

Journal of Southeast Asian Language Teaching



A refereed journal of the Council of Teachers of Southeast Asian Languages

ISSN 1932 3611

Journal of Southeast Asian Language Teaching, Volume 14, 2008

Thai Pop Music: Corpus Analysis and Second Language Learning

Teddy Bofman (T-Bofman@neiu.edu)

and

Paul Prez (P-Prez@neiu.edu)

Northeastern Illinois University

Abstract

This paper examines a corpus of 400 popular songs downloaded from www.ethaimusic.com on the Internet. The techniques of corpus linguistics were used to analyze this body of music in an effort to better prepare language teachers to use these songs in the classroom.

Corpus linguistics is the study of language using large bodies of text. While linguists have always been interested in data, that interest waned considerably during the heyday of Noam Chomsky. The modern computer has refocused our attention and allowed us to examine large bodies of authentic data.

The analysis of pop songs that we present allows us to focus on key linguistic features found in this music. Using Oxford WordSmith Tools 4.0, we have looked at word frequencies and concordances in order to get an overall picture of the language found in these songs. We examine features such as pronoun usage, references to emotions and body parts, elaborate expressions, slang spellings, Sanskrit vocabulary and English terminology found in these texts.

Our findings provide guidance to language teachers in a number of ways. For example, they point out which vocabulary is high frequency and therefore essential to teach. They also indicate major collocational usage, essential knowledge for the language learner. The lyrics we analyzed allow students to engage in data-driven learning. These same tools applied to other genres could have tremendous impact on the way Thai and other languages are taught in the United States.

This study examines a corpus of four hundred popular Thai songs found on the Internet at www.ethaimusic.com. The techniques of corpus linguistics were used to examine this body of music in an effort to help language teachers better prepare to teach this material. We believe that studies such as this one could be applied to various types of discourse such as newspaper articles,

academic writing, business reports, news broadcasts, and sitcoms, as a means of helping language instructors gauge appropriate levels of material and also as a means of helping teachers choose what to target before, during, and after actual lessons.

Our discussion covers several areas, including a brief description of corpus linguistics, the tools we have used in our corpus study, the kinds of information we have learned about the corpus we examined, and finally, suggestions for working with this corpus, given the information we have gleaned from our research.

What is corpus linguistics? It is the study of language using large corpora or bodies of text, whether written, spoken, or signed (Hunston 2002). Linguists have been using corpus techniques for decades. Bofman's doctoral dissertation, for example, completed at the University of Michigan in 1978, was a corpus study of a sixty-page segment of the *Ramakian*, the national epic of Thailand. In this study, she examined pragmatics and word usage without the aid of a computer. A portion of that study has been made available on www.seasite.niu.edu/thai. During the heyday of Noam Chomsky interest in language data waned to some extent, with the focus of study based on the intuitions of native speakers, rather than looking at bodies of naturally-occurring data.

Advances in computer technology have reignited an interest in corpus studies, which allow researchers to look at very large bodies of data in an efficient manner. Corpus studies value the use of authentic texts. The songs we have looked at, popular Thai music found on the Internet, are just that. Authentic texts, valuable to all teachers, are of particular importance to teachers who are non-native speakers and may have doubts about their intuitions.

By using corpus techniques to examine a body of texts, teachers are better able to judge whether or not the material is at the appropriate level for the class being taught. Corpus techniques also enable teachers to determine what aspects of the language need to be targeted in actual teaching. Corpus linguistics is one of many tools available to teachers. It is not meant to replace existing techniques, but to complement them.

Corpus linguists investigate computer-readable corpora with the aid of analytical software. Ideally, such software provides the types of analyses that will highlight and profile the features and characteristics of a corpus that are of interest to the individual researcher. For the present study, we used Oxford WordSmith Tools 4.0, published by Oxford University Press. This software package comprises a suite of analytical tools each of which aids in building an overall picture of a corpus. The WordSmith tools that were of most interest to us were WordList and Concord. WordList generates word frequency lists in ascending or descending order as well as in alphabetical order. This same tool also provides statistical data regarding range of use of each lexical item. This information is useful because a word that occurs thirty times in a single song is of less pedagogical significance than a word that occurs thirty times but in a number of songs throughout the corpus. The WordSmith tool, Concord, provides lexis-in-context data, i.e., concordances. Making extensive use of these two functions and the various options included in each, we gained insight into the language of Thai pop music that would be difficult to obtain relying on intuition alone.

Why did we choose this particular corpus to study? The most obvious reason, perhaps, was its accessibility. We had access to hundreds of songs, amounting to roughly 103,000 running words. The corpus provides authentic material using modern colloquial Thai. It is not material written with the goal of teaching in mind. Rather, it is material that stems from the culture itself. Of equal importance, perhaps, is the fact that the topic is inherently of interest to college-age students who enroll in language classes.

