

A CAI PROGRAM FOR TEACHING FILIPINO
Curtis D. McFarland
Waseda University School of Science and Engineering

Introduction. I have developed a computer-assisted program for teaching reading in Filipino (Tagalog) which I am currently using in Japan. I realize that this conference is primarily for linguistics and syntactic analysis, and not for language-teaching materials. However this program is directly related to syntactic analysis in several ways. In recent years Tagalog and other Filipino languages have become favorite sources for data in linguistic arguments such as that about subject and topic. Unfortunately in many cases the linguists citing Filipino data have very limited knowledge of the language, and even some of those specialized in the language sometimes go astray.

One problem has to do with the difference between relativized and non-relativized environments. For example Filipino speakers would say (I shall avoid the word 'can'; almost anything can be said):

Siya ang pumatay kay Mario. 'He is the one who killed Mario.'

but not

*Pumatay siya kay Mario. (He killed Mario)

Again they would say

Ang Toyota ang pinagtatrabuhan niya. 'Toyota is where he works.'

but not

*Pinagtatrabuhan niya ang Toyota. (He works at Toyota)

Again

Ang Quiapo ang pinuntahan niya. 'Quiapo is where he went.'

but not

*Pinuntahan niya ang Quiapo. (He went to Quiapo)

The distinctions at first glance seem very slight, but the constraints underlying them are very strong. And they (along with many other little rules) are the source of many errors not only in linguistic arguments, but even in descriptions of Filipino and pedagogical materials. We would therefore like to appeal to anyone making use of Filipino data, for any reason, to acquire more than a superficial knowledge of the language. And we offer this program as one means for accomplishing this task.

Underlying Philosophy. As a language teacher of both English and Filipino, I feel that the best language materials are 'real' language materials, that is, language that has actually been used to communicate, whether to transmit information, tell a story, or whatever. In other words, it is not material which was created for the purpose of teaching the language. For that reason this program has been built upon two 1,000,000-word corpuses of written Filipino: one drawn from a wide range of Tagalog literature, the other from 50 popular mini-novels published in the 1990s. A list of the materials in the first corpus can be found in my *A Frequency Count of Pilipino* (Manila, 1989, Linguistic Society of the Philippines; the books in the second corpus are listed in Appendix A.

A second basis of my philosophy is that the key to language learning is repetition: repeated exposure to and/or use of the same linguistic elements in a range of environment. Repetition of these elements reinforces our understanding of them, while at the same time building a context into which subsequent elements will be introduced as our learning

progresses. It is a waste of time to memorize words and phrases which will not be encountered until 25 lessons on, or in some cases, ever again. And the best way to achieve effective repetition is to base the lessons on the frequency of words in the language. Frequency counts of the two corpuses were conducted and the 2,500 most frequent words were identified. These 2,500 words are the targets of these lessons. (The most frequent 100 words are listed in Appendix B.)

The concordances. A central factor in the development of this program are concordances of the two corpuses. These concordances differ from most concordances with which you may be familiar in that they are interactive. To go back for a moment, the first step in the frequency-count program was actually the creation of an index of words in the corpus, before the actual count was made. This index is a primary component in the concordance program (Figure 1). In running the program, the user first searches the index for the item he is interested in (or he might just browse the index looking for interesting items).

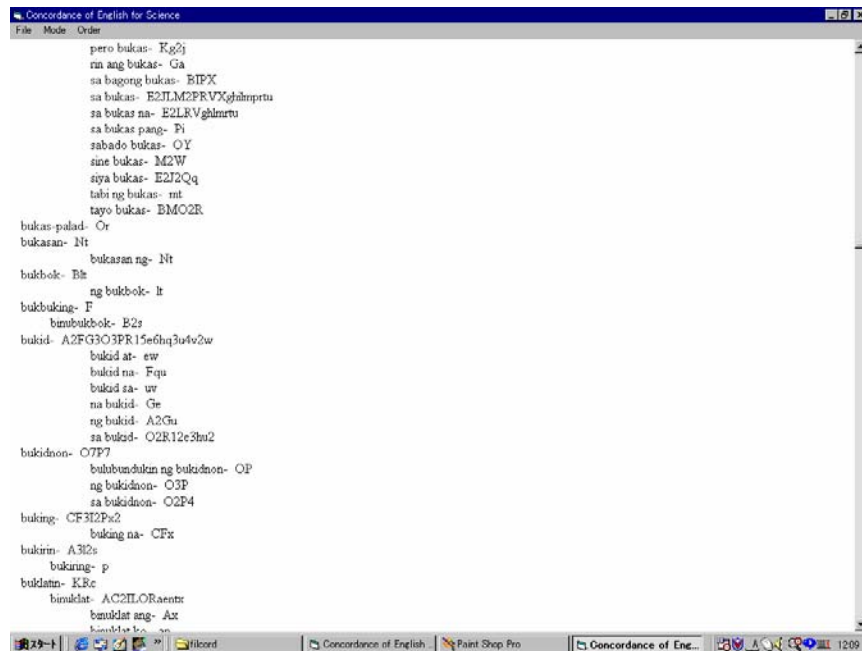


Figure 1

When he 'clicks' the mouse on the desired item, the sentences in which that item occurs are presented on the computer screen. For example, if he clicks **bukid**, the screen shown in Figure 2 appears.

These concordances are valuable tools in linguistic analysis. Large numbers of examples of the occurrences and collocations of individual words and various constructions. Examination of the examples allows us to describe the constructions and the conditions for their occurrence with a high degree of precision. On the other hand we can observe the non-occurrence of other constructions. I concede that non-occurrence is in a sense not real evidence. However non-occurrence, or even rare occurrence, of a linguistic feature suggests that that feature is subject to some constraint.

These concordances are not generally available for distribution, for a number of reasons, not the least of which is its length. The indexes alone, if printed out, would be about 2000 pages. However, they can be shared, under certain conditions, with academic colleagues.

III. The readings

A set of readings was then constructed from the original corpuses. The first step in this process was to run a computer program which searches each corpus for connected sequences of words which are all above a certain level of frequency. For Chapters 1A and 1B, the level was 100 words. For Chapters 2A and 2B, the level was 200 words, etc. (The chapters designated A are drawn from the first corpus; B chapters, from the second corpus.)

