CHAPTER III

RESEARCH METHODOLOGY

This chapter aims to discuss general information of data and research methodology used in this study. The first section describes the process of corpora building. The second section mentions the steps of the data analysis starting from the linguistic analysis to the quantitative analysis and to the qualitative analysis respectively. In the final section, a summary of all important points is provided.

3.1 Building of corpora

3.1.1 Selection of corpus

In this study, feature articles in English language magazines published in Thailand and in Britain are collected as representatives of Thai and British English. Since the focus of this study is on describing the point that Thais very much prefer to provide lots of information, the magazine genre seemed to be appropriate. In comparison with other written genres, in which standard English is commonly used and they are available in Thailand such as newspaper and academic prose, magazine feature articles seem to provide more space and freedom (in terms of language and style) to the writers. Therefore, the expecting characteristics of Thai English can be highly assumed. Also, the language used in magazine articles tends to be more natural and relaxing than those in newspapers or academic papers – where fixed styles, expressions, vocabulary and so on are commonly required.

Apart from magazines produced by Thai publishers, almost all English language magazines sold in Thailand are from two major native speaking countries: the USA and the UK. In this comparative study, magazines published in Britain are selected as representatives of British English variety. British English was used in this comparative study because firstly, it is the origin of the English language. It has been developed alongside with its original culture by the majority of its native people. In contrast, with a metaphor comparing USA as a salad bowl – people from everywhere of any races and cultural backgrounds mix together; American English might be less homogeneous. Since relationship between language and culture was a point of

concern of this study, British English was probably more suitable. Additionally, according to the mixing of people in USA, it seems to be less reasonable to assume the nationality or the language community of writers. Furthermore, when looking at the literature, comparative studies between Thai and British English are rare. As in those studies reviewed in Chapter II, all of them compare Thai English with American English. Hence, findings from this study can widen the proof for distinction of Thai English from another native variety, namely British English.

3.1.2 Selection of magazines

The selection of magazines and types of columns is based on English magazines published in Thailand. Then comparable columns from comparable types of magazines published in Britain are selected accordingly.

For the selection of Thai English magazines, first, the magazines should contain articles that are only written in English, not the Thai language articles with their English translated versions. Contents of the articles should be general so that any readers can read and understand them. In other words, they should not be those articles requiring some particular knowledge of some specific fields like computers, cars, technology and so on. Also the English language used in the magazines should be general English that is meaningful for general readers. The articles should not contain too much jargon or slang. Moreover, they should not be written with simplified English for student readers. According to these criteria, Tropical Art Culture & Travel and Metro: are two magazines selected as representatives of Thai English magazine expository prose.

Tropical Art Culture & Travel magazine is published by a publishing company in Thailand. The focus of the magazine is on travel, art, and culture of Thailand and neighboring countries. Target readers of the magazine are executive and managerial level people with various occupations, who are interested in travel, art, and culture. The readership covers both Thais and foreigners living in Thailand and abroad. It is a monthly magazine with a circulation of 40,000 copies a month (Tropic of Cancer Co., Ltd., n.d.).

Metro: magazine is also published by a publishing company in Thailand. It is a variety magazine that focuses on the trendy city life of Bangkok. It claims itself as "an upmarket product, appealing to an upmarket audience" and Bangkok's best-

selling English language magazine. Thus, its target readers are wealthy well-educated Thais and foreigners living in Thailand, especially in Bangkok. It is a monthly magazine with a circulation of 35,000 copies a month. (City Media Group Co., Ltd., n.d.)

In comparison with Thai English magazines, Conde' Nast Traveller magazine and HELLO!, the two magazines published in Britain, are chosen as representatives of British English.

Conde' Nast Traveller magazine, under the same name, is published in many countries and in more than one language. For the content in the magazine from different places of publishing, it seems that the publishers work separately. They only share the theme of the magazine which is on travel, adventure, culture and new ideas. So as to be a representative of British English magazine, Conde' Nast Traveller magazine in English version published in Britain is particularly selected for this study. Contents of articles in this magazine are reasonably comparable to those in Tropical Art Culture & Travel. They introduce traveling destinations in several countries. In addition, the UK Conde' Nast Traveller is a monthly magazine with a circulation up to 80,000 copies a month. Its readers include not only people in Britain but also people in other countries (CondéNet UK Ltd., n.d.). Its high circulation and wide range of readership can confirm that the magazine can be a good representative of standard British English.

