Collocations of Keywords Found in Insurance Research Articles: A Corpus-Based Analysis

Tanu Khamphairoh[*]

Montri Tangpijaikul[**]

Abstract

The objectives of this study are to identify keywords most frequently used in insurance research articles, and to explore the patterns of collocations of the keywords used in insurance research articles. One hundred and fifty-five research articles from the "Journal of Risk and Insurance" and the "Risk Management and Insurance Review" published between 2007 and 2010 were selected. The data was analysed by the application of the software AntConc 3.2.1w developed by Anthony (2007).

The results derived from the present study through a corpus-based analysis method revealed a list of keywords which were frequently used in the field of insurance. From the selected top one hundred keywords, the collocates to each keyword were identified. Only collocations that met specified criteria were selected and categorised into lists of 2-, 3- and 4-word clusters. Many of the keywords and collocations found in this present study could be useful to ESP learners and teachers who are involved in the field of

[*] Graduate Student in English for Specific Purposes Program, Graduate School, Kasetsart University

[**] Lecturer, Department of Foreign Languages, Faculty of Humanities, Kasetsart University

insurance. The findings obtained from this research could be adapted to classroom materials and also references. Keywords in the field of insurance and their collocational behaviours found in this study would be beneficial to insurance students, teaching professionals, and material designers in the field of insurance.

Keywords: collocations; corpus linguistics; research article

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อวิเคราะห์หาคำหลักที่ปรากฏบ่อยที่สุดและคำปรากฏร่วมของคำหลักในบทความวิจัยด้านการประกัน กลุ่มตัวอย่างประกอบด้วยบทความวิจัย จำนวน 155 ฉบับ ที่ได้รับการตีพิมพ์ระหว่าง พ.ศ. 2550 - 2553 ในวารสารทางวิชาการ Journal of Risk and Insurance และ Risk Management and Insurance Review เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลคือโปรแกรม AntConc เวอร์ชัน 3.2.1w พัฒนาโดยแอนโทนี (2007)

ผลการศึกษาจากการวิเคราะห์โดยวิธีหลักการคัดคำโดยใช้คลังข้อมูล หรือ corpus-based analysis ปรากฏกลุ่มคำหลักที่ใช้ในด้านการประกัน และคำปรากฏร่วมของคำหลัก 100 คำแรก ซึ่งเข้าหลักเกณฑ์การประกอบกลุ่มคำปรากฏร่วมตั้งแต่ 2 ถึง 4 คำ คำหลักและคำปรากฏร่วมหลายคำที่ปรากฏในงานวิจัยนี้จะมีประโยชน์ต่อผู้เรียนและผู้สอนในด้านการประกัน โดยสามารถนำผลลัพธ์ที่ได้จากการศึกษานี้ไปปรับใช้เป็นสื่อการเรียนการสอนและเพื่อการอ้างอิง

คำสำคัญ: คำปรากฏร่วม; ภาษาศาสตร์คลังข้อมูล; บทความวิจัย

Introduction

It has been well noted that 'collocation' plays a major role in language lexical acquisition (McCarthy, 1990). Learning collocation will assist language learners in the way of improving fluency and native-like ability to

select appropriate words in context (Shin and Nation, 2008), so it is crucial to learn and possess skills in the use of collocation.

In the past decade, there have been many studies carried out employing a corpus-based approach with an effort to generate a listing of high-frequency collocations particularly in an academic text. Learners need to know not only words that are important or ones that appear frequently in their academic field, but also their collocational behaviours (Durrant, 2009). Learning collocation will assist language learners in the way of improving fluency and native-like ability to select appropriate words in context (Shin and Nation, 2008), so it is crucial to learn and possess skills in the use of collocation. Although many research studies of collocation in academic texts have been carried out, such as in science (Gledhill, 2000), medicine (Marco, 2000), and engineering (Ward, 2007), so far the study of collocation in the field of insurance is still limited. The insurance Industry is one of the major business activities and 'Insurance Management' is one of the academic programs in many universities. The vocabulary of English texts in the insurance industry is the key to understanding and learning the essential language and terminologies used in the insurance industry for both students and insurance professionals.

For this reason, this study aims to explore keywords that ESP students in the field of insurance need to know; as well as their most frequent collocational patterns as appeared in insurance research articles, so as to offer some useful suggestions for language learners in this particular field. The research questions are:

- What are the keywords found in insurance research articles?
- What are the most frequent collocations of the keywords found in insurance research articles?