At each level, several hundred selections were produced. (This number can be moderated in terms of the minimum length of selection required, and the number or frequency of lower-frequency words allowed.) Of these selections, 20 were selected, on the basis of a number of criteria. One criterion was coherence or readability: Is the selection understandable as a unit within a conversation or part of a narrative; can its context be deduced or imagined without additional explanation. And other abstract criteria were used: Was the selection interesting, amusing, etc? (The first pages from Lesson 1A, Lesson 5B and Lesson 10A are shown in Figures 4-6.)

Translations are provided for each reading. At present these are English translations. A second version with Japanese translations is now in preparation. Students are encouraged to learn inductively by observing for themselves how the various words are used in different contexts.

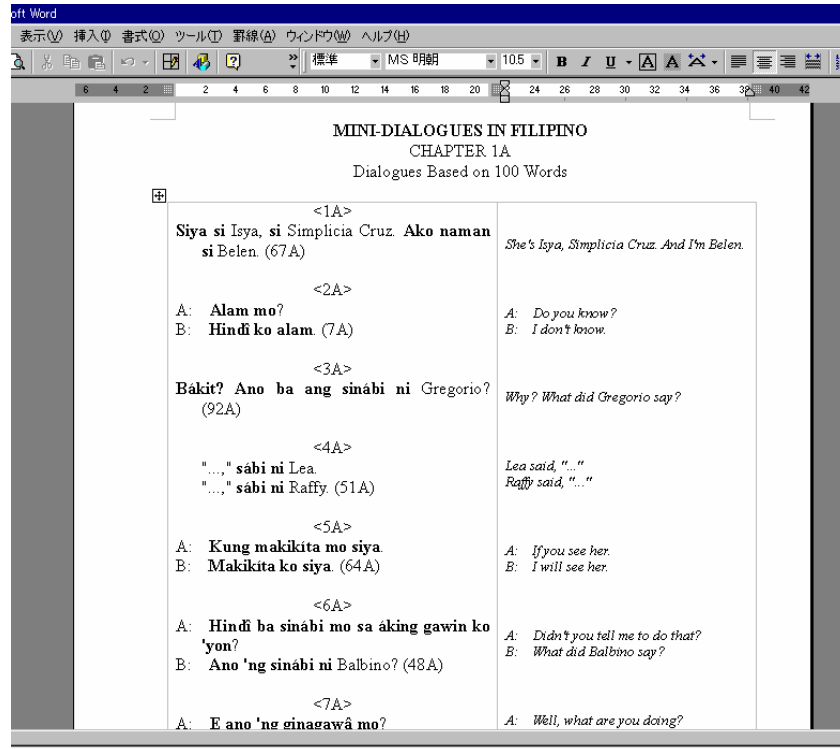


Figure 4

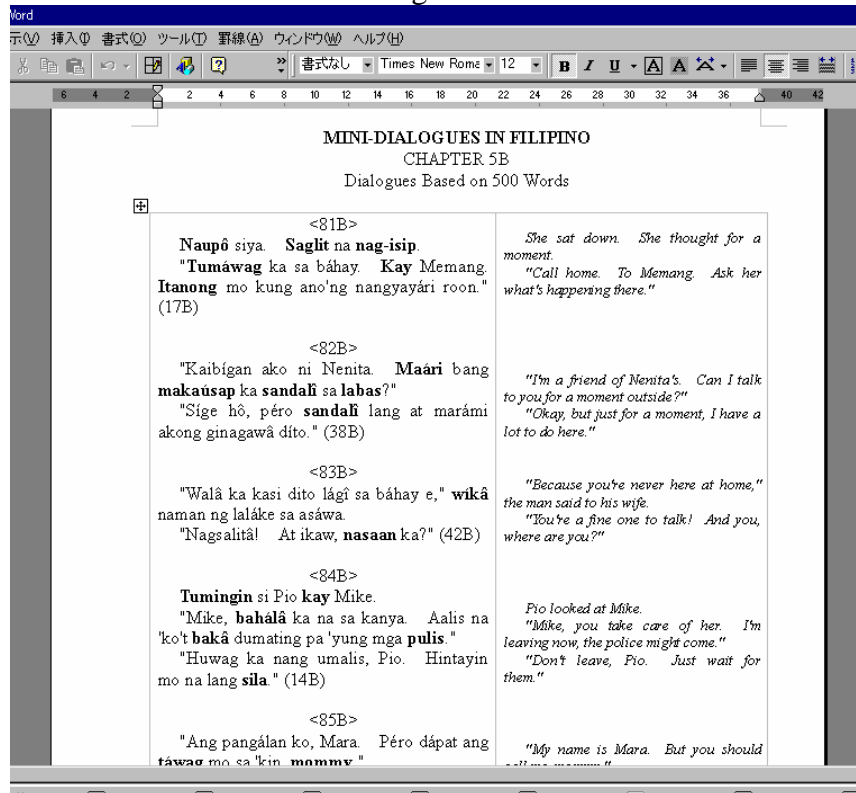


Figure 5

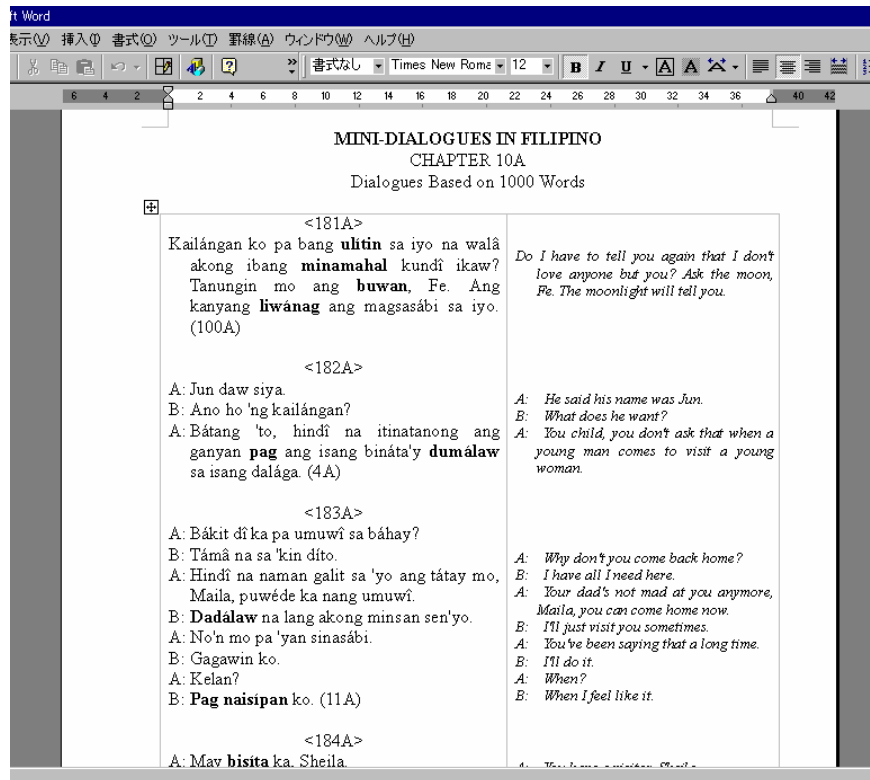


Figure 6

The main advantage of this approach is that we have a high degree of control of the rate at which new vocabulary is introduced. And by following the frequency, we ensure that the first words introduced will occur repeatedly in subsequent readings. This does two things: (1) it reinforces the students' understanding of the meanings and uses of these words; (2) these words become part of the 'familiar' semantic and syntactic context for the introduction of other words in later lessons. Of course the order of introduction of words does not follow the frequency absolutely, for various reasons. For one thing, some high-frequency words tend to combine with much lower-frequency words, and thus tend not to be found in the connected sequences of high-frequency words described above. In fact, in fifty lessons only about half of the top 2500 words in the rank list are introduced. Nonetheless the words that are introduced basically follow the frequency order.

One possible disadvantage of this approach is that it is not always possible (especially in the early lessons) to select readings for which the contexts are completely clear. Even this might be an advantage if it induces the students to examine the linguistic and semantic clues to think of possible contexts for the readings.

Students are not expected to memorize new words upon their first introduction. If they can do it, or know the words already, so much the better. However, our approach encourages the students to become gradually familiar with the words as they occur repeatedly.

The first time a given word occurs in the readings, it is highlighted (bolded) in the text, and placed in the New Words section of the glossary for that lesson.