HELLO! magazine is a variety magazine chosen as a representative of British English. The content of the magazine covers certain topics ranging from well-known people, fashion, health and beauty, and travel. 75 per cent of the content is about people and the other 25 per cent is about general interest. HELLO! magazine is published in many languages by different groups of publishers. The magazines from different publishers share the same theme but the inside contents depend on the publication house. HELLO! in English is sold worldwide, including Thailand. From the year 2006, there is also a Thai version of HELLO! available in Thailand. Unlike other English language magazines available in Thailand which focus mainly on fashion, models, and gossips, HELLO! magazine presents various view points of people's lives and perspectives in the form of long feature articles. Target readers of the magazine are educated and working people who possess high purchasing power. The majority of the UK readers are those upmarket audiences. It is a weekly

magazine with a circulation of 412,807 per week that makes an average of 1.8 million copies each month (HELLO Ltd., n.d.). According to information about the content and target readers, articles from this magazine are comparable to those taken from Metro.

3.1.3 Selection of articles

The selection of articles was done firstly from the Thai English magazines focusing on two types of column: places and people. The researcher scanned through all articles in the magazines and bookmarked all possible articles to be chosen for the analyses. Those bookmarked articles were checked for the name of the writers in order to get the initial knowledge that the writers were Thai or at least they have some relation to Thai family, language, and culture. Next, the potential articles were read thoroughly in order to know the contents and to make sure that the contents were appropriate: the articles should be written as expository features for information presenting purpose, not for commercial or advertising purposes. Moreover, for people articles in Thai magazines; although a writer is Thai, some articles – which are full of direct quotations of a non-Thai interviewee answering short questions of a Thai interviewer or writer – were not selected. The length of the article was also a point of concern. To get enough information, the articles should not be too short. The length of the articles should range approximately from 800 - 2000 words for articles about people and 1000 - 3000 words for articles about places.

To obtain more information confirming the proper selection of Thai English representatives, questionnaires were sent to the writers and the editors. For the writers' questionnaire, the answers concluded that the Thai writers are Thai and they are competent users of English. All of them have some experience, at least 3 years, being in English native speaking countries. However, when asking them to rate their English, most of them mentioned that they can use English grammatically correct but still they feel that their English is rather different from the native speakers. Since the Thai writers are, at the time, living in Thailand, all of them answered that Thai is a major language for communication in their everyday life. Concerning the editorial process, a questionnaire was also sent to editors. In general, editors are both Thai and English native speakers. All of the editors answered that Thai writers have high proficiency in English and their English is near-native. They considered that

Thai writers' exposure to English language communities is very important. In terms of editing articles, all editors agreed that they primarily focus on comprehension and correct grammar. According to the information gathered from the Thai writers and the magazine editors, it could ensure that the Thai English articles could be proper representatives of Thai English.

After getting Thai English articles, British English articles were then selected. Similar to the selection process of Thai articles, articles in the two magazines were scanned and any possibly comparable articles to those of Thai articles were bookmarked. To check whether they were really comparable, the articles are read thoroughly and only the suitable articles were chosen for the analysis. Generally, most of the British articles were longer than the Thai articles in terms of number of words. However, in the selection of British English articles, the length of the texts was controlled at around 800-2000 words for articles about people and 1000 - 3000 words for articles about places..

Findings from the pilot study indicated that, in average, an article with an approximate length of 400 words contained about 25 prominent head nouns and 19 sentences. To make sure that the main study would have enough data, these numbers were considered. Therefore, with an estimation based on the shortest length of articles (1000 words); if one aimed to get a sample size of at least 1,000 prominent head nouns and 1,000 sentences for each groups of writers, a corpus of at least 25 articles should be a good number. According to the calculation, 25 articles could provide around 1,562 prominent head nouns and 1,187 sentences.

As a result, 28 articles from Thai English magazines and 28 articles from British English magazines were collected. The table below describes general information of sample size of the two corpora.

Table 3.1 Sample size in Thai and British corpora

Magazine articles	Places	People	Total	Head nouns	Sentences
Thai	15	13	28	2027	2150
British	15	13	28	2632	2581

More elaborated information of each article is provided in Appendix C.

3.2 Data analysis

This section presents the process of conducting this research. It covers data preparation, research methods and analysis, and descriptions of research instruments used in each particular step.

3.2.1 Linguistic analysis

3.2.1.1 Compilation of corpora

All of the selected articles were converted into machine-readable format: namely Thai English articles (TEA) and British English articles (BEA), for Thai and British English corpora respectively. Titles, pictures, writer's names and other notes (e.g. pictures' description) were taken off. Only the real content of articles was kept for the analysis. Then, all of the sentences in the articles were separated from each other by adding a blank line between them. This method could help eliminate confusion and mistakes in the analysis. Also, it facilitates the counting of sentences.