What we found in this body of songs is abundant linguistic creativity, a feature that is the hallmark of classical Thai literature and obviously of modern Thai as well. Additionally, the songs are a rich source of cultural information. They make reference to modern life in Thailand, providing an excellent starting point for discussion and explanation of current life in the country. To the best of our knowledge, there is still no frequency list of vocabulary for Thai comparable to the General Service List (GSL) that exists for English (West 1953). These lists provide a rank order of vocabulary in terms of usefulness and frequency of usage. These high frequency vocabulary items are the basic building blocks of language and should be a major focus of language pedagogy. It is our hope that as other genres of Thai are analyzed using corpus techniques, a general service list of Thai will emerge. Our study is perhaps a first step in that direction.

Our vision is that studies such as ours need to be applied to texts representative of various domains in a given language. Corpus studies of Buddhist sermons, classical poetry, newspaper articles, and other types of texts would provide language teachers with information about various genres that could be used as authentic material for the classroom. By sharing our information about these corpora, teachers of Thai, Indonesian, Tagalog, and other Southeast Asian languages would have pedagogical knowledge of a rich and varied supply of authentic material at their disposal.

The website www.ethaimusic.com allows the viewer to see the song written in Thai, transliterated into English, and translated by non-native speakers of English into English. One can hear the song sung and can also consult an online glossary of the vocabulary occurring in the lyrics. These features alone make the music highly accessible to students, whether for classroom study or independent work.

The Thai language is written without spaces between words. In order to be processed digitally, the text has to be segmented into words. This fact provides a tremendous challenge to those who do corpus-based studies of Thai. We went through each song and manually inserted spaces as word boundaries. Decisions were made in a variety of ways, using our knowledge of the language as a starting point. The online glossary provided by the website gave us additional clues. Dictionaries were consulted as well as native speakers. We found that native speakers did not always agree as to what constitutes a word in Thai. In fact, from one day to the next, a native speaker might have a different opinion as to whether something was or was not a single word in Thai. This process of segmentation was laborious and time-consuming and likely to be the source of inconsistencies. However, this particular study does not depend on having perfect counts; instead, we look at trends and therefore feel confident that our data is acceptable. It is to be hoped that as corpus-based work on Southeast Asian languages enters into its own, this particular problem of segmentation will be resolved by the community of Thai scholars.

A positive aspect of the segmentation process is that it allows the researcher to gain initial familiarity with the corpus. It is through this first reading that the researcher begins to make mental notes of the nature of the text. What kinds of vocabulary are most frequent? What are the basic grammatical features used in these songs? What themes are found within the corpus? These questions formed the basis of our analysis.

Throughout the segmentation process, the researchers made notes of trends in the text. In this particular corpus we noted, for example, that, not surprisingly, pronoun usage was largely informal, and we planned to track this usage. We saw that the mood within the songs was largely melancholy; many words describing sadness, confusion, and loneliness were used. While we noted a relatively limited mention of parts of the body, we did note numerous references to nature. We saw abundant usage of doublets, reduplication, internal rhyming, alliteration, and elaborate expressions, structures found in so many Southeast Asian languages. These features will be discussed subsequently.

Once the segmentation was completed, we inserted tags for features that we wished to track throughout the text. One such feature, for example, was doublets or word pairs. Each time a doublet was encountered, we tagged it in the following manner: <doub>. Tagging allowed us to generate a list of all the doublets occurring within the corpus. Armed with such a list, the language teacher can target which doublets need to be pre-taught for maximum comprehension of the texts. Choosing which items to tag depends on the nature of the corpus and the interests of the researcher.

In general, the work of preparing and analyzing a corpus is not a straightforward, linear process. With each reading and rereading, researchers add to their knowledge of the corpus and thus might add to the areas they wish to explore.

After the corpus has been prepared, the next step is to run a frequency list of the vocabulary found in the text. This list is crucial in determining the level of the material. It also provides data allowing the teacher to know which vocabulary the student will most frequently encounter. It is this vocabulary that should be targeted in teaching.

We then generated an alphabetical listing of the vocabulary. This list allows the teacher to look up any given lexical item to determine its frequency within the corpus. Parenthetically, version 4.0 of WordSmith has a less-than-ideal system of alphabetizing Thai. Since it starts with the symbol farthest to the left, it first alphabetizes all words that have vowels that occur to the left of the first consonant, rather than integrating these words the way Thai is actually alphabetized. For example, the word เรียน /rian/ ‘to study’ occurs before กลาง /klaaŋ/ ‘in the middle’ in this list. However, the opposite would be true in a Thai dictionary. This is a relatively minor inconvenience, and it is easily dealt with. Secondly, there are over 450 different words written in English throughout the corpus. These words are alphabetized first, before any of the Thai words. They pose no particular challenge to the learner of Thai, since they are English and they are written in English.

Yet another feature of WordSmith, Concord, allows the researcher to make concordances, which are listings of all occurrences of a given word, with text appearing to the left and text appearing to the right of that word. The benefits of concordances are considerable, including the ability to generate collocational data. Collocates are words that tend to co-occur. By examining concordances, students are able to see authentic examples of how a word is actually used. Word-in-context examples expand the learners' depth of vocabulary knowledge. With concordances, the teacher does not have to create artificial examples to use in teaching. From these concordances, language exercises can be constructed using material actually found in the texts.