Review of Literature

Based on word frequency, Nation (2001: 11-12) divides words in any text into four groups: (1) high-frequency words, (2) academic vocabulary, (3) technical vocabulary, and (4) low-frequency words. High-frequency words are the 2,000 most frequently used words in English or General Service List (GSL) which cover around 80 per cent of running words used in any text (West, 1953). Academic words are 570 words families from the Academic Word List (AWL), most commonly found in academic texts of different genres and fields (Coxhead, 2000). They account for appropriately 10 per cent of running words in any academic texts. Technical words are those that do not belong to GSL and AWL. They could be found in specific subject areas and account for around 5 per cent of the texts. Finally, low-frequency words cover up the smallest proportion of the texts.

Nation's division of words into four groups is useful as words that are considered 'key' should not belong to the first high frequency group. A keyword is defined as "a word which occurs with unusual frequency in a given text". (Scott, 1997: 237) as compared to its number of occurrences in a reference corpus (Scott, 1997). On this ground, keywords that result from the comparison of the two corpora cannot be guaranteed for their semantic specificity in the field of insurance. To identify keywords that are also technical in the field of insurance, the rating scale approach suggested by Chung and Nation (2004) is applied in this study.

As mentioned earlier, a word can be best acquired when studied with their typical collocational patterns. Collocation is the habitual co-occurrence of two or more words in a text within a space of a four words 'span' (+/- 4) to the left and right of the 'node' which represents an item

whose collocations are studied, and those words found within the span are called 'collocates' (Sinclair, 1991). By implementing a corpus-based approach, researchers can study massive number of texts and identify various types of collocates which co-occurred with the node in the form of chunks or collocations (Sinclair, 1991).

Ward (2007) investigated behaviour of collocations in relation to the lexical technicality in engineering. The results revealed the importance of technical terms and collocations in the form of complex noun phrases commonly found in the study, for example, *partial molar enthalpy* and *heat transfer coefficients*. Cortes (2004) investigated the presence of lexical bundles in two corpora: research articles and student writings from the history and biology disciplines. The results showed the relatively low frequency of use of lexical bundles in student writings in both disciplines when they were compared with the frequency of use in published writings. For students to be familiar with and able to use more lexical bundles appropriately, it was suggested that students be exposed to their usage. Sangvaree (2004) found that one of the common problems faced by employees in insurance companies in Thailand is the knowledge of technical terminologies in the insurance field. Taking the above studies into consideration, the importance of collocations in academic texts is obvious. This demonstrates the necessity of students' knowledge of a set of collocations.

Methodology

Data Collection and corpus creation

One hundred and fifty-five research articles (electronic versions) from the "Journal of Risk and Insurance" and the "Risk Management and Insurance Review" published between 2007 and 2010 were selected. The entire text articles were selected in original length. However, there are some parts to be removed because they are not contents of the articles, including charts, diagrams, tables, bibliographies, numbers, references, and appendices. The data was analysed by the software AntConc 3.2.1w which was developed by Anthony (2007). The selected articles were compiled and stored electronically in the corpus under study which I shall call the IRAC (Insurance Research Articles Corpus). There were 980,121 word tokens and 19,476 word types found subject to the analysis with AntConc. The IRAC is a small corpus, and its sampling of the data is aimed to represent English which is used in insurance research articles.

Data Analysis

The present research employs a corpus-based approach in order to explore the list of keywords and the list of collocations of keywords found in insurance research articles. To achieve that, the two steps are: (1) Compiling Corpus and Identifying Keywords and (2) Identifying Collocations of Keywords.

Regarding the compilation of corpus and keyword identification, first a wordlist tool was used to count word frequency in order to identify the word list of the IRAC. In *Wordlist Tool*, the *Stop List* feature was used to filter out function words. After that, the *Keywords* feature was used to identify a list of words in the IRAC which are statistically significant in their frequency of

occurrence by comparing with the reference corpus British National Corpus (BNC) Sampler - one million word collection from written text domain. The BNC Sampler was chosen not only because it has been widely used as a reference corpus, but it also represents general English. After this process, a listing of the top 100 keywords was created. It is worth noting that only those keywords that have a specialised meaning related to the insurance field retrieved by applying "rating scale approach" (Chung and Nation, 2004) were selected for further analysis. At this stage, words were rated manually. The researcher and raters who have expertise in the field of insurance decided on the meanings of individual words and whether they have specialised meanings in the insurance field.