The second time the word occurs, it is highlighted in the text, and placed in the Review Words section of the glossary. The glossary entry contains a reference to the first occurrence of the word, allowing the student to compare the two occurrences.

The third time the word occurs, it is highlighted in the text, and placed in the Words to Memorize section of the glossary, again with references to previous occurrences. Finding a word in this section should encourage the student to memorize the word at this stage and to warn him/her that subsequent occurrences of this word will not be entered in the glossary.

Following the glossary is a section called 'Sentences to Remember'. This section lists all the sentences (from Chapter 1A to the current chapter), which (1) contain at least one word included in the 'Words to be Memorized' part of the glossary and (2) consists completely of words which have previously occurred in three or more readings. The implication is that at this stage should be able to understand these sentences without translation, and perhaps to understand them more fully than they could when they were first encountered. (Parts of the glossary and 'Sentences to Remember' section for Chapter 5 are shown in Figure 7.)

A new section currently being added is the 'Grammatical Structures'. This section also follows the 'Rule of Three'. The first time a particular structure occurs in a reading, students are encouraged to figure it out inductively, or at least to make note of it without explanation. The third time the same structure occurs, it will be explained in the 'GS' section, together with ten example sentences (from past and subsequent chapters) of that structure. For structures for which there are fewer than ten examples in the entire set of 50 chapters, additional examples will be added from the concordances. The lessons will contain no created sentence examples

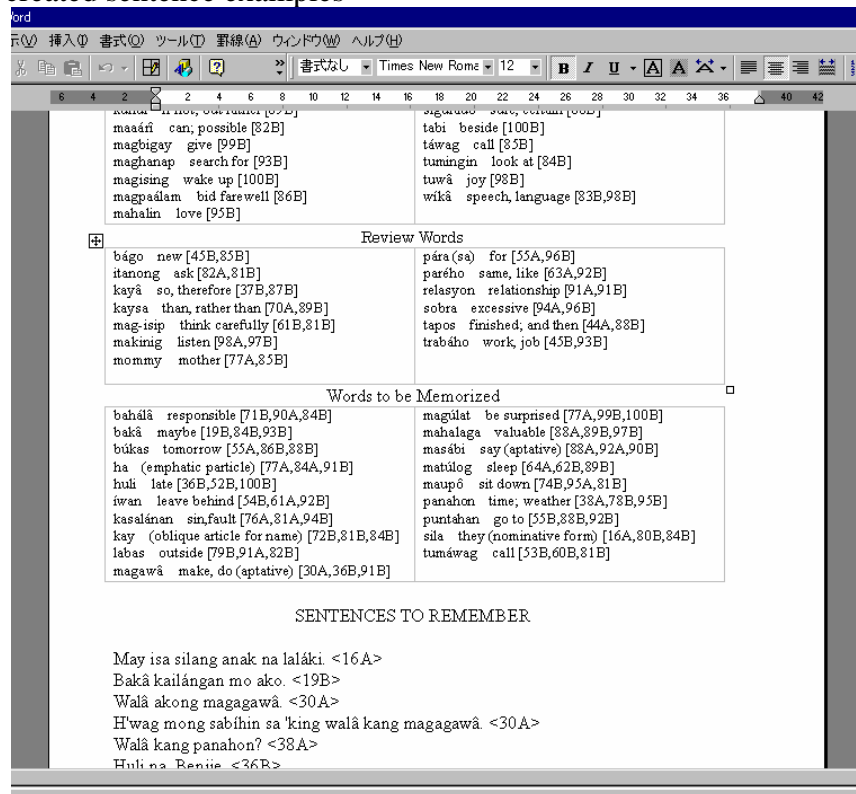


Figure 7

The readings are seen as a primary tool for classroom teaching, but they could also serve as an independent self-study reading program. However, they have proved to be very difficult for students beginning at the zero level (no previous knowledge of Filipino),

and thus seem more appropriate at the intermediate level, or in conjunction with a more standard type of instruction. We are expecting to publish the readings within this year.

IV. The computer programs.

The readings are accompanied by a package of ten computer programs (one reading program, four drills, four games, and a glossary browser). The computer programs are seen as a self-study tool, either as a supplement to classroom work, or for independent study.³

The first program is a simple reading program which reproduces the reading texts in a computerized format. Some students may prefer to read the text in the computer or the printed text may not be available to them.