3.2.1.2 Identification of types of sentence construction and noun modifiers

Types of sentence construction

In this step, each sentence was analysed in isolation. All clauses in the sentence were considered, based on Quirk et al. (1972). Then, the type of the sentence (simple, compound, complex, or compound-complex sentence) could be classified and then coded. In short, a simple sentence consists of one independent clause while a compound sentence contains two or more. A complex sentence comprises an independent clause that embeds other dependent clause(s). Last, a compound-complex sentence occurs as a structure combining a complex sentence with others (e.g. complex + simple, complex + compound, or complex + complex). In other words, a compound-complex sentence is the one that consists of two or more independent clauses with at least one dependent clause. (More elaborated descriptions can be seen in Chapter 2, pp. 41-44) Incomplete sentences such as short answers, run-on expressions, and so on, were also coded. However, they were marked that they were incomplete. Those yes/no answers were abandoned in the analysis. A code

marking type of the sentence was placed at the beginning of each sentence. A list of codes used for types of sentence construction is presented in the following table.

Table 3.2   Codes for types of sentence constructions

Codes	Types of sentence constructions
[Sim]	Simple sentence
[Cp]	Compound sentence
[Cx]	Complex sentence
[Cc]	Compound-complex sentence
[i]	Incomplete sentence

For example

[Sim] Khom civilization virtually disappeared.

[Cp] Most of the time you can go up and take a picture, but he doesn't really like it and every once in a while he takes a piece out of someone.'

[Cx] When Buddhist kings reined, Buddhism was supported by the establishment of temples all over the region.

[CC] [i] Amble along the ridge taking in views of the south China Sea, then descend to the village of shek O, where those who have remembered to pack their kit can enjoy a refreshing swim.

When a sentence holds a direct quotation, the sentence analysis could be done as shown in the following example.

Extract:

"I like to be here at different times of the day to be with them. I don't want to be one of those fathers who is just there in the evenings and weekends," he says.

The above extract could be separated into three sentences which were;

1. [Sim] He says "_____".

2. [Cx] I like to be here at different times of the day to be with them.

3. [Cx] I don't want to be one of those fathers who is just there in the evenings and weekends.

Noun modifiers

After a sentence type was identified, prominent head noun(s) and noun modifier(s) of that sentence were then examined. The process could be described as in the following steps.

Step 1: *Identification of prominent head noun*

To identify the most prominent head noun(s) of a sentence, the concept of theme and rheme was considered. According to the Prague school, "Theme is what the sentence is about and rheme is what is said about it" (Connor, 1996 p. 81). In other words, theme can be said as an element functioning as a subject of the sentence while rheme as a part adding further information to theme – the predicate. In general, a head noun in theme was considered as most prominent since it is a topic of the sentence. However, in cases that the head noun in theme was not a good source for the analysis of noun modifiers, a head noun in rheme was considered instead. This includes the following cases.

1. *The subject or theme of a sentence was a pronoun.* Since pronouns do not take any modifiers, they may not really provide any useful information to this study.

2. *The sentence was a cleft.* In English, cleft or pseudo cleft sentences are used when the focus of the sentence is shifted from the subject to the subjective complement. Therefore, the head noun in rheme (the subjective complement) should be more prominent in this case. So, it is more likely for writers to add modifiers to the head noun in rheme than that in theme. Also, in cleft sentences, a pronoun '*it*' is normally used as a subject. Hence, in this case, it is obvious that the theme does not contribute useful information to the study.

For example

Cleft sentence: It was Thomas Edison who invented the electric lamp.

3. *The subject of a sentence was a nominal clause.* The nominal clause can be both finite and non-finite. When it is placed in a subject position, the whole clause behaves as a nominal subject. In this case, the noun in theme is not really prominent since it is only a part of the nominal clause.

> *For example*
>
> **That-clause**: *That she is still alive* is a consolation.
> **Nominal relative clause**: *What he is looking for* is a wife.
> **To infinitive clause**: *To go to Japan* is a good idea.
> **ing-clause**: *Eating people* is wrong.

It should be noted that there is a special class of verbal noun in the English language: the gerund. Sometimes ing-clauses and gerunds look similar (for the distinction of the two, see literature review). When the -ing construction occurred, the identification would be done carefully in order to justify whether it was a gerund or ing-clause. Since a gerund can have the same manners as general nouns (especially it can take modifiers); when that -ing construction is proven to be a gerund, it should be identified as a prominent head noun. This means the identification does not have to shift to rheme.