The second step in the analysis involves identifying collocations of the keywords obtained from the first step. The *Cluster* function of AntConc Tools was used to create listings of the common collocates of each keyword within four word span (Jones and Sinclair, 1974). The cutoff point of the target collocations was set with a minimum number of four occurrences; the cluster that appears less than 4 is excluded from the count (Nelson, 2000). This study focuses only on lexical cluster, so function words at the first or last position of the cluster were excluded from the list. Through completing the second step the collocations of the keywords were identified and vocabulary exercises that could make good use of these collocational patterns suggested.

Results and Discussion

The results of the data analysis from the IRAC were presented. The first was the list of the top 100 keywords. The second was the lexical collocations of the keywords.

1. The list of the top 100 keywords

The quantitative results have provided a list of the first 100 keywords whose semantic meanings are related to insurance as appeared in the two selected journals. It is not surprising to see the word *insurance* ranked at the top of the list followed by *risk* and *insurer*. As it can be seen, keywords from the corpus (Table 1) give an idea what the corpus in the study is about.

Table 1: The top 100 keywords whose meanings related to insurance

1	insurance	21	contract	41	distribution	61	rating	81	condition
2	risk	22	company	42	purchase	62	group	82	fund
3	insurer	23	capital	43	wealth	63	private	83	mortality
4	market	24	measure	44	information	64	reinsurer	84	product
5	model	25	financial	45	factor	65	account	85	impact
6	loss	26	year	46	hedge	66	care	86	policyholder
7	firm	27	case	47	management	67	portfolio	87	incentive
8	premium	28	period	48	investment	68	growth	88	deductible
9	cost	29	probability	49	annuity	69	business	89	structure
10	asset	30	life	50	benefit	70	problem	90	pension
11	rate	31	return	51	stock	71	decision	91	security
12	price	32	liability	52	term	72	strategy	92	exposure
13	value	33	reinsurance	53	income	73	adverse	93	medical
14	coverage	34	function	54	bond	74	equity	94	regulation
15	variable	35	analysis	55	property	75	control	95	debt
16	policy	36	utility	56	average	76	aversion	96	allocation
17	claim	37	optimal	57	pay	77	insure	97	volatility
18	health	38	ratio	58	consumer	78	payment	98	reserve
19	individual	39	plan	59	equilibrium	79	interest	99	regulatory
20	level	40	demand	60	selection	80	industry	100	tax

These 100 lexical keywords can be distributed into five grammatical classes. Eighty percent of them were nouns, which made up the major proportion, followed by adjectives (7%), nouns/adjectives (4%). Verbs and nouns/verbs equally make up 2% each. This is not surprising as the most important terms in insurance have meanings related to money and its value (e.g. *cost, price, value, property, fund, asset, equity, liability, wealth*) as well as with measures or regulations (e.g. *contract, measure, claim, average*). Meanwhile, words such as *purchase* and *control* can be both a noun and a verb, as in clusters such as *purchase full coverage* (v.), *mandatory purchase* (n.), *control moral hazard* (v.) and *risk control* (n.).

From grammatical class distribution, the top 100 keywords can be categorised further into five main areas according to their associated meanings as followed:

1. Keyword referring to people, industries and institutes
   e.g. insurer, firm, policyholder and reinsurer
2. Keyword referring to economics & finance
   e.g. fund, investment, value, price, capital and asset
3. Keyword referring to business activities
   e.g. insurance, rating, payment, allocation and analysis
4. Keyword referring to products & contracts
   e.g. health, life, coverage, policy, care, deductible and premium
5. Keyword referring to measures & conditions
   e.g. level, model, period, annuity, regulation and aversion.

It is important to note that the categorisation of these keywords may not be ideal or absolute, since one word may fit into two or more categories and others may have a different view or perspective in classifying

the meanings of words. For example, *deductible* can be defined as insurance contract provision and therefore was categorised under keyword referring to products & contracts. However, others may see *deductible* as a condition of the insurance policy as it refers to an amount of covered loss deducted from loss payment.

2. Most frequent collocations of the keywords found in IRAC

For 2-word clusters, only lexical collocates are included. For 2-4 word clusters, function words at the first and/or last position of the cluster are excluded on the ground that lexical items rather than grammatical ones are the focus of this study.