New vocabulary items which are bolded in the printed texts are highlighted in color (blue) in the screen text. The meanings of a new item can be seen by clicking on a given word. The glossary entry for that word appears at the bottom of the screen. (Some students will definitely prefer this to searching for a word in the glossary.) If the student would like to see additional examples of the same word occurring elsewhere in the set of lessons, he can do so by using the examples menu. He can see up to eight examples of the same word

The four drills consist of a vocabulary-in-context drill, a collocation-grammar drill, and two vocabulary flashcard drills. These drills are designed for individual study, but of course students may choose to work together to do them.

The first drill is a drill for the learning of vocabulary in context (Figure 8). The text appears on the screen as in the reading program, with new items highlighted. Most of the items are highlighted in blue, but one word is highlighted in red, and ten definitions appear on the right side of the screen. The student tries to select the correct definition for the red item by clicking on that definition. He/she receives immediate evaluation of the answer (right or wrong). If the answer is wrong, the correct answer is shown. Whether right or wrong the complete glossary entry is displayed at the bottom of the screen.

If the answer is correct, the highlight of the red item changes to green, otherwise it goes back to blue. The next new word is highlighted in red, and the drill continues. At the end of the lesson, or when the student terminates the drill, he/she is shown his/her score of rights and wrongs.

One strength of this drill is that it encourages the student to consider the semantic and syntactic environment in 'guessing' the definitions of unfamiliar words, both in terms of their meanings and the part of speech which would fit in a particular position.

Vocabulary in Context - CHAPTER 5B
File Examples

CHAPTER 5B
<81B>

Naupo siya. Saglit na nag-isip.
"Tumawag ka sa bahay. Kay Memang. Itanong mo kung ano'ng nangyayari roon." (17)

<82B>

"Kaibigan ako ni Nenita. Maari bang makausap ka sandali sa labas?"
"Sige ho, pero sandali lang at marami akong ginagawa dito." (38)

<83B>

"Wala ka kasi dito lagi sa bahay e," wika naman ng lalake sa asawa.
"Nagsalita! At ikaw, nasaan ka?" (42)

<84B>

Tumingin si Pio kay Mike.
"Mike, bahala ka na sa kanya. Aalis na 'ko't baka

sleep
sit down
mother
(title for man)
be where
when
time; weather
for
same, like
policeman

SKIP maupo /maqupoq/ sit down NEXT

Figure 8

The second drill is considered the most difficult and most important program in this set of programs. It is a drill on grammar and collocation of words (Figure 9).

After a given word has appeared three times as a new vocabulary item, this word becomes a Target Item for this drill for its next five occurrences. Thus each chapter contains, in addition to its regular glossary, a hidden glossary containing these Target Items.

As in the vocabulary drill the text appears on the screen. The New Items are no longer highlighted. The Target Items are replaced by blank spaces (surrounded by brackets). Most of the brackets are highlighted in blue, but one set of brackets is highlighted in red, and ten Target Items appear on the right side of the screen. The student tries to find the Target Item that fits in the highlighted blank space and click on that item.

The Target Items are listed in their basic forms. If the particular word can appear in different forms (aspect, focus, etc.) the student must first choose it correctly from the list and then click again on the correct form.

As in the vocabulary drill the student gets an evaluation of each answer and a score at the end of the drill.

Grammar Drills - CHAPTER 5B
File Examples

CHAPTER 5B
<81B>
Naupo siya. Saglit na nag-isip.
"Tumawag ka sa [?]. Kay Memang. Itanong mo kung
ano'ng nangyayari [?]." (17)

<82B>
"[?] ako ni Nenita. Maari bang makausap ka sandali
sa labas?"
"[?] ho, [?] sandali lang at [?] akong
ginagawa dito." (38)

<83B>
"Wala ka [?] dito [?] sa [?] [?],"
wika naman ng [?] sa [?].
"[?]! At ikaw, nasaan ka?" (42)

<84B>
Tumingin si Pio kay Mike.
"Mike, bahala ka na sa [?]. [?] na 'ko't baka

agad
asawa
ayaw
bahay
buhay
bumalik
dapat
diyos
doon
dumating

SKIP bahay // house NEXT

Figure 9

This drill is considered important in encouraging the students to examine the semantic and syntactic environment of the blank space in choosing an appropriate word both in terms of meaning and class of word. The student must also choose the correct form of the word.

This drill can be done before the lesson is read, as a challenge to the students. It is also effective when done after reading the lesson, to help the students internalize the content of the lesson.

The last two drills are simple vocabulary memorization drills. Nowadays, we prefer practice to memorization in language learning. Yet we must admit that language learning does entail a certain degree of memorization. Learning vocabulary or at least some parts of the vocabulary is especially conducive to memorization. We used to (or maybe still do) use flashcards for vocabulary memorization. These drills are a more efficient and less cumbersome (no dropping of the cards) version of flashcards.

For the third drill (Figure 10) the computer randomly selects ten items from the glossary for the given chapter. One of these ten items (the Filipino term) is displayed in the center of the screen. The definitions of the ten items are displayed on the right side. The student tries to click the correct definition. As in the other drills, the answer is evaluated, and the complete glossary entry appears at the bottom of the screen.

If the answer is correct, that item is removed from the list and replaced by another item from the glossary. If the answer is wrong, the item remains in the list and will be asked again later in the drill. The drill continues until all glossary items have been correctly answered (or the student terminates the program).