For English, there is a published list of the two-thousand most "useful" words, listed in the order of their frequency. This list, called the General Service List, was compiled in 1953 by Michael West. Although it has not been updated, it is still widely consulted by linguists and educators. Not surprisingly, a large number of the most frequently used words are function words, the words that are part of the grammar, such as 'the,' 'be,' 'of,' 'and,' 'a,' 'to,' and 'in.' They are an absolute necessity to learn, and need to be taught early on.

Our corpus comprises 4,931 types, or different words. Thirty-nine types make up 50.02% of the total running words. These thirty-nine words are basic vocabulary items and are likely to be learned in a first-year Thai class. In order to understand this corpus, it is crucial that students have a command of these words. These thirty-nine words are เธอ /thəə/ 'you, informal,' ไม่ /mây/ 'negative,' ฉัน /chǎn/ 'I, informal,' ที่ /thî/ 'that, which,' จะ /cà/ 'future,' ให้ /hây/ 'causative, to,' ก็ /kôw, kə/ 'so,' ไป /pay/ 'go,' รัก /rák/ 'love,' ว่า /wâa/ 'that, direct or indirect quotation marker,' มี /mii/ 'have,' คน /khon/ 'person,' ได้ /dây/ 'get, get to do, can,' เป็น /pen/ 'be,' มา /maa/ 'come,' กัน /kan/ 'each other,'ใจ /cay/ 'heart,' รู้ /rúu/ 'know,' มัน /man/ 'it,' แต่ /têɛ/ 'but,' นี้ /níi/ 'this,' อยู่ /yùu/ 'be, stay,' อยาก /yàak/ 'want,' ใคร /khray/ 'who,' ยัง /yan/ 'still,' แค่ /khêɛ/ 'as far as,' ๗ 'etc.' ทำ /tham/ 'do, make,' ต้อง /tôn/ 'must,' วัน /wan/ 'day,' ดี /dii/ 'good,' ความ /khwaam/ '-ness,' เรา /raw/ 'we,' กับ /kâp/ 'with,' บอก /bòk/ 'tell,' แล้ว /léɛw/ 'then, past tense marker,' นั่น /nán/ 'that,' ใน /nay/ 'in,' and อย่าง /yàaŋ/ 'like, kind.'

What is important here is the fact that fifty percent of the vocabulary in this corpus is likely to be covered very early on in a Thai class. This fact alone suggests that the corpus is highly suitable for a beginner's class, since the majority of its vocabulary consists of words that are of high frequency.

Having established the accessibility of the corpus to beginning learners, let us look at other significant features of these songs. The corpus we studied consists of four hundred songs and roughly 103,000 running words. Well over half of the songs are sung by males. Roughly three-fourths of the songs are slow ballads, a fact which makes this corpus particularly useful for relative beginners in a language, since the slow pace makes listening comprehension easier. Those songs that are not slow ballads include rap music, jazz, hip hop and dance music.

The songs are overwhelmingly melancholy in nature. They deal with themes such as lost love, shyness, romantic yearning, and scorn. Typical lines from these songs are "We've been apart for a long time," "I will dry your tears," "Do you miss me when you are lonely?" and "I know you

don't need me.” A small number of these songs (less than one quarter) deal with other topics, such as fun, sexuality, and politics.

Words dealing with sadness far outnumber words dealing with happiness in this corpus. Words meaning ‘sad,’ ‘lonely,’ ‘disheartened,’ ‘confused,’ and ‘worried’ appear in abundance throughout the corpus and need to be targeted in teaching. In contrast, words for ‘happiness,’ such as ยินดี /yindii/ and สุข /sùk/, are sparsely scattered throughout the corpus.

Nature is a theme that one might expect to find in a corpus of songs dealing primarily with love. The most frequently mentioned word related to the theme of nature is ฟ้า /fáa/ ‘sky’ or ‘heaven,’ occurring by itself ninety-nine times. It also occurs in compounds such as ท้องฟ้า /thóŋfáa/ ‘sky’ (11 times) and นางฟ้า /naaŋfáa/ ‘angel’ (6 times). Also frequent is the word น้ำ /náam/ ‘water,’ (50 times), occurring also in compounds such as น้ำตา /námtaa/ ‘tears’ (91 times), น้ำใจ /námcay/ ‘considerate of the feelings of others’ (13 times), and น้ำลาย /námlaay/ ‘saliva’ (7 times). The Thai word ลม /lom/ ‘air’ or ‘wind,’ occurs by itself forty-eight times and in compounds such as ลมหายใจ /lomhăaycay/ ‘breath’ (8 times). Other natural phenomena include ดาว /daaw/ ‘star,’ (34 times), จันทร์ /can/ ‘moon,’ (26 times), อาทิตย์ /aathít/ ‘sun,’ (3 times), นา /naa/ ‘rice field,’ (36 times), ทุ่ง /thûŋ/ ‘meadow,’ (14 times), ทะเล /thálee/ ‘sea,’ (5 times), นก /nók/ ‘bird,’ (14 times), ปลา /plaa/ ‘fish,’ (8 times), งู /ŋuu/ ‘snake,’ (10 times), แมว /mɛɛw/ ‘cat,’ (5 times), and หมา /măa/ ‘dog,’ (1 time). Clearly these words, many of which are high frequency words, need to be addressed pedagogically.