Shown in Table 2 is the list of the first 10 keywords with their top three 2-word clusters to the left and right of the node.

Table 2: The top 10 keywords with their first six 2-word clusters

No.	Keywords	2-word clusters	No.	Keywords	2-word clusters
1	**insurance**	life insurance	2	**risk**	low risk
		health insurance			high risk
		purchase insurance			mortality risk
		insurance market			risk management
		insurance company			risk aversion
		insurance coverage			risk free
3	**insurer**	life insurer	4	**market**	insurance market
		stock insurer			capital market
		international insurer			stock market
		insurer financial			market share
		insurer participation			market value
		insurer performance			market concentration

No.	Keywords	2-word clusters	No.	Keywords	2-word clusters
5	model	feedback model	6	loss	expected loss
		basic model			earnings loss
		pricing model			aggregate loss
		model framework			loss coverage
		model parameters			loss ratios
		model includes			loss distribution
7	firm	insurance firm	8	premium	insurance premium
		new firm			pure premium
		individual firm			risk premium
		firm size			premium growth
		firm specific			premium income
		firm characteristics			premium rate
9	cost	total cost	10	asset	risky asset
		high cost			optimal asset
		low cost			illiquid asset
		cost sharing			asset allocation
		cost function			asset risk
		cost efficiency			asset reconstitution

Insurance: *insurance market* is the most frequent of the 2-word clusters with collocates to the right and *life insurance* with collocates to the left of the keyword "insurance" in the IRAC. *Insurance* displayed strong meanings when describing the type of insurance products, e.g. *life insurance*, *health insurance*, and *insurance coverage*. Combined with collocates to the right, it is found the formation of words as compound nouns, e.g. *insurance market* and *insurance company*. This is why it is good to learn these as one lexical item.

Risk: learners need to know that risk conveys negative meaning, and that it is something to be avoided, as can be seen in *risk aversion*, or

managed with care, as in *risk management*. If not it could be fatal, as seen in *mortality risk*. Risk comes in degree and appears with the adjectives *high* and *low*.

Insurer: collocates to the left of *insurer* describe the types of insurer – *life insurer*, *stock insurer* and *international insurer*. Insurer is the principal institute in the insurance industry and engages in activities in the insurance business. This engagement is called *insurer participation*, and insurers' performance is known as *insurer performance* for clients.

Market: in the IRAC, *market* is particularly concerned with the types of market in the field of insurance, as can be seen in *insurance market*, *capital market* and *stock market*. When it functions as an adjective, it refers to the prices of things that are not fixed. For example, *market value* refers to the value determined by the demand and supply of the economy system.

Model: insurance is an industry that involves risk and possibility of losses, so there are standardised models in predicting future events, as seen in *feedback model*, *basic model*, *pricing model*, *model framework* and *model parameters*.

Loss: loss is a direct result of risk. In this study, *loss* is found with meanings which refer to the types of losses, such as *expected loss*, *earnings loss* and *aggregate loss*. We purchase an insurance policy so as to have our losses covered (*loss coverage*) and distributed (*loss distribution)* by insurers.

Firm: the word *firm* mostly collocates with adjectives that suggest its qualities, such as *new firm* and *individual firm*. Firm is also referred to in terms of its size and characteristics, as in *firm size* and *firm characteristics*.

Premium: the word *premium* is found to collocate with a wide range of words related to the terms of an insurance policy, as in *insurance*

premium, *pure premium* and *premium rate.* It also collocates with words involving financial investment, as in *premium growth* and *premium income*.

Cost: it can be seen that the word *cost* is likely to be discussed as something that needs to be managed, as in *cost sharing* and *cost efficiency*. Cost is often showed in degree adjectives such as *high* and *low*.

Asset: the word *asset* is found to collocate frequently with words whose meanings are related to economy and finance, e.g. *asset allocation* and *asset reconstitution*. The combination of *asset* with its collocates also showed the types of assets, e.g. *risky asset*, *optimal asset* and *illiquid asset*.

The following Table 3 presents the list of examples of 3- and 4-word clusters of the first 10 keywords from the IRAC. Only the four clusters of each keyword are shown.