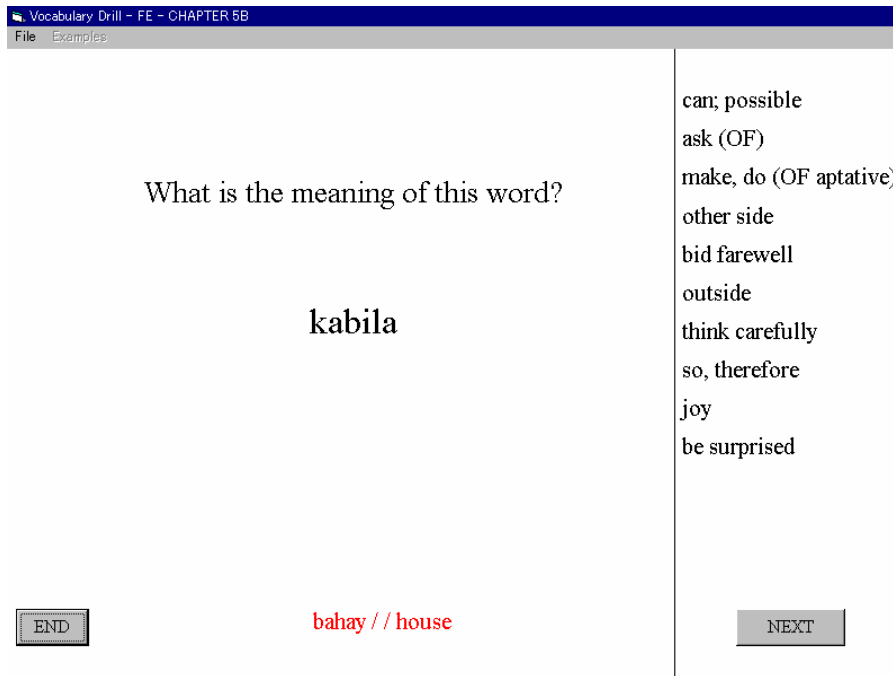


Figure 10

The fourth drill (Figure 11) is exactly the same as the third, except in reverse. English (or Japanese) definitions are displayed in the center and the Filipino items are placed on the right side as the answers.

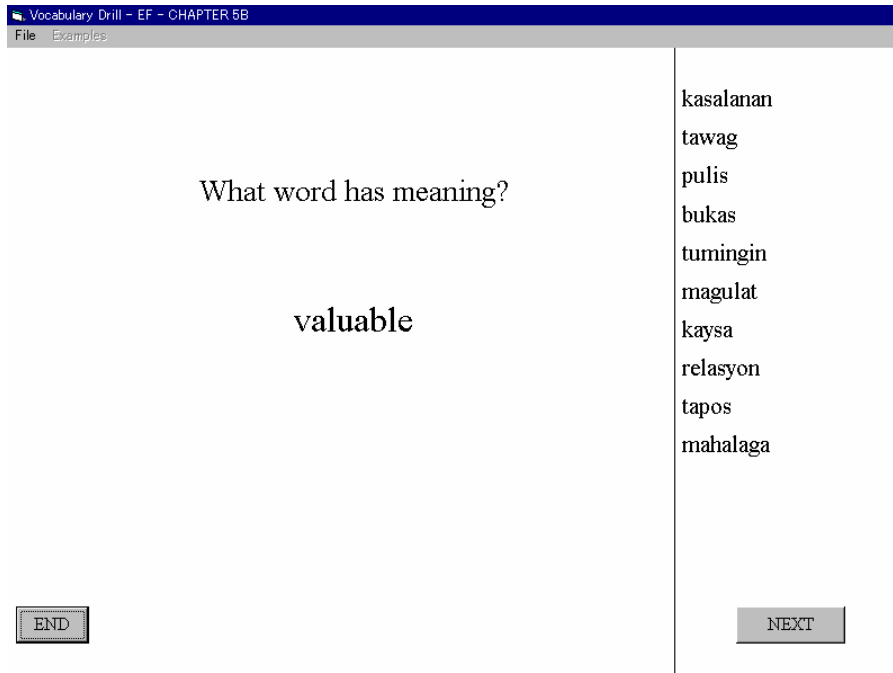


Figure 11

The four games are based on common (at least in the U.S.) children's games, which have been adapted to CAI applications. These are Tictactoe, Anagrams, Hangman, and Concentration. Games are for fun, but they are also for serious language learning, in that they call upon students to employ many of the same language skills they need for communication. Games present a more relaxed, less threatening situation for the students,

such that many students learn more, without even being aware that they are learning. These games are designed for two students to play and study together; they can also be used by one student playing alone.

The first game is *Tictactoe*. In the children's game players alternately place X's and O's in a grid of nine squares (three by three). To win you have to place your X's or O's in three squares in a line (horizontally, vertically, or diagonal). Of course, if both children are expert at the game, neither can ever win.

In our version of the game (Figure 12), you have to answer questions about vocabulary in order to get the squares. The first student selects the square he wants by clicking on it. An English word will appear in that square and ten definitions appear on the right side of the screen. Actually the routine is exactly the same as in the vocabulary drills, except that now the students get rewards for answering correctly.

The student tries to click the correct definition. If he/she is correct, he/she gets that square, and his/her name appears on it. If he/she is wrong, the square remains blank. In either case, the correct glossary entry is displayed on the screen. As you know, the center square is the strongest square as it participates in more winning combinations than any other square. For this reason, if a student selects the center square, he/she must get two items correct to get that square.

Play alternates between the two students until one captures three squares in a line, or until all the squares are taken with no winner.

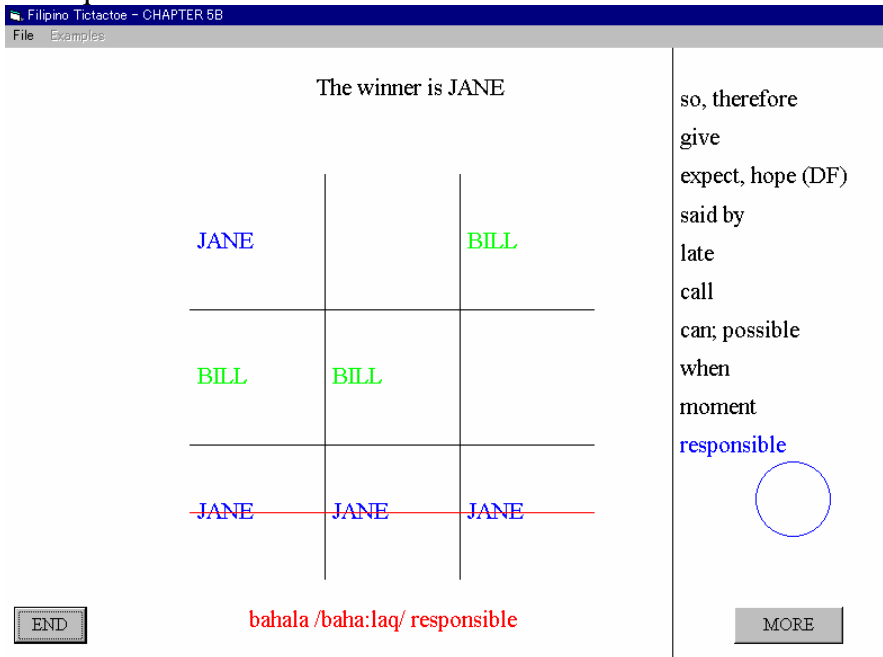


Figure 12

In the children's version of *Anagrams* (a forerunner of SCRABBLE), players get a number of letters at random, and try to make words out of them. In our version (Figure 13), we reverse the procedure. We start with a word, one of the words from the glossary for the selected chapter, and let the computer scramble the order of the letters. For example, if we start with "bahay", the computer might give us "ayhba". The object is to find the original word.

