable difference in frequency, no obvious difference in usage was found between the usage of *the Agreement* and *this Agreement*. For example, *under the Agreement* (frequency: 10) was the most frequently occurring three-word collocate of *the Agreement*. However, *under this Agreement* (frequency: 47) had a substantially higher frequency than *under the Agreement* (frequency: 10) in the corpus. The sentential and textual environments of under *this Agreement* and those of *under the Agreement* were compared to determine possibly differentiating characteristics in usage patterns of the two expressions. Each of the 57 concordance lines was examined. One instance of under *the Agreement* was found to be *under the [*] Agreement*, indicating a noun modifier that referred to confidential information; this instance of *the Agreement*, therefore, did not refer to the text itself. Another nine instances of under *the Agreement* were from the text of a certain licensing agreement regarding intellectual rights, which could have been the idiosyncratic expression of a particular writer and not shared by others.

When the legal instrument needed to refer to itself, *this Agreement* was always preferred over *the Agreement*, although with a capitalized *Agreement* the referent of the word having been fixed in the definition section.

The clustering and concordance results of *this Agreement* also indicated that 84% of occurrences were in a formulaic expression with the pragmatic function of specifying the legal basis of the provision concerned. The formulaic expressions were *subject to the conditions and terms of this Agreement*, *in compliance with all terms of this Agreement*, *under this Agreement* or *in accordance with all terms of this Agreement*. The concordance results also showed that these formulaic expressions generally appeared in the provision articulating the granting of a license. Serving as a qualification of the provision, these formulaic expressions jointly formed a provisionary clause to constitute a "speech act of declarations."

3.4 Any

The fourth word on the keyword list was *any*. Because *any* is generally the first word of a phrase or an expression, a list of words was generated based on the probability of *any* as the first word in the cluster preceding the remaining words. The top two clusters of *any* were *any third party* and *any such*.

3.4.1 Any Such

The two-word combination *any such* must be followed by a noun referring back to something stated previously. The semantic set following *any such* frequently contained sublicense-related nouns to form clusters such as *any such sublicense*, *any such sublicensee*, *any such grant of a sublicense*, and *any such permitted sublicense*. The data showed that 42% of nouns following *any such* could be categorized as sublicense-related, whereas the rest belonged to other diverse groups. Finding that a substantial number of the nouns occurring after any such were sublicense related was expected; after all, the collected texts were licensing agreements. The tendency of the pragmatic function associated with most of the *any such* provisions was instructive. Regardless of whether they were sublicense-related, all occurrences of *any such*… were found in provisionary clauses with verbs or verb groups that indicated the pragmatic function of making a promise or a demand. Verbs and verb groups associated with a*ny such + Noun* included *guarantee*, *agree to*, and *shall + (the base form of a verb)* followed by *any such + Noun*, as shown in this example:

GNE guarantees the making of all payments due to SG by reason of Net Sales of any
Licensed Products by any such sublicensee (Onecle, 2014)

3.4.2 Any Third Party

Any third party as a noun phrase was found to collocate frequently with verbs or verb groups such as *sublicense*, *distribute*, *deliver/transfer/make any material*, *authorize*, and *make available*. Concordance results for the textual environments of sentences with the noun phrase *any third party* showed that the term usually served, with limited exceptions, as the objective of these verbs and verb phrases and tended to appear in a provisionary clause imposing restrictions on or granting permissions to third-party rights. Furthermore, 70% of verbs or verb phrases that occurred with *any third party* in a provision were sublicense related, which indicates that most frequently the right of the licensee to sublicense to a third party was the concern being addressed when the formulaic expression appeared. The corpus data also showed that 77% of occurrences of the formulaic expression appeared in a negative provisionary clause regarding third-party rights or one limiting third-party rights. That most licensing agreements restrict the right of the licensee to sublicense rather than grant unrestricted sublicense rights is to be expected; a licensing agreement granting unrestricted sublicensing rights to the third-party licensee rarely occurs. In the examined corpus, the major pragmatic function of provisions in which *any third party* occurred most frequently was to limit or even forbid certain sublicensing rights.

Collectively, the lexical primings of the studied keywords substantially converged both semantically and pragmatically. They showed how the keywords, their major collocations, and the semantic associations of these collocations work together to realize the associated pragmatic functions (i.e., to specify the scope of the rights granted, to make a promise or a demand, and to specify the legal basis). These pragmatic functions expressed the essential communicative purposes of normative texts in a context in which "licensing" was involved.

3.5 Pedagogical Implications

A summary of the major findings discussed in the result section are shown in Table 2.