Table 3: The top 10 keywords with their first 3-word clusters

No.	Keywords	3-word clusters	No.	Keywords	3-word clusters

1	insurance	cost of insurance
		lines of insurance
		risk management and
		   insurance demand
		for life insurance
			2	risk	aggregate mortality risk
					risk minimization
					   hedging
					risk management and
					   insurance
					coefficient of risk
					   aversion

3	insurer	insurer stock returns
		insurer investment
		   policy
		life insurer stock returns
		volatility of an insurer
			4	market	life insurance market
					cat bond market
					equilibrium in a
					   market
					market for medical
					   malpractice

No.	Keywords	3-word clusters	No.	Keywords	3-word clusters
5	model	asset pricing model	6	loss	excess of loss
		adverse selection model			industry loss warranties
		capital asset pricing model			changes in the loss
		model of insurance demand			increase in loss coverage
7	firm	firm specific characteristics	8	premium	pure premium rate
		relation between firm			actuarially fair premium
		value of the firm			effect of premium growth
		take over the firm			equity premium of stocks
9	cost	cost of capital	10	asset	optimal asset allocation
		cost of insurance			asset risk taking
		cost of public funds			capital asset pricing model
		cost of equity capital			asset allocation and liquidation

It can also be seen from the table above that when keywords are combined in a bigger cluster, they seem to take on meanings more specific to the insurance field.

Insurance: in the IRAC, there appears a typical pattern of a noun + preposition (of) + the word *insurance*, as in *cost of insurance* and *lines of insurance*. In the insurance business there are four major *lines of insurance*, including property, casualty, life, and health and disability. The most frequently

used 4-word cluster in the IRAC is *risk management and insurance*. This could be due to the fact that *'Risk Management and Insurance Review'*, which is the name of the selected journal for this study, is frequently mentioned.

Risk: this word collocates with clusters which are related to a strategy in the prediction of the mortality rate, as in *aggregate mortality risk*, and a risk measure in financial trading strategy, as in *risk minimization hedging*. It is also found in the cluster *coefficient of risk aversion* which is the method of measuring the variables of the degree of taking risks.

Insurer: the collocates of *insurer* indicate a focus on insurance activities, especially in terms of money and investment, e.g. *insurer stock returns* and *insurer investment policy*.

Market: the word *market* is commonly used to indicate the types of insurance markets, such as *life insurance market* and *cat bond market*. It is worth noting that *cat* is the abbreviation for 'Catastrophe bonds', which is one of the debt instruments in risk protection. The word *market* also has meanings related to insurance products for professional liability in the medical field. For example, *market for medical malpractice* is one of the types of insurance products available for professionals such as doctors who might be exposed to the risk of medical malpractice lawsuits.

Model: frequently used 3- and 4-word clusters of the word *model* generally have meanings focused on the names of models, as in *asset pricing model*, *adverse selection model*, *capital asset pricing model*, and *model of insurance demand*.

Loss: when a loss is over the limit to be covered by the primary insurer - *excess of loss* - it is then shifted to the reinsurer. Among many

products of reinsurance, *industry loss warranties* is found in the IRAC. This product provides coverage of catastrophe loss to an industry. *Loss* with its collocate *coverage* in *increase in loss coverage* seems to express positive connotation as something that can be increased or improved.

Firm: the word *firm* is accompanied by words that suggest characteristics of the firm such as *firm specific characteristics* and *value of the firm*. An activity referring to the firm can be seen in *take over the firm*. *Firm* is found in the cluster whose meanings refer to relations, such as *relation between firm*.

Premium: in the IRAC *premium* collocates with words that refer to the calculation of premium rate, e.g. *pure premium rate* and *actuarially fair premium*. For example, *pure premium rate* is the estimate of the future costs of losses arising from claims as indicated in historical experience. It also combined in the clusters that have meanings related to finance and investment, as in *effect of premium growth* and *equity premium of stocks*.

Cost: in the IRAC there is also found the collocational pattern of the word *cost* + preposition (of) + a noun. For example, *cost of capital* refers to the opportunity cost of an investment; *cost of insurance* literally refers to the rate of premium buyers need to pay.

Asset: it collocates with words referring to economy and finance, as in *optimal asset allocation*. This cluster has meanings concerned with the selection of the best way to allocate assets in an investment portfolio. It is found the 4-word clusters *capital asset pricing model* with the meaning refers to the name of a type of method in determining rate of return of an asset.