Given that so many of these songs are about love, an American audience might expect body parts such as eyes and lips to play a large role in the lyrics. But in this corpus, fewer than twenty terms for body parts are used, and used sparingly. They include เอว /eew/ ‘waist,’ แขน /khɛɛn/ ‘arm,’ หลัง /lǎŋ/ ‘back,’ ปาก /pàak/ ‘mouth,’ มือ /muu/ ‘hand,’ ฟัน /fan/ ‘tooth,’ เขี้ยว /khíaw/ ‘fang,’ หน้า /nâa/ ‘face,’ หู /hûu/ ‘ear,’ อูรา /ùraa/ and อก /òk/ ‘chest,’ ตา /taa/ ‘eye,’ คอ /khoo/ ‘neck,’ หัว /hûa/ ‘head,’ ขา /khăa/ ‘leg,’ ท้อง /thóŋ/ ‘belly,’ and แก้ม /kɛɛm/ ‘cheek.’ Overwhelmingly, the body part most frequently mentioned in the corpus is ใจ /cay/ ‘heart;’ it occurs alone roughly one thousand times, and in compounds roughly two thousand times. These compounds frequently refer to moods and personality traits, such as ใจดี /caydii/ ‘kindness’ and ใจหาย /cayhăay/ ‘frightened.’ On the other hand, จิตใจ /cítcay/, rather than being a physical body part, refers to the mind. In contrast, the next most frequently mentioned body part is มือ /muu/ ‘hand,’ which occurs fewer than one hundred and fifty times. Clearly, an essential focus of teaching should center around the word ใจ /cay/ and its compounds.

A great deal of subtlety is expressed through the use of pronouns in Thai. Pronoun choice indicates levels of formality and intimacy and thus needs to be carefully understood when listening to love songs. It comes as no surprise that the pronouns typically found in this corpus tend towards the informal. The two most frequently occurring pronouns – เธอ /thəo/ ‘you, informal’ and ฉัน /chǎn/ ‘I, informal’ – occur over four thousand times (4,356) and over three thousand times (3,081), respectively. The other pronouns which need to be taught are มัน /man/

‘it’ (973 times), เรา /raw/ ‘we’ (598 times), เขา /khǎw/ ‘he’ or ‘she’ (501 times), ผม /phǒm/ ‘I, masculine, formal’ (129 times), คุณ /khun/ ‘you, formal’ (127 times), น้อง /nǒng/ ‘I, you, younger sibling, informal’ (35 times), ไอ /ay/ ‘I, informal’ (12 times),หนู /núu/ ‘I or you, literally rat, informal’ (16 times), กู /kuu/ ‘I, informal’ (8 times), พี่ /phîi/ ‘I or you, elder sibling, informal’ (58 times), and ยู /yuu/ ‘you, informal’ (1 time).

Two highly productive grammatical morphemes occur in abundance in this corpus among those that occur in spoken and written Thai. They are ความ /khwaam/, which changes verbs to nouns and which occurs 661 times, and การ /kaan/, which corresponds roughly to a gerund in English and occurs 48 times. Specific examples are ความรัก /khwaam rák/ ‘love’ and การคิดถึง /kaan khítthǎng/ ‘feeling homesick.’ While not terribly difficult, these two constructions might not be covered in a first year Thai class; consequently, they would need to be presented before the corpus is used in a beginners’ class.

Two structural features which we did not tag are worth mentioning. The first is the structure *X negative Y*, such as นอนไม่หลับ /nǒn mâi lâp/ ‘lie down, not sleep,’ where the second verb is the unfulfilled result of the first. It is our impression that this structure occurs with moderate frequency, and thus it would be worth mentioning before teaching this corpus.

The second structure, also not tagged, is serial verbs. It is our impression that this structure is not very prominent in the corpus. Were we to do another round of tagging, we would definitely look at these structures to assess their significance within the corpus, rather than relying on our impressions.

Lewis (1993) talks of the advantages of teaching language in meaningful chunks. Thai lends itself well to this strategy. A ubiquitous feature of Thai is the pairing of words, often in rhyming doublets or words paired through alliteration. Sometimes the words are synonyms. At other times a word may occur only as a part of a pair and does not appear to have a meaning of its own. An example of a rhyming doublet is เปรียบเทียบ /pĕrĭap thĭap/, in which both words singly and together mean ‘compare.’ An example of a pair using alliteration is มากมาย /mâak maay/, in which both words, singly and together, mean ‘a lot.’ The word มาก /mâak/ is heard more frequently in everyday Thai than its pair มากมาย /maay/. There are well over one thousand such doublets occurring throughout the corpus, in almost every song. The fact that these words often rhyme or share alliteration makes it easier to recognize a doublet. It pays for the classroom teacher to teach these word pairs. They are important chunks found in everyday spoken language as well as in poetry and song. Ability to recognize and use them makes one a more comfortable and more natural user of Thai.