The game consists of ten rounds. In each round the computer randomly selects and scrambles one of the words from the glossary, and presents the scrambled version on the

screen. The score at the beginning is (number of letters - 1) x 15. E.g. for "bahay" the score is (5 - 1) x 15 = 60. With each second that passes, the score is reduced by one.

The first student tries to put the letters in the correct order. If he/she cannot answer correctly within 15 seconds, the first correct letter is shown, and the turn passes to the other student. Play continues until one student gets the correct answer or the score falls to 0. The student answering correctly collects the score remaining on the 'clock'. The complete glossary entry is displayed on the screen. At the end of ten rounds, the student with the higher score is the winner.

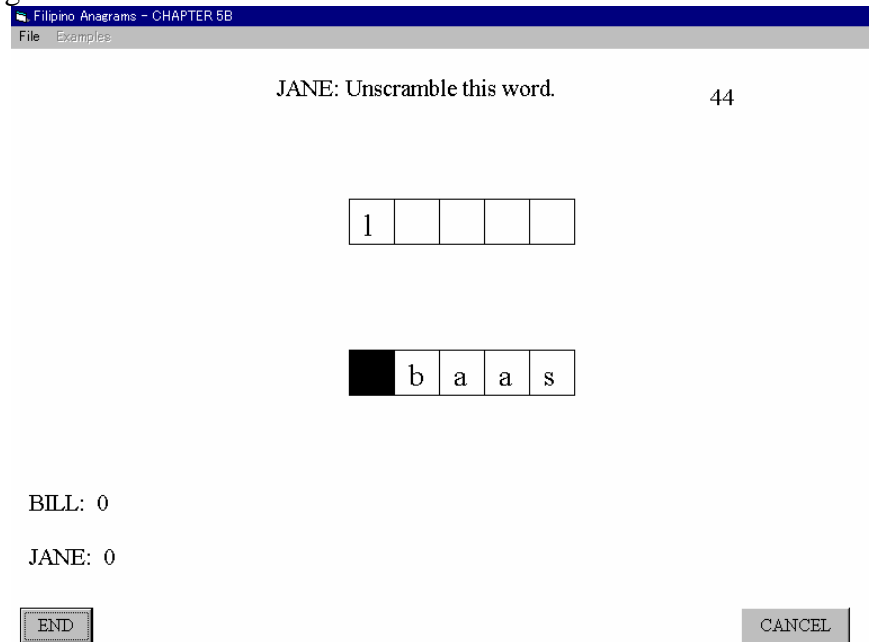


Figure 13

This game strengthens the skill of word recognition, the simple ability to recognize that certain sequences of letters form words that are part of the Filipino language, and of the lessons they are studying.

In the children's game of *Hangman*, one player thinks of a title (book, movie, song, etc.), name, etc., and writes blank spaces (underlined) for the letters in the title or name. For example, he/she might write " _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ " for "Gone with the Wind". That player also draws a hangman's scaffold. The other player tries to guess letters that appear in the title, and finally the whole title. Each time the second player guesses wrong, the first player draws the head or another body part hanging from the 'rope' on the scaffold. If the hanging body is complete (about six misses) before the second player gets the complete answer, the first player wins.

Our version of the game (Figure 14) is much less bloody. No one gets hanged. In fact the two players compete to answer the puzzle, which is one of the sentences from the selected lesson. Another difference is that in the children's game, if the second player asks for a letter that appears several times in the title, the first player has to write all the occurrences of that letter. In the above example, there are two T's. If the second player asked for 'T', the first player would have to write all three T's. Our puzzles, being ordinary English sentences, are much longer (as many as 60 or 70 letters), and contain most of the letters in the alphabet. It is thus too easy if you are given all occurrences of a

given letter. Rather, each time you ask, you'll only see one occurrence. In the above example, you would have to ask for 'T' twice. The third time you asked would be a miss.

The computer selects one sentence from the text, and presents it on the screen as a string of blank spaces (one blank for each letter). The first student selects a letter of the alphabet. If that letter appears in the sentence, it will appear in the proper place in the sentence, and the student selects another letter. If there are more than one occurrence of a given letter in the sentence, only one of those occurrences will be shown. If there are none (or no more) of the selected letter in the sentence, play passes to the other student. Play continues until one student successfully completes the whole sentence.



BILL: Choose a letter.

	a	h		t		s				b		l					u
		y	,		k		w		p		r	i					g
	m							n			o	.					

a	b						g	h	i		k	l	m
n	o	p			r	s	t	u		w		y	

END

NEXT

Figure 14

This game provides practice in a variety of language skills, as well as reinforcing knowledge of the content of the chapter.

The last, and most difficult, game is *Concentration*. The children's version of this game is a card game. The entire deck of 52 cards are dealt out on the table (or floor), face down. One player turns over two cards. If they match (e.g. both aces), he takes those cards, and turns over two more cards. If the cards don't match, they are turned face down again, and the other player turns over two cards. Play continues until all the cards are gone, and the player who has taken the most cards wins.

In the television version of the game, there is a grid of thirty-six (six by six) numbered squares). Under this grid are a number of prizes, each located twice somewhere in the grid. There are also "Wild Cards", "Take A Prize", and "Lose A Prize" squares. One player asks for two squares to be opened, and two prizes or other cards are revealed. If they match, that player gets the prize in question, and two pieces of a puzzle (consisting of letters and pictures) are revealed, and the same player opens two more squares. Play continues until someone can guess the complete puzzle, and only that person can keep the prizes he/she has been capturing.