Among those lists, it was found that a smaller cluster is often a part of a bigger cluster (Sinclair, 1991). The present study has found a similar pattern as shown in Figure 1.
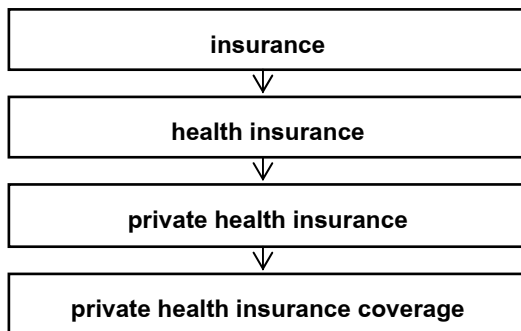
```
┌─────────────────────────────────────────┐
│              insurance                   │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│           health insurance               │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│        private health insurance          │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│    private health insurance coverage     │
└─────────────────────────────────────────┘
```

Figure 1: Hierarchy of lexical bundles with 'insurance' as the node

In Figure 1, the node *insurance* occurs first with the collocate *health* to form the collocation of the 2-word cluster *health insurance*. This cluster appears 651 times. The number of occurrences is lower when the combination is enlarged to the 3-word cluster *private health insurance*, which occurs 82 times, and the 4-word cluster *private health insurance coverage*, which appears 13 times. From this example, we see a clear semantic association between *"private"*, *"health"*, *"insurance"* and *"coverage"*. The connection requires knowledge in the insurance area to understand that this is one of many types of insurance products.

In a specialised field, abbreviation is also seen as a key element. In the present study, the node *bond* collocates with the abbreviation *CAT* as in *CAT bond*. To understand the meaning of this collocation, specific understanding of the field is required. It is significant for learners to know that *CAT* stands for "Catastrophe", so *CAT bond* actually refers to a kind of bond

that insurance companies issue to investors in order to raise funds in case a catastrophe occurred.

The Pedagogical Implications

To help English students acquire vocabulary effectively, the Lexical Approach (Lewis, 1997) has raised awareness of how important it is for learners to recognise and be able to use lexical structures in the forms of clusters or chunks of words. Having recognised the keywords in insurance as well as their collocational behaviours, the findings of this research allow for practical application in creating a collection of exercises focused on the use of collocation in the insurance field. Using authentic resources in the collocation exercises from target corpora is not exactly a new idea; prominent examples can be seen in works as English Collocations in Use (Advanced) (O' Dell and McCarthy, 2008) and Collocations Extra (Walter and Woodford, 2010). With respect to the need to learn vocabulary in the insurance field as suggested by Sangvaree (2004), this study has created a set of collocation exercises based on the presented outcome of the concordances shown by the Concordance tool AntConc (Anthony, 2007). Only the top three keywords "insurance", "risk", "insurer" and their collocates are selected to be included in the four different types of exercises: (1) Collocations Match-UP, (2) 3 – 4 Clusters Challenge, (3) Collocations Hunting, (4) Skill Practice.

1. Collocations Match-UP: Match the two parts of these collocations

1. primary   insurance
2. risk   aversion
3. health   insurer

2. 3 – 4 Clusters Challenge: Complete the collocations using the words below.

insurance   risk   insurer

1. private health ……………
2.………...management and insurance
3.…………and reinsurer

3. Collocations Hunting: Choose the correct collocation.

1. Over the course of multiple rounds, the potential insureds attempt to purchase …………… from the potential insurers.

   a. *insurance*   b. cost   c. loss   d. premium

2. Companies with very low risk exposures solve their………........ management issues without external help and rely on brokers only for transaction services.

   a. investment   b. internal   c. damage   d. *risk*

3. Stockholders supply capital to a stock ………….only when they expect to earn a fair return on their investment.

   a. money   b. firm   c. *insurer*   d. model

4. Skill Practice: Complete each sentence using the given words below

insurance   risk   insurer

1. Some people may lack private health…………coverage during a particular time period because they become frictionally uninsured. (insurance)

2. ……..aversion refers to the tendency to prefer to pay a defined sum of money that is known with certainty instead of being exposed to the risk of suffering a larger and uncertain financial loss in the future. (risk)

3. To back up their promises to pay claims under the terms of their policies, both the primary…………and the reinsurer hold equity capital to provide sufficient funds in case of adverse loss or investment fluctuations. (insurer)

The set of collocation exercises was ordered in degree of difficulty. Learners could start off easy by matching two-word collocations to raise awareness of their collocational patterns before moving on to a longer set of three- and four-word clusters, respectively. Exercises 3 and 4 are more difficult and challenging for learners, who will need to apply their cognitive skill in reading and inferring meaning from context. Vocabulary exercises are most fruitful for learners only when they are designed to be recyclable, and ordered in steps of difficulty which can help scaffold learners during their learning stages.