When the identification of the prominent head noun went to rheme, the following criteria were applied.

1. The prominent head noun can be a direct object or subjective complement of the sentence.

> *For example*
>
> He wears spurs, a leather jerkin and tinted glasses.
>
> He is a very nice man.
>
> When the complement is an adjective phrase, a noun in the adjective complement is taken.
>
> *For example*
>
> He is afraid of ghosts.

But the noun in an adjunct would never be considered, even if there was no other head noun available.

For example

In that spooky *hotel*, he was so scared.

The noun '*hotel*' is not considered as a prominent head noun.

2. When the verb is ditransitive, direct object is firstly considered to be the most prominent head noun. In cases that the direct object is not available (i.e. a pronoun), the indirect object is then considered.

For example

I gave it to that wonderful <u>lady</u>.

3. The objective pronoun or the objective nominal clause would not be identified as a prominent head noun.

For example

You can find *them*.

We knew *there was trouble coming.*

From the above criteria, a hierarchy of head noun identification can be concluded. First, subject of the sentence is considered. Then, if it does not serve the requirement, the identification shifts to direct object or complement. If direct object and complement are not available, the identification goes to indirect object.

According to all these criteria, it was possible for a sentence to have more than one prominent head noun. This was, for example, when there were more than one noun acting as a subject of the sentence or when the sentence consisted of more than one independent clause as in a compound sentence. On the other hand, it was possible for a sentence to contain no prominent head noun if there was no noun neither in theme nor rheme matching the criteria.

Step 2: *Identification of modifiers*

In this step, modifiers surrounding each prominent head noun were identified and classified according to their positions and types, based on a framework of Quirk et al. (1972). For positions, noun modifiers were divided into premodifiers and postmodifiers. Then, types of modifiers were identified. A code marking type of each noun modifier was placed in front of it. Followings are some examples.

[art] A [ADJ] small <u>walkway</u> [PP] next to the Phu Soi Dao waterfall is the starting point for the trek through the shady forest.

They have [det] their [det] own <u>rooms</u> [PP] in her house.

The table below lists the codes used for all types of noun modifiers.

Table 3.3   Codes for noun modifiers

Premodifiers	
Code	**Meaning**
[predet]	predeterminer
[art]	article
[det]	determiner
[ord]	ordinal
[car]	cardinal
[quan]	quantifier
[ADJ]	adjective
[N]	noun
[S]	's-genitive
[OTH]	other minor types

Postmodifiers	
Code	**Meaning**
[OF]	general of-phrase
[genOF]	genitive of-phrase
[PP]	other prepositional phrase
[REL]	relative clause with relative pronoun
[RELw/o]	relative clause without relative pronoun
[ED]	past participle clause
[ING]	present participle clause
[TO]	infinitive clause

The codes for postmodifiers can be added with 'nr' and 'app' to mark that the postmodifier is non-restrictive and appositive respectively, e.g. [PPnr] is for a non-restrictive prepositional phrase and [RELapp] is for an appositive relative clause.

3.2.1.3 Reliability check for the identification

Reliability checks for the identification of prominent head noun, noun modifiers, and types of sentence construction were done in order to confirm that the researcher's coding was consistent and unbiased. Thus, an external inter-rater was invited. The inter-rater, during the period of this study, was a Ph.D. student in English applied linguistics and had taken some classes in basic English linguistics. Also, he had quite a long-time experience in teaching English at the university level. Therefore, it could be claimed that he was knowledgeable about the English language. So, he was adequately qualified for being an inter-rater of this study.

The inter-rater was asked to code six articles (10% of the data) which were randomly selected from both Thai English and British English articles. Before doing so, criteria for the identifications and a sample of a coded article, which was a different one from those six articles, were given to him. He was asked to study them thoroughly. After that, a short training was arranged. The researcher verbally explained all criteria to him and checked for his understanding. Also, the inter-rater was allowed to discuss his comments or suggestions and to ask for clarity of some particular points. Next, still with the researcher, the inter-rater was asked to code his first article. That was to see if the inter-rater had any problem with the real coding process.

After the inter-rater finished his six articles, the coded articles of the inter-rater and the researcher were compared. The following tables show details on the coding done by the inter-rater and the researcher.