Also ubiquitous in both Thai and in our corpus, but less of a challenge to students, is the use of reduplication. Adjectives and adverbs can be repeated, making the meaning less precise or intensifying the meaning. Nouns, too, are sometimes repeated, as a form of pluralization. In some cases, the word is written twice, while in others it is written once and followed by ๆ, a marker which signals reduplication. Once students recognize the marker, they need give little

additional attention to this form other than to determine whether the reduplication is meant to intensify or make the meaning less precise.

Although a relatively small portion of the corpus, English words written in Thai letters pose a challenge to non-Thai readers. Scattered throughout the corpus are over one hundred words of this nature. They include words like มิคืน /mítnay/ ‘midnight,’ ลิปสติก /líp̄sətik/ ‘lipstick,’ บูมเมอเรจ /buummeeræŋ/ ‘boomerang,’ เบบี๋ /beebîi/ ‘baby,’ and แคร์ /khæɛ/ ‘care.’ Many of these words, though clearly not all, are easily associated with western culture and modernity. What is important is that, even for relatively experienced students of the language, these words are difficult to recognize when encountered in the text. There is nothing that marks them as non-Thai, and one tends to try to interpret them along Thai lines until the realization hits that the word is really an English word. A worthwhile exercise to give before these songs are studied would be one that matches the English words written in the two different alphabets, Thai and English. That way, when encountered, the student is already familiar with the word and less likely to be caught off guard. An exercise on this phenomenon can be found in the Appendix.

Sanskrit words are sparsely scattered throughout this corpus. The word กรรม /kam/ ‘karma,’ for example, occurs thirteen times. The word ยุติธรรม /yútitham/ ‘justice’ occurs three times, and อัศจรรย์ /àtsacan/ ‘miraculous,’ occurs just once. These words are best dealt with as they arise within a given song, since they are relatively infrequent.

Given the highly modern content of this particular corpus, it is interesting to note that certain features found in classical Thai poetry are also found in this modern genre. Classical Thai poetry is known for an abundance of internal rhyme, where multiple words rhyme in a given line and sometimes extend into the next line as well. Our corpus is filled with examples of internal rhyme such as มีใจเก็บไว้จะไห้เธอ /mii cay kèp wáy cà hây thəə/, where ใจ /cay/, ไว้ /wáy/, and ไ้ /hây/ all rhyme.

Similarly, alliteration, where multiple words begin with the same sound, is rampant throughout the corpus. An example of this phenomenon is ทุกถ้อยคำที่เธอเอ่ย /thúk thŏy kham thîi thəə əəy/, where ทุก /thúk/, ถ้อย /thŏy/, ที่ /thîi/, and เธอ /thəə/, all begin with a /th/ sound.

A hallmark feature of many Southeast Asian languages, including Thai, is the use of elaborate expressions, in both spoken and written language. This form of word play gives beauty and elegance to the language, and demonstrates a certain linguistic virtuosity. Two patterns are highly typical, ABAC and ABCB, although variations on these two patterns can and do occur. An example of the ABAC pattern found in this corpus is ยิ่งไกลยิ่งชิด /yîŋ klây yîŋ chít/, where the first and third syllables, ยิ่ง /yîŋ/, are repeated. An example of the ABCB pattern found in the corpus is ตัวเองที่ตัวเอง /tua eej thîaw eej/, where the second and fourth syllables, เอง /eej/, are repeated.

The frequent use of doublets, reduplications, internal rhyme, alliteration, and elaborate expressions all point to the need for the classroom teacher to discuss these features before introducing the songs in this corpus. Students who go on to study advanced Thai are likely to

come across these forms. These popular songs provide a reason to introduce these features of the language before the student reaches advanced levels of language study.

A final feature worth noting involves the written form of certain words in this corpus. Throughout the corpus we have found variant spellings for certain words. They are spoken transliterations of spelled forms, such as รึ/รึ / instead of หรือ /rǎu/ ‘or,’ ไร /ray/ instead of อย่างไร /yàaŋray/ or ไรไร /yaŋŋay/ ‘how,’ and เล้า /kháw/ for เขา /khǎw/ ‘he, she, they. While not terribly frequent, they can be confusing to the learner, and they are worth pointing out to students of these songs.

Having explained the process and some of our findings, let us look at some language exercises based on high-frequency vocabulary highlighted in this study.

Appendix I

EXERCISES

Pt. I. Match the Thai (1-4) with the English equivalent (A-D).

1. แล้วเธอจะเกลียดไหม [C]
2. เธอให้สัญญาว่าเธอจะมีฉันรักกันตลอดในชาตินี้ [A]
3. อยากจะขอให้เธอทบทวน [B]
4. หากเธอได้เห็นว่าในบางทีก็มีที่หวั่นไหว [D]

- A. You promised that we would love each other for the rest of our lives.
 - B. I want you to think it over.
 - C. Would you hate me?
 - D. If you found that sometimes there would be uncertainty...
-

Pt. II. Use your knowledge of Thai to help you answer these questions.