Table 2. Major findings regarding the lexical priming of the keywords

Keywords	Major collocations	Semantic associations of the collocations	Pragmatic associations of the collocations
License	- *to grant to + [the licensee] + a license + under ...*	Intellectual property or knowledge or the ownership of them (e.g., *to grant to ICB a license under Patent Rights and Technology*)	To restrict or to specify the scope
Licensed	- *Licensed products* - *Licensed technology* - *Licensed know-how*	Nominalized verbs related to manufacturing, trading or using (e.g., *importation of the Licensed Product*)	To specify the scope; to limit
Agreement	- *under this Agreement* - *subject to the terms and conditions of this Agreement*	N/A	To specify the legal basis of the following regulations
Any	- *any such + [a noun]*	Verbs associated with promises or demands (e.g., *GNE guarantees the making of all payments due to SG by reason of Net Sales of any Licensed Products by any such sublicensee ...*)	To make a promise or a demand
	- *any third party*	Sublicensing-related verbs or verb phrases [e.g., *In Categories B, C, D, and E and, subject to Section 8.3, Category F, Codexis may grant sublicenses pursuant to this Section 2.1(c) to any Third Party*]	To specify or limit the rights relating to a third party

Because these lexical primings are essential to licensing agreements, novice drafters of licensing agreements must be familiar with them. Teachers may use Table 2 as a reference in developing teaching materials that are relevant for a course on licensing agreement English, particularly when it is for beginners of legal English. Exercises and activities should be developed to assist students in acquiring the listed formulaic expressions as well as the sentence patterns in which the keywords are frequently embedded. Students should also be encouraged to develop an awareness of the pragmatic functions of the formulaic expressions.

Because they occur rarely in general English, some expressions and usages in this study deserve particular attention from ESP teachers. For example, formulaic expressions such as *subject to the conditions and terms of this Agreement, in compliance with all terms of this Agreement, under this Agreement*, and *in accordance with all terms of this Agreement* are commonly written in legal texts to specify the legal source of the regulations as the legal instrument itself. However, they may occur infrequently in general English. Expressions such as *subject to the terms of + [a legal or a diplomatic agreement]* and *subject to the terms and conditions of + [a legal or a diplomatic agreement]* are often observed in legal or diplomatic documents. However, they are not the formulaic expressions frequently encountered by English learners who are not students of law or international relationships. The two expressions occurred 83 times in the investigated corpus of only 60,503 words. By contrast, *subject to the terms of or subject to the terms and conditions of occurred* only once (an extremely low frequency) in the COCA, a general English corpus with 450 million words. Thus, ESP teachers teaching legal and political genres may need to acquaint their students with this formulaic expression. Another example is the formulaic expression *(to grant to [the licensee]) a license under + [specific intellectual creation or knowledge or the ownership of them that are being licensed in the agreement]*, which typically constitutes parts of the provision granting the license. The chance of encountering such a formulaic expression in general English may be extremely low; developing the psychological priming of a similar collocational pattern may require additional instruction. However, because ESP teachers may not be familiar with legal English (despite having reached an advanced level of general English), they must learn these expressions as collocational patterns first.

The contents of Table 2 can also serve as the minimum standard for evaluating the appropriateness of basic teaching materials for licensing agreement English in high-tech industries. Although the table does not comprehensively include all materials for an ESP course on licensing agreements (additional studies must be conducted to disclose other critical linguistic knowledge on licensing agreements), the listed contents are essential for a beginner learning licensing agreement English. Teachers can use the table to determine whether the basic lexicon of licensing agreements used in high-tech industries have been included in a course at the primary level.

4. Conclusion

The keyword approach adopted in this study identified several linguistic features of the English used in licensing agreements, thereby validating the keyword approach as appropriate for yielding useful information on the linguistic features of the target genre regarding the lexical priming of the major keywords of the genre.

The findings of this study support the speculation that licensing agreement English exhibits specific linguistic features; the linguistic features in lexis and lexical primings characterize the target genre as a genre realized by a varity of English which has deviated from general English. These findings also confirm that instruction on the linguistic features of the target genre is necessary because of the linguistic differences between general English and licensing agreement English. Without proper instruction, even advanced EFL learners may not be equipped with the requisite linguistic knowledge to enable them to function effectively in legal professions.

References

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings.* New York: Longman.