Table 3.4   Inter-rater's and researcher's coding

Samples	Prominent head noun		Total agreement	r	Noun Modifiers		Total agreement	r	Sentence constructions		Total agreement	r
	I	R			I	R			I	R		
1	64	57	50	88%	91	85	75	88%	68	70	68	97%
2	61	65	61	94%	58	69	58	84%	51	51	50	98%
3	98	101	98	97%	146	166	146	88%	106	106	106	100%
4	66	66	66	100%	67	65	65	100%	55	54	48	89%
5	76	74	72	97%	119	121	116	96%	51	51	47	92%
6	43	39	35	90%	84	82	80	98%	43	43	43	100%
Total	408	402	382	95%	565	588	540	92%	374	375	362	97%

I = coding done by interrater   R = coding done by researcher   r = reliability

In order to confirm that the researcher's identification of prominent head nouns, types of noun modifiers, and types of sentence constructions could be trusted, reliability of the coding was tested. To do so, the following formula is applied.

$$\textit{Coefficient of reliability} \quad = \quad \frac{\text{number of total agreement of all judges}}{\text{total number of tokens coded by researcher}}$$

(Scholfield, 1997, in Singhasiri, 2001 pp. 164-165)

With this formula, reliability coefficiency of the coding could be calculated. Then, the degrees of reliability for the coding of the three categories (prominent head nouns, noun modifiers, and types of sentences) could be determined. The results from the calculation are shown as follows.

Coefficient of reliability for prominent head noun = 382/402 = 0.95

This means the coding of prominent head nouns is 95% reliable.

Coefficient of reliability for noun modifiers = 540/588 = 0.92

This means the coding of noun modifiers is 92% reliable.

Coefficient of reliability for sentence constructions = 362/375 = 0.97

This means the coding of noun modifiers is 97% reliable.

According to the degrees of reliability for coding between the researcher and the inter-rater, ranging from 100% to 84% as shown in Table 3.4 above, it can be firmly claimed that the linguistic analysisi of this study are trustworthy.

3.2.2 Quantitative analysis

After the identification process had been complete, the quantitative analysis was done in the following steps.

Step 1: *Frequency count*

Within this step, a frequency count of all tokens appeared in Thai and British English corpus was done. Therefore, a total number of

1. prominent head nouns,
2. each type of noun pre- and postmodifiers, and

3. each type of sentence construction

were obtained.

Step 2: *Comparison between Thai and British English corpus*

In this step, every total frequency of the same type of token was compared across the two corpora. This was to test whether each of them was used significantly different, in terms of frequency of occurrence, by Thai and British writers. To do so, chi-square test was applied.

Chi square statistics

Chi square is a non-parametric test of statistical significance. It can be used to estimate the difference of the variables. Unlike those parametric tests e.g. t-test and analysis of variance, by the terms 'non-parametric', chi square accepts weaker, less accurate data as input. That is, things like normality (the bell-shape distribution), homogeneity of variance, and so on are not expected. Therefore, with its strength, chi square can be used in a wide variety of research contexts.

Chi Square Requirements

1. The sample must be randomly drawn from the population.
2. Data must be reported in raw frequencies (not percentages).
3. Measured variables must be independent.
4. Values/categories on independent and dependent variables must be mutually exclusive and exhaustive.
5. Observed frequencies cannot be too small.

(Connor-Linton, 2003)

To calculate chi square value (χ^2), observed frequency (O) and expected frequency (E) are needed.

Chi square value can be calculated by using the following formula,

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \dots + \frac{(O_n-E_n)^2}{E_n}$$

The observed frequency is obtained from real observations while the expected frequency can be a theoretical value (if available) or a calculated value of the following formula.

$$E = \frac{(O) \times (\text{number of sample})}{(\text{total number of sample})}$$

Chi square calculation for noun modifiers and interpretation

The comparison of each type of noun modifiers between TEA and BEA can be done separately. This is because there is no relation among the occurrences of all types of noun modifiers. In other words, the occurrence of one type cannot block the occurrence of any other types. This kind of data is called one-way classification.

Key (1997) mentions that for one way classification of chi square test (or sometimes referred to as the Single Sample Chi Square Test), the calculation needs to be corrected with Yate's correction. To do so, a value of 0.5 is subtracted from each absolute value of the difference between observed frequency and expected frequency. Thus, the formula for computing chi square value of a single set of data becomes:

$$\chi^2 = \sum \frac{(|O - E| - 0.5)^2}{E}$$

The following is an example of chi square calculation comparing the use of relative clauses as postmodifiers in Thai and British English corpus.