1. เธอให้สัญญาว่าเธอจะมีฉัน รักกันตลอดในชาตินี้
(You promised we'd love each other for the rest of our lives.)

Which word means promise? (50 times)

2. แล้วเธอจะเกลียดไหม
(Would you hate me?)

Which word means hate? (29 times)

3. อยากจะขอให้เธอทบทวน
(I want you to think it over.)

Which word means think it over? (21 times)

4. หากเธอได้เห็นว่าในบางทีก็มีที่ **หวั่นไหว**
(If you found that sometimes there would be uncertainty...)

Which word means uncertainty? (34 times)

CLOZE EXERCISE

Fill in the blanks in the sentences with the correct word from the list of 5 below.

ทบทวน คบ สัญญา หวั่นไหว เกลียด

1. แล้วเธอจะ **เกลียด** ไหม
Would you hate me?
 2. เธอให้ **สัญญา** ว่าเธอจะมีฉัน รักกันตลอดในชาตินี้
You promised we'd love each other for the rest of our lives.
 3. อยากจะขอให้เธอ **ทบทวน**.
I want you to think it over.
 4. หากเธอได้เห็นว่าในบางทีก็มีที่ **หวั่นไหว**.
If you found that sometimes there would be uncertainty...
-

PRONOUNS

Choose the best pronoun. (Answers in **red**.)

1. ฉันรัก _____ (I love you.)
ผม (I) เรา (we) **เธอ (you)**
 2. _____ เป็นสาวอีสาน (I'm a girl from the northeast.)
ผม (I-masculine) เธอ (you) **น้อง (I-younger one/term of endearment)**
-

PHRASE COMPLETION

Join phrase segments from columns A and B.

COLUMN A

คนแรก

ไม่ต้องรักฉันมาก

ยืมเป็นพี่ชายได้

COLUMN B

กว่าใครๆ

ไหมชั่วคราว

ที่รัก

Write out the complete phrase.

1. The first person I've ever loved.

คนแรกที่รัก

2. May I borrow you to be my big brother for a while?

ยืมเป็นพี่ชายได้ไหมชั่วคราว

3. No need to love me more than anyone else.

ไม่ต้องรักฉันมากกว่าใครๆ

CONCORDANCE

Study the concordance lines. Then answer the questions that follow the list.

- | | | |
|------------|-----|----------------------|
| 1. นั้น | คิด | ถึงเธออยู่ตลอดเวลา |
| 2. เขาไม่ | คิด | มีใจ |
| 3. นั้นไม่ | คิด | ว่าเธอจะมีอีกคน |
| 4. ยังคง | คิด | ถึงเธอตลอดมา |
| 5. นั้น | คิด | ว่าคงหน้าตาเหมือนเธอ |
| 6. เพราะ | คิด | ถึงเธอไม่อยากไปไหน |
| 7. ไม่ | คิด | ไปจริงจังจริง |
| 8. ไม่ | คิด | ว่าเธอจะมีอีกคน |

1. In this concordance what words typically followed คิด? ถึง ว่า
2. What kind of word typically followed ถึง? A pronoun
3. Is it formal or informal? Informal
4. What kind of structure typically followed ว่า? A clause
6. How would you translate ว่า into English? That

WORD PAIRS

Make word pairs using items from the list below. (Answers in red.)

ตาย นอน มาย สน อุ่น

มากมาย (a lot)

แน่นอน (certain)

อบอุ่น (good-natured)

เดี่ยวตาย (lonely)

สับสนสน (confused)

FEELINGS EXPRESSED WITH ใจ

Underline the best answer.

1. Your lover just left you. How do you feel?

ดีใจ (happy)

ประทับใจ (impressed)

หมดใจ (disheartened)

2. Your lover has found somebody new. How do you feel?

ช้ำใจ (hurt)

จริงใจ (sincere)

ไว้ใจ (trusting)

3. You are not sure what to do. How do you feel?

วุ่นวายใจ (confused)

ไว้ใจ (trusting)

อุ่นใจ (secure)

OPPOSITES

Draw lines to connect the opposites in columns A and B.

COLUMN A

จน (poor)

เลว (bad)

เกลียด (hate)

เศร้า (sad)

COLUMN B

ยินดี (happy)

รัก (love)

ดี (good)

รวย (rich)

FINDING ENGLISH WORDS

Each of these lines has an English word written in Thai letters. Underline the English word. Choose from the following:

professor, care, technology, concert, BMW, World Trade, cocaine

1. โปรดอย่าทำลายด้วย**เทคโนโลยี**
2. มันไม่**แคร์**และไม่**หวัง**
3. **เวิร์ล**เทรคก็**คูดี**
4. เปลี่ยนให้เขาได้นั่ง**บีเอ็มดับเบิลยู**
5. อยากเปลี่ยน**โคเคน**เป็น**แดงโม**
6. ใต้เป็น**โปรเฟรชเซอร์**
7. ขอให้ชีวิตสดใสยิ่งกว่าแสงไฟ**คอนเสิร์ต**