Our version of the game (Figure 15) resembles the television version. The 'game board' consists of three levels of thirty-six (six by six) squares each. The top level (which appears at the beginning) consists of blank squares. The second level consists of eighteen

items from the glossary, each of which appears twice, scattered randomly about the 'board'. There are no wild cards, etc. The lowest level contains one of the sentences from the selected lesson.

The first student opens (clicks) two squares. Two glossary items (from the second level) will appear on those squares. If the two squares match (they display the same glossary item), then two bits of the sentence at the lowest level will be revealed, and the same student can open two more squares. If the squares don't match, they are closed again, and play passes to the other student.

After each correct match, that student is given the opportunity to solve the puzzle, by filling in the missing parts of the sentence. Play continues until one student successfully completes the sentence at the lowest level.

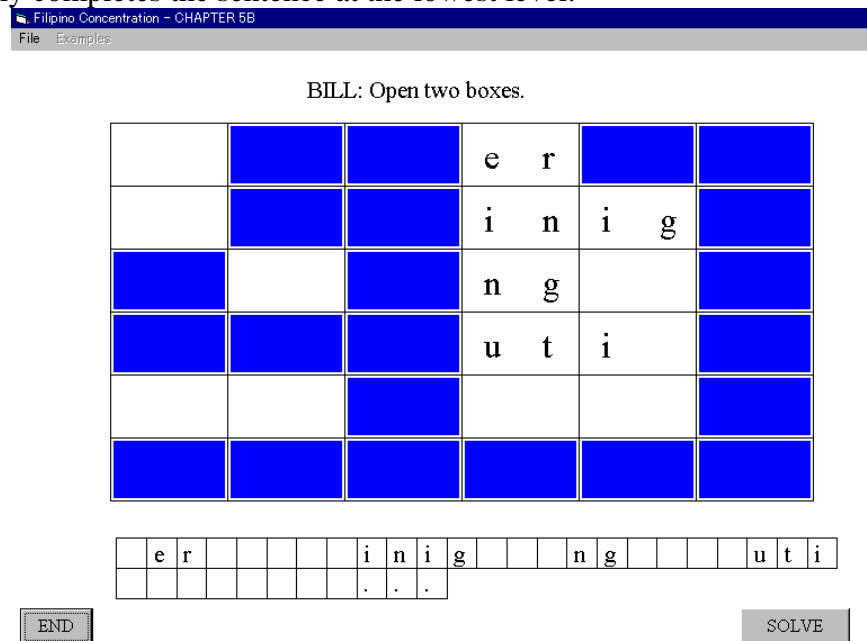


Figure 15

There is no fixed order in which the programs are to be done, nor is it necessary that they all be done. This should be left up to the student. He/she may want to read first, then do the drills and games as a review. Or he/she may want to play first, as a challenge, and go back to the readings and drills later. Similarly, the choice of level should depend on the student's degree of progress. In general, if the drills and games at one level seem too easy, the student should advance to the next lesson. If they seem moderately difficult, he/she should repeat at the same level. If they seem very difficult, he/she may want to go back to an earlier lesson.

The final sub-program is the Glossary Browser (Figure 16). This glossary contains all 2,500 highest-frequency words, arranged in alphabetical order. To see an example of a particular word, the student clicks on that entry. Additional examples (up to 8) can be accessed with the Examples menu.

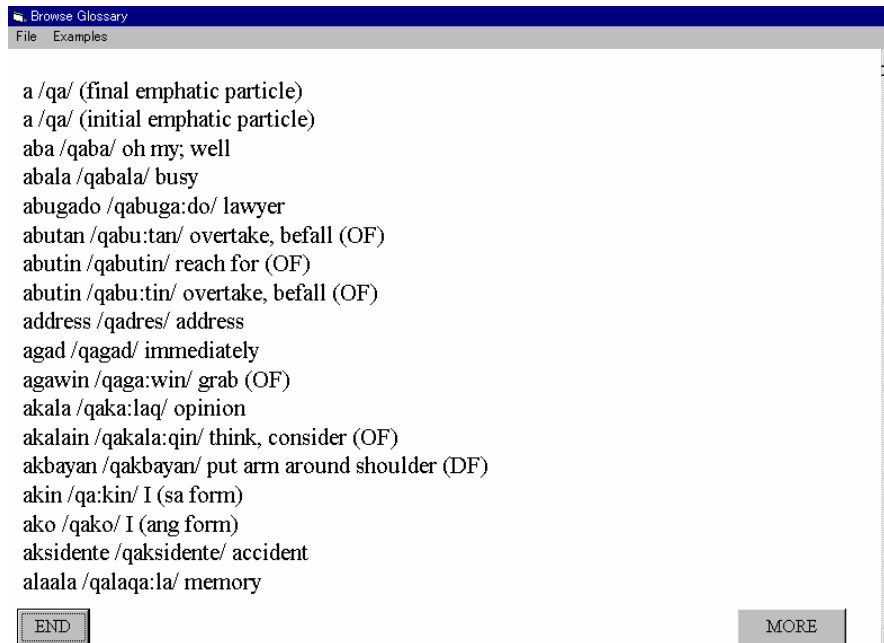


Figure 16

V. Conclusion

The method presented in this paper are adaptable for use with other corpuses and other languages. In fact I also have a version of this program for studying English for science. I would be very eager to work with other scholars in developing similar programs with other corpuses.