Table 3.5   Sample of data for chi-square calculation

	Relative clause (O)	Total Head noun
Thai	170	2027
British	154	2632
Total	324	4659

First, the expected values for the use of relative clause by Thai writers (E_t) and by British writers (E_b) are needed.

So, $E_t = (324 \times 2027)/4659 = 140.96$

$E_b = (324 \times 2632)/4659 = 183.04$

Then, all values can be substituted to the chi square formula.

$$\chi^2 = \frac{(|O_t-E_t| - 0.5)^2}{E_t} + \frac{(|O_b-E_b| -0.5)^2}{E_b}$$

$$= \frac{(|170-140.96| - 0.5)^2}{140.96} + \frac{(|154-183.04| - 0.5)^2}{183.04}$$

$$= 5.78 + 4.45 = 10.23$$

To interpret the chi square value from the chi square table, the degree of freedom (df) and a probability value (P) are needed. For the data of one-way classification, the degree of freedom is equal to a number of values/categories on the variable minus one. Therefore, in this case df is, $(2 - 1)$, one. Probability value means the chance that the estimation may have some errors. For example if $P = 0.05$, it can be said that it is 5 per cent chance for the interpretation to be wrong. In other words, we allow 95 per cent of confidence for the interpretation. Statistically, the probability value is accepted at $P \leq 0.05$. It should be noted that when less probability value is set, the more accuracy in the interpretation is obtained.

The critical chi square value when $df = 1$ and at $P \leq 0.05$ (from the table) is 3.84. Since the calculated chi square value from the data (10.23) is larger than the critical value (3.84), it can be interpreted that we can be 95% sure that the use of the relative clause by Thai and British writers is significantly different. Moreover, the observed and expected frequencies of relative clauses in Thai and British corpora indicate that Thai writers use higher number of relative clauses than it is expected (O = 170 and E = 141), while British writers do it differently (O = 154 and E = 183). This information can be interpreted that Thai writers use relative clauses significantly more often than British writers do.

Chi square calculation for sentence constructions and interpretation

Unlike the occurrence of noun modifiers, the occurrence of each type of sentences is restricted. One sentence cannot be classified in more than one type of construction. Thus, this in-group relation plays a role in the comparison. This kind of data is a two-way classification. For the two-way classification, chi square

calculations have to be done as a whole. This means each category (each type of sentences) cannot be separately compared across the two corpora.

Thus, a comparison for the use of sentences across the two corpora has to be done with the whole set of data. As a result, only one chi square value can be obtained.

The chi square value can be computed with the equation:

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \ldots + \frac{(O_n-E_n)^2}{E_n}$$

Thus, the calculation of the chi square test for types of sentence constructions can be done as follows.

Table 3.6 Data for chi-square calculation of sentence constructions

Sentence Types	TEA (O)	BEA (O)	Total
Simple	546	823	1369
Compound	137	274	411
Complex	1237	1094	2331
Compound-complex	230	390	620
Total	2150	2581	4731

First, all expected values for each cell are calculated with a formula:

$$E_i = \frac{\sum (O) \times (number\ of\ sample)_i}{(total\ number\ of\ sample)}$$

Therefore, expected frequency of simple sentence in TEA $= \frac{1369 \times 2150}{4731}$

$$= 622.14$$

expected frequency of simple sentence in BEA $= \frac{1369 \times 2581}{4731}$

$$= 746.86$$

The same method of calculation is repeated in order to obtain all expected frequencies presented in the following table.

Table 3.7   Expected frequencies for types of sentences in TEA and BEA

Sentence Types	TEA (E)	BEA (E)
Simple	622.14	746.86
Compound	186.78	224.22
Complex	1059.32	1271.68
Compound-complex	281.76	338.24

All of the values are substituted to the chi square equation mentioned above. Then, the chi square can be obtained. And for this test, it is 133.45.

Next, the degree of freedom should be indicated. For two-way test,

df = (numbers of category – 1) x (numbers of comparison group – 1)

Thus, for this case, df = (4-1) x (2-1) = 3

Again, the probability value is set at $P \leq 0.05$. From the chi square table; the critical value of chi square, where df = 3 and $P \leq 0.05$, is 7.82. Therefore, the chi square value obtained from this test indicates a very high degree of significant difference in the use of sentences between the two groups of writers.

3.2.3 Qualitative analysis

After the significant types of noun modifiers and sentence constructions were identified from the quantitative analysis (the chi square test), the qualitative analysis was further conducted in order to learn more about their use and construction in TEA and BEA. The process of the analysis can be described as follows.