Appendix II

TAGGING EXAMPLES

ไม่ ต้อง แงะ <redup>ติด ๆ<redup> กัน ไว้

<elab>จะ กอด ไม่ กอด<elab>

กอด กัน นะ ตาม <Sanskrit>ธรรมเนียม<Sanskrit>คน

ระหว่าง เรา<doub>มากมาย<doub>

วัน <eng>วาเลนไทน์<eng> จะ ส่ง ดอกไม้ เรา จะ ไป ให้ ใคร

<englet>Welcome to 5 club<englet>ยินดี

เก็บ ไว้ ใน <body>อก<body> นี้ มัน อัด อยู่ ใน นี้

<rhyme>ไม่ ว่า เธอ จะ อยู่ ที่ ไค ก็ ยัง ส่ง ใจ ไป<rhyme>ถึง

<alit>ถ้อยคำ ที่ เธอ<alit> เคย บอก นั้น

ELABORATE EXPRESSIONS

ยั้ง ไกล ยั้ง ชิด

A B A C

or

ตัว เอง เทียว เอง

A B C B

ALTERNATE SPELLINGS

รี for หรือ

ไม้ for ไหม

ไฉ for ยังไฉ or อย่างไร

เค้า for เขา

Appendix III

Word Frequency List: 400 Most Frequent to Less Frequent

1	เธอ	8	รัก	15	ใจ	22	อยู่
2	ไม่	9	ไป	16	มา	23	อยาก
3	ฉัน	10	ว่า	17	กัน	24	ทำ
4	ที่	11	มี	18	แต่	25	ๆ
5	ให้	12	คน	19	รู้	26	ใคร
6	จะ	13	ได้	20	มัน	27	แค่
7	ก็	14	เป็น	21	นี้	28	วัน

29	ดี	63	อย่า	97	หน้า	131	ไหว
30	ต้อง	64	เพียง	98	คือ	132	ข้าง
31	ยัง	65	เอา	99	หาก	133	รับ
32	ความ	66	เจอ	100	พบ	134	หนึ่ง
33	เรา	67	เมื่อ	101	อาจ	135	เก็บ
34	นั้น	68	ใช่	102	ต่อ	136	ผม
35	กับ	69	นาน	103	ทาง	137	คุณ
36	แล้ว	70	พูด	104	กว่า	138	นอน
37	คง	71	ขอม	105	เปล่า	139	จำ
38	ใหม่	72	ที่	106	ช่วย	140	ถาม
39	บอก	73	เจ็บ	107	ไกล	141	เพื่อน
40	เคย	74	สัก	108	ครั้ง	142	หญิง
41	ใน	75	เอง	109	อีก	143	ชีวิต
42	อย่าง	76	พอ	110	จน	144	หวัง
43	เขา	77	มอง	111	สิ่ง	145	เหงา
44	ขอ	78	เดียว	112	ผ่าน	146	ซัก
45	และ	79	ลืม	113	มาก	147	ฟัง
46	หัวใจ	80	หรือ	114	ทิ้ง	148	ลง
47	คิด	81	ตา	115	หมด	149	สาย
48	ทุก	82	ฝัน	116	คอย	150	ขึ้น
49	เลย	83	นะ	117	เกิน	151	ฟ้า
50	ของ	84	กลับ	118	เหลือ	152	ชาย
51	จริง	85	ด้วย	119	ใด	153	หาย
52	ถึง	86	รอ	120	ยังไง	154	แม่
53	เหมือน	87	ทั้ง	121	คืน	155	น่า
54	ไว้	88	เข้าใจ	122	ชอบ	156	ปล่อย
55	คำ	89	ผู้	123	แม่	157	ลม
56	ตัว	90	ยัง	124	มือ	158	ต้องการ
57	อะไร	91	เรื่อง	125	หา	159	ไป
58	เพราะ	92	ตรง	126	เข้า	160	หวัง
59	จาก	93	เห็น	127	เดิม	161	จบ
60	ถ้า	94	เวลา	128	เพื่อ	162	ตาย
61	ดู	95	เดิน	129	ทำไม	163	น้ำ
62	ไหน	96	เปลี่ยน	130	ทน	164	แสน