Bhatia, V. K., Langton. N. M., & Lung, J. (2004). Legal discourse: opportunities and threats for corpus linguistics. In U. Connor & Thomas A. Upton (Eds.), *Discourse in the professions: perspectives from corpus linguistics.* (pp. 203-226). Amsterdam/Philadelphia: John Benjamins. http://dx.doi.org/10.1075/scl.16.09bha

Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing, 5*, 257-269. http://dx.doi.org/10.1093/llc/5.4.257

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243-257. http://dx.doi.org/10.1093/llc/8.4.243

Black's Law Dictionary: Free Online Legal Dictionary (2nd ed.). (2015, April). Retrieved from http://thelawdictionary.org/

Cheek, A. (2003). Writing user-friendly documents. Retrieved fromhttp://www.fda.gov/ohrms/dockets/dailys/03/Mar03/033103/8005a2d

Elliott, D. C. (1989). *Constitutions in a modern setting—the language of the practice of law.* A paper presented at Lawasia Conference, Hong Kong, September 1989.

Fan. (2009). An exploratory study of collocational use by ESL students—a task based approach. *System, 37*, 110-123. http://dx.doi.org/10.1016/j.system.2008.06.004

FindLaw. (2014). Development and license agreement. Retrieved from http://corporate.findlaw.com/contracts/operations/development-and-license-agreement-gilead-sciences-inc-f.htm

Handford, M., & Matous, P. (2011). Lexicogrammar in the international construction industry: A corpus-based case study of Japanese–Hong-Kongese on-site interactions. *English for Specific Purposes, 30*, 87-100. http://dx.doi.org/10.1016/j.esp.2010.12.002

Hoey, M. (2005). *Lexical priming: a new theory of words and language*. London and New York: Routledge. http://dx.doi.org/10.4324/9780203327630

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: MA: Harvard University Press.

Koester, A. (2010). Building small specialized corpora. In A. O'Keeffe & M. McCarthy (Eds.), *Routledge handbook of corpus linguistics* (pp. 66-79). London; Routldege. http://dx.doi.org/10.4324/9780203856949.ch6

Lauchman, R. (2002). Plain Language: A handbook for writers in the U. S. federal government. Retrieved from Http://www.mindspring,com/~rlauchman/PDFfiles/PLHandbook.PDF

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker et al. (Eds.), *Text and technology* (pp. 157-176). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/z.64.11lou

Maine Legislative Drafting Manual. (2003). Retrieved from http://janus.state.me.us/legis/ros/manual/Contents.htm

Martinez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*, 183-198. http://dx.doi.org/10.1016/j.esp.2009.04.003

Meyer, C. (2002). *English corpus linguistics.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511606311

Meyer, C. (2005). Corpus and text—Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: a guide to good practice* (pp. 1-16). Oxford: Oxford Books.

Mudraya, O. (2006). Engineering English: a lexical frequency instructional model. *English for Specific Purposes 25*, 235-256. http://dx.doi.org/10.1016/j.esp.2005.05.002

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: role of inhibitionless spreading activation and limited capacity attention, *Journal of Experimental Psychology: General, 106*, 226-254. http://dx.doi.org/10.1037/0096-3445.106.3.226

Neely, J. H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner & G.W. Humphreys (Eds.), *Basic processes in reading: visual word recognition* (pp. 264-336). Hillsdale, NJ: Lawrence Erlbaum.

Nelson, M. (2006). Semantic associations in business English: a corpus-based analysis. *English for Specific Purposes, 25*, 217-234. http://dx.doi.org/10.1016/j.esp.2005.02.008

Onecle. (2014). License agreement. Retrieved from http://contracts.onecle.com/seattle-genetics/genentech-license-cd40-2003-03-06.shtml

Peacock, M. (2012). High frequency collocations of nouns in research articles across eight disciplines. *Iberica, 23*, 29-46.

Scott, M., & Tribble, C. (2006). *Textual patterns: key words and corpus analysis in language education*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. http://dx.doi.org/10.1075/scl.22

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University.

Stubbs, M. (1996). *Text and corpus analysis*. Oxford: Balckwell.

Stubbs, M. (1995). Corpus evidence for norms of lexical collocation. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 245-256). Oxford: Oxford University Press.

Swale, J. M., & Bhatia, V. K. (1983). An approach to the linguistic study of legal documents. *Fachsprache, 5*(3), 98-108.

Tognini Bonelli, E. (2001). *Corpus linguistic at work*. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl.6

Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes, 30*, 101-112. http://dx.doi.org/10.1016/j.esp.2010.12.003

Willims, C. (2005). *Tradition and change in legal English: verbal constructions in prescriptive texts*. Bern: Peter Lang.

Copyrights