3.2.3.1 Noun modifiers

For further study of noun modifiers, a concordancing programme, KWIC, was used. With this programme, concordance lines of a particular type of noun modifiers could be executed. Followings are illustrations and explanations of how concordancing lines for genitive of-phrase, as an example, could be compiled.

First, a key word for the search should be put in a box provided in the programme. As in this figure, a code for genitive of-phrase was input.
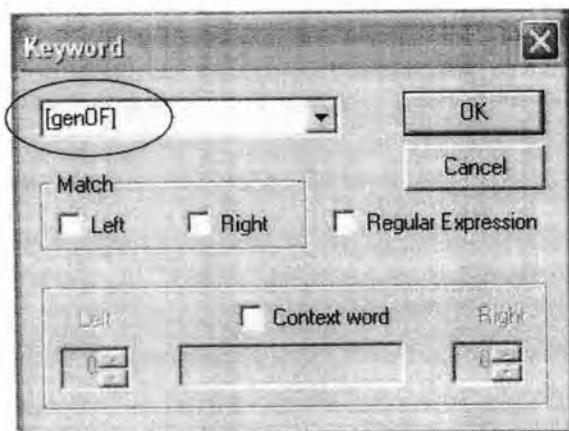
Figure 3.1 KWIC programme for concordance lines

After pressing 'OK' button, the programme then displayed all concordance lines gathered from each corpus with the searched word in the middle.
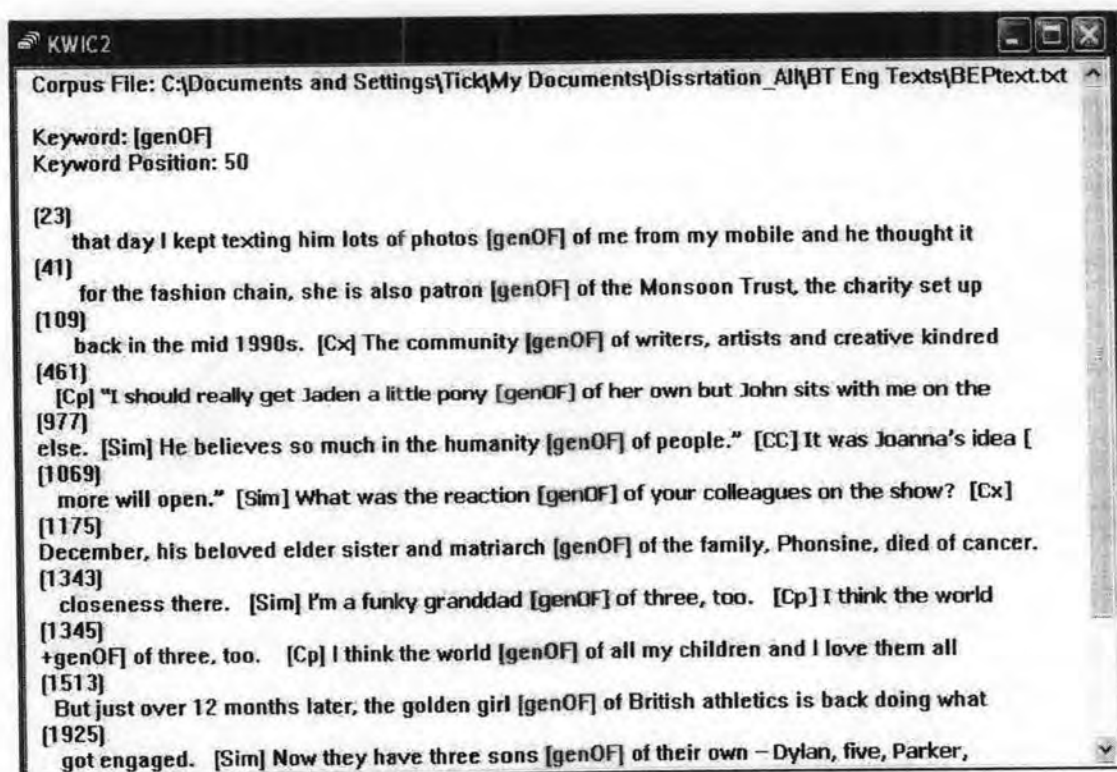

Figure 3.2 Concordance lines obtained from KWIC

Then, the qualitative analysis in terms of the use and construction of genitive of-phrases across the corpora could be done. However, sometimes, the information provided in concordance lines was not enough (e.g. those for long relative

clause). So, the researcher had to go back to the text file and search for more information.