165	ตอน	199	ร้อง	233	บ้าน	267	เต็ม
166	ร้าย	200	ฝาก	234	สวย	268	กล้า
167	บาง	201	ตลอด	235	อย่างไร	269	ร้อย
168	บ้าง	202	ยื่น	236	แถม	270	สัญญา
169	ลา	203	โลก	237	อ่อน	271	ค่อย
170	ซ้ำ	204	กลัว	238	เท่าไร	272	ราว
171	หน่อย	205	นั่ง	239	เฝ้า	273	เหนื่อย
172	เสียใจ	206	ลอง	240	เริ่ม	274	แทน
173	แบบ	207	สุดท้าย	241	ตก	275	โทษ
174	สอง	208	เจ้า	242	สุด	276	ง่าย
175	ผิด	209	โดน	243	เปิด	277	นะ
176	เท่าไร	210	ถูก	244	ยิ้ม	278	หยุด
177	ใหม่	211	สั่ง	245	ภาพ	279	ดัง
178	เก่า	212	แอบ	246	หรรษา	280	ราว
179	ใกล้	213	รู้สึก	247	ห้วน	281	ถูก
180	จึง	214	หมาย	248	เสียง	282	เช่น
181	ตาม	215	เท่านั้น	249	แท้	283	ขาด
182	ตั้ง	216	เรื่อย	250	ช่าง	284	คู่
183	พร้อม	217	ไร	251	นี่	285	ซึ่ง
184	แรง	218	น้ำตา	252	สำคัญ	286	สาว
185	ต่าง	219	พัน	253	สุข	287	ห้าม
186	ห่าง	220	เท่า	254	เล่น	288	โย
187	กอด	221	เมื่อไหร่	255	การ	289	ดวง
188	ก็	222	เหมียว	256	นั่น	290	คาย
189	ตอบ	223	ทัก	257	มาย	291	ผล
190	พอ	224	พี่	258	เหตุผล	292	จง
191	ออก	225	รี	259	ปวด	293	นาที่
192	เพลง	226	เพิ่ง	260	พวก	294	นาย
193	เสมอ	227	บน	261	สั้น	295	หลอก
194	แฟน	228	กลับ	262	เรียก	296	เค้า
195	ได้ยิน	229	คุย	263	นิด	297	เลือก
196	กัด	230	เพื่อ	264	ว่าง	298	โทร
197	เกิด	231	เสีย	265	สนใจ	299	น้อง
198	ก่อน	232	ใช้	266	เชื่อ	300	ยาก

301	ส่ง	335	เคียง	369	อุ
302	ติด	336	ยื่น	370	เกลียด
303	ทัน	337	เรียน	371	เนี่ย
304	ผูก	338	เลื่อน	372	แก้ง
305	สบาย	339	ดอกไม้	373	แมว
306	หนาว	340	ปี	374	แรก
307	กลาง	341	ลอย	375	จี้
308	กลาย	342	หวาน	376	จ้ง
309	นั่ง	343	อ้าง	377	ซึก
310	ปาก	344	เจอะ	378	ตั้งแต่
311	เข้า	345	เฉย	379	ถือ
312	เดา	346	โชค	380	ท้อง
313	โดย	347	โปรด	381	ทุ
314	อื่น	348	คอ	382	นาง
315	เล็ก	349	สน	383	ลบ
316	กาย	350	คบ	384	ว่าง
317	ชะ	351	ค่า	385	หมดใจ
318	ดาว	352	ซ้ำ	386	อาการ
319	น้อย	353	ลำบาก	387	เต็ม
320	นับ	354	สิทธิ์	388	เมือง
321	มัย	355	โกรธ	389	กราว
322	รู้จัก	356	จับ	390	จาง
323	หายใจ	357	ซ้ำ	391	ซิด
324	แคว	358	ตื่น	392	ดวงใจ
325	แพ้	359	พัด	393	ตัน
326	ใส่	360	ละ	394	บังเอิญ
327	ไฟ	361	หน	395	ฝน
328	ควร	362	อุ่น	396	มัน
329	งาม	363	เดอ	397	ร้องไห้
330	นา	364	เมื่อไร	398	วอน
331	ปิด	365	ใหญ่	399	เขี้ยว
332	พวง	366	ท้าย	400	เท
333	ซ้ำ	367	พา		
334	ลึก	368	มาลัย		

Acknowledgements

We would like to thank Carol J. Compton, Rhodalyne Crail, John Hartmann and Amara Vatanakul for their kind assistance with this article.

References

Electronic References

eThaiMusic.com, last update 07/19/2006, viewed 01/2005 – 07-2006.
<<http://www.ethaimusic.com/>>.

Print References

Aston, Guy, Silvia Bernardini, and Stewart Dominic. (2004). *Corpora and Language Learners*. Philadelphia: John Benjamins North America.

Bofman, Theodora Helene. (1978). *The Poetics of the Ramakian*. The University of Michigan. Doctoral dissertation.

Ghadessy, Mohsen, Alex Henry, and Robert L. Roseberry. (2001). *Small Corpus Studies and ELT: Theory and Practice*. Philadelphia: John Benjamins North America.

Hunston, Susan. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Iwasaki, Shoichi and Preeya Ingkaphirom. (2005). *A Reference Grammar of Thai*. New York: Cambridge University Press.

Lewis, Michael. (1993). *The Lexical Approach: The State of ELT and a Way Forward*. Hove, England: Language Teaching Publications.

_____. (1997). *Implementing the Lexical Approach: Putting Theory into Practice*. Hove, England: Language Teaching Publications.

_____, ed. (2000). *Teaching Collocation: Further Developments in the Lexical Approach*. Hove, England: Language Teaching Publications.

McEnery, Tony and Andrew Wilson. (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Moore, Christopher. (1992). *Heart Talk*. Bangkok: White Lotus.

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

_____. (2005). "Teaching and Learning Vocabulary" In Eli Hinkel (Ed.) *Handbook of Research in Second Language Teaching and Learning*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Smyth, David. (2002). *Thai: An Essential Grammar*. New York: Routledge.

West, Michael. (1953). *A General Service List of English Words*. London: Longman, Green.