Conclusion

This present study was carried out with an aim to identify and analyse a list of most frequent keywords and their collocations. To achieve this, a corpus-based approach was employed. The IRAC was created to compare with the reference corpus, the 1 million written words of the BNC Sampler corpus. The electronic versions of published academic articles were collected, compiled and analysed by means of the software tool AntConc with the log-likelihood statistical test of lexical keyness used to detect keywords and their collocational patterns.

The list of the top 100 keywords and their collocations certainly has fundamental importance to the study of English in the field of insurance. It is worth introducing these frequently used lexical items to ESP students in the insurance field – and others who are involved in this industry – to learn, and for teachers to teach. It is recommended that the keywords and their collocational behaviours found in this study be incorporated into the vocabulary lessons of insurance as well as its teaching materials. Through suggested vocabulary exercises, learners can become more familiar with the lexical collocations of important keywords frequently used in the insurance field.

Limitations and Recommendations

The findings of this study reveal the vital role of implementing a corpus-based approach in vocabulary learning in which learners can gain access to authentic material and actual usage in context. Teaching concordances is yet another issue that is worth researching, but this is beyond the scope of this study. Future researchers could consider using a larger reference corpus or selecting more corpora which may be more effective in terms of finding frequency and keyness. While the present study has put more emphasis on providing the list of frequently used keywords and their collocations, other aspects of language features and usage have not fallen within the scope of this study. It is hoped, however, that the lexical findings in this study will prove valuable for ESP learners who study in the insurance field.

References

Anthony, L. 2007. **AntConc 3.2.1w (Windows)**. Available at: www.antlab.sci.

Chung, T.M. and Nation, P. 2004. "Identifying technical vocabulary." **System**, 32: 251-263.

Cortes, V. 2004. "Lexical bundles in published and student disciplinary writing: Examples from history and biology." **English for specific purposes**, 23: 397-423.

Coxhead, A. 2000. "A new academic wordlist." **TESOL Quarterly**, 34(2): 213-238.

Durrant, P. 2009. "Investigating the viability of a collocation list for students of English for academic purposes." **English for Specific Purposes**, 28: 157-169.

Gledhill, C. 2000. **Collocations in science writing.** Tübingen: Gunter Narr
          Verlag.

Jones, S. and Sinclair, J. M. 1974. "English lexical collocations. A study in
          computational linguistics." **Cahiers de Lexicologie**, 24: 15-61.

Lewis, M. 1997. **Implementing the Lexical Approach: Putting Theory Into
          Practice**. Hove: Language Teaching Publications.

Marco, M. J. L. 2000. "Collocational frameworks in medical research papers:
          A genre- based study." **English for Specific Purposes**, 19:
          63-86.

McCarthy, M. 1990. **Vocabulary**. Oxford: Oxford University Press.

Nation, I.S.P. 2001. **Learning Vocabulary in Another Language**.
          Cambridge: Cambridge University Press.

Nelson, M. A. 2000. **Corpus-Based study of business English and
          business English teaching materials**. (Electronic version).
          Unpublished PhD. Thesis. Manchester: University of Manchester.

O' Dell, F. and McCarthy, M. 2008. **English Collocations in Use
          (Advanced)**. Cambridge: Cambridge University Press.

Sangvaree, K. 2004. **A Study of the English Language Used in Insurance
          Business.** Master of Arts in English for Business and Technology,
          The University of the Thai Chamber of Commerce.

Scott, M. 1997. "PC analysis of key words – and key key words." **System**,
          25-2: 233-245.

Shin, D. and Nation, P. 2008. "Beyond single words: the most frequency
          collocations in spoken English." **ELT Journal**, 62(4): 339-348.

Sinclair, J. M. 1991. **Corpus, Concordance, Collocation**. Oxford: Oxford
          University Press.

Walter, E. and Woodford, K. 2010. **Collocations Extra. Multi-level Activities for Natural English**. Cambridge: Cambridge University Press.

Ward, J. 2007. "Collocation and technicality in EAP engineering." **Journal of English for Academic Purposes**, 6: 18-35.

West, M. 1953. **A general service list of English words**. London: Longman.