The same process of the qualitative analysis as shown above is repeated for all other noun modifiers which are used significantly different in TEA and BEA.

3.2.3.2 Sentence constructions

For the qualitative analysis of sentence constructions, the sentences of the same type were grouped together. To do this, Microsoft excel was used. With this application, full information of sentences could be retained so it was rather appropriate for this analysis than the concordancing programme. Since the sentences in the text files had already been separated; when inputting them to the spreadsheet, each line nicely contained one sentence. With the codes placed in front, grouping of sentences in accordance with their types could be easily done by using the sort ascending tool.
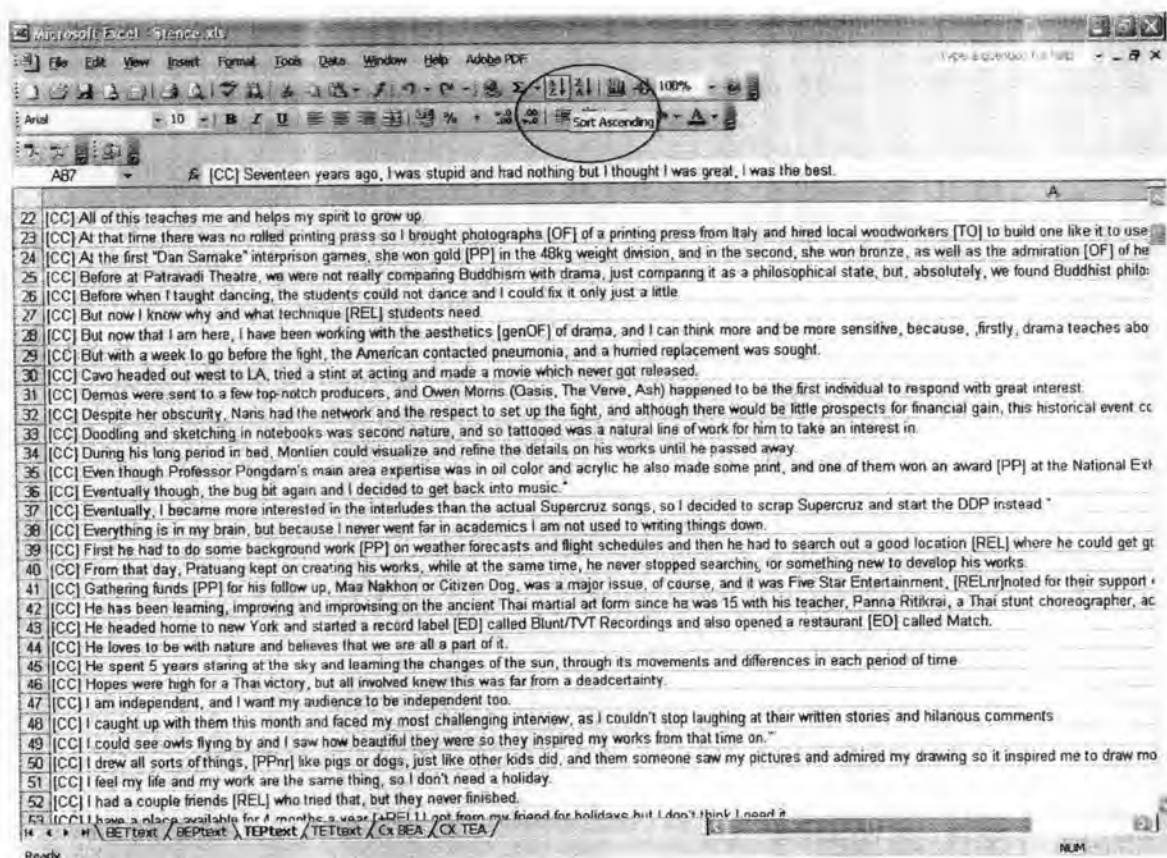

Figure 3.3 Analysis of sentence construction

Next, each type of sentence in Thai and British English corpus was further studied in terms of their use and construction. In addition, when more contextual information of a particular sentence was required, the researcher went back to the original text and used the search tool to help capture the sentence.

3.3 Conclusion

In this chapter, research methodology has been reported. The first section describes the process of data collection and information of sample size. In the second section, elaborated explanations for the data analysis are provided. The section firstly discusses the method for compilation of the corpora. Then, details on the linguistic element identification and coding together with reliability checking process are given. After that, steps in the quantitative analysis and the instrument used (chi square test) are presented. Last, methods for the qualitative analysis and the instruments (computer programmes) are explained.