APPENDIX A
BOOKS INCLUDED IN THE SURVEY

- Magkahiwalay na Langit*, Veronica Fabian, 1992 (Love Souvenirs -- Scarlet Books)
Huwag Mong Sukatin Ang Puso, Ernildo Bueno, 1992 (Love Souvenirs -- Scarlet Books)
Sa Init ng 'yong mga Yakap, Dodi Anzures, 1992 (Love Souvenirs -- Scarlet Books)
Ikaw Lamang Kahit Kailan, Ervie Nangca-Antonio, 1992 (Love Souvenirs -- Scarlet Books)
- Ano Man Ang Naisin Mo*, Romie Del Moro, 1992 (Love Souvenirs -- Scarlet Books)
Hello... I Love You, Santiago Diokno, 1992 (Love Souvenirs -- Scarlet Books)
Kahit Limos kung Pag-ibig, Loida M. San Diego, 1992 (Diamond Romance -- Emarija Publishing)
- Babasaging Puso*, Maritess M. Armero, 1992 (Diamond Romance -- Emarija Publishing)
Teen Love Triangle, Ysobelle Avila, 1992 (Teen-Age Romances -- Emmica Publishing)
Too Young to Love, Loreta E. Baltazar, 1993 (Teen-Age Romances -- Emmica Publishing)
- Story Book Girl*, Imelda Quijano, 1992 (Teen-Age Romances -- Emmica Publishing)
Unang Gloria, Crisostomo Papa, 1992 (Rosas -- Anvil Publishing)
Ang Lalaki sa Panaginip ni Teresa, Joi Barrios, 1991 (Rosas -- Anvil Publishing)
Di Iidlip sa Magdamag, Joey Papa, 1992 (Rosas -- Anvil Publishing)
Sanlibong Alitaptap, Wilfredo O. Pascual, Jr, 1992 (Rosas -- Anvil Publishing)
Sanga-sangang Landas, Lourdes Vidal, 1992 (Rosas -- Anvil Publishing)
Dahil Mayroong Ikaw, KC Cordero, 1992 (18 Karat Romance -- Pacyno Publishing)
Sa Ngalan Ng Pananagutan, Anya dela Pena, 1993 (Queen of Hearts Romances -- Emmica Publishing)
- Atin Ang Kasalanan*, D. G. Salonga, 1993 (Queen of Hearts Romances -- Emmica Publishing)
- Bad Shot & Good Shot sa Puso mo*, Virginia Casimiro, 1992 (Young Heart -- S.G.E. Printers)
- Pusong Malikot*, Daisy Mendoza, 1992 (Young Heart -- S.G.E. Printers)
Pagluha mo ang hangad ko, Lexer Ann Cruz, 1992 (Love & Romance Stories -- S.G.E. Printers)
- Paalam... Karla Mikaela*, G. Dava-Labarda, 1992 (Love & Romance Stories -- S.G.E. Printers)
- Pilitin Mong Magmahal, Kahit Konti*, Veronica Fabian, 1992 (Moonlight Romance -- Den-Mar Publishing)
- Trahe De Boda*, Jessica Jacinto, 1992 (Moonlight Romance -- Den-Mar Publishing)
Lahat Ng Ito, At Higit Pa!, Estela Rimorin Rosales, 1992 (Moonlight Romance -- Den-Mar Publishing)
- Puso Ko'y Huwag Sugatan*, Sally Eugenio, 1992 (Mr. Cupido -- Den-Mar Publishing)
Kailangan Kita, Lily De Luna, 1992 (Flame -- S.G.E. Printers)
Maaari Ba Akong Mangarap?, Elena M. Patron, 1992 (BMC Executive Homes)
Pagod na Puso, Carla Regala, 1992 (Secrets of the Heart -- Bookware Publishing)
Sa Lilim Ng Pagmamahal, Joy Reyes, 1992 (Purple Shadows -- Harrel Publishing)
Yakap Ng Kahapon Ang Ngayon, Leira Bersamina, 1993 (Valentine Romances -- Books for Pleasure)

Ikaw Pa Rin, Ofelia Concepcion, 1992 (Valentine Romances -- Books for Pleasure)
Gising Tayo Nang Mangarap, Maria Elena Cruz, 1993 (Valentine Romances -- Books for Pleasure)
Paalam, Kaibigan, Gilda Olvidado, 1993 (Valentine Romances -- Books for Pleasure)
Kung Bulag Ang Pag-ibig, Zeny Garcia Villa, 1992 (Valentine Romances -- Books for Pleasure)
Takas Sa Nitso, Rodie Metin, 1983 (Hiwaga Mystery Novels -- Books for Pleasure)
Hiram na Katawan, Arielle, 1992 (Pag-ibig at Kababalaghan -- S.G.E. Printers)
Nagkita Tayo Sa Panaginip, Renato Custodio, 1992 (Pag-ibig at Kababalaghan -- S.G.E. Printers)
May Simula... May Wakas, Jinky E. Jamolin, 1992 (Pag-ibig at Kababalaghan -- S.G.E. Printers)
Sa Bawat Kislap, Arlene de Guzman, 1992 (Pag-ibig at Kababalaghan -- S.G.E. Printers)
Ang Mangingibig ni Corazon, Juvy V. Castillo, 1992 (Pag-ibig at Kababalaghan -- S.G.E. Printers)
Ang Hitman, Roger Nicolas, 1991 (Pinoy Suspense -- Anvil Publishing)
Fantasia, Leo Dee Rogierro, 1992 (Pinoy Suspense -- Anvil Publishing)
Kulay Dugo Ang Rosas, Edgar Reyes, 1991 (Pinoy Suspense -- Anvil Publishing)
Sino?, Lualhati Bautista, 1990 (Pinoy Suspense -- Anvil Publishing)
Lilipulin Ko, Ruben R. Marcelino, 1991 (Pinoy Suspense -- Anvil Publishing)
Ang Pinagpala, Enrique De Jesus, 1990 (Pinoy Suspense -- Anvil Publishing)
Correctional, Wilfredo Pa. Virtusio, 1990 (Pinoy Suspense -- Anvil Publishing)
Three P (Puso, Pera at Pag-ibig), Vic Macapagal, 1990 (Pinoy Krosword)

APPENDIX B

100 MOST FREQUENT FILIPINO WORDS IN THE TWO CORPUSES

(The first figure indicates the raw frequency; the second figure is the frequency after adjustment for distribution within the corpuses.)

1	ang (ang article)	63040	62407.05	51	dalawa 'two'	2321	2216.35
2	sa (sa article)	61374	60688.18	52	ákin 'I (sa form)'	2341	2122.82
3	ng (ng article)	41458	40740.27	53	laláki 'man; male'	2404	2119.33
4	na (linker)	32678	31337.31	54	maging 'become'	2237	2118.57
5	ni (ng article for name)	21700	20850.22	55	gawin 'make, do'	2154	2080.16
6	na 'already'	21139	19670.51	56	lámang 'only, just'	2789	2048.13
7	at 'and'	19309	18747.88	57	iba 'other, different'	2102	2019.39
8	si (ang article for name)	18456	17862.16	58	pára 'so that'	2237	2004.81
9	niya 'he/she (ng form)'	18024	17415.65	59	doon 'there (far)'	2102	1980.52
10	hindí 'no, not'	17322	17051.47	60	káhit 'even if'	1996	1897.36
11	siya 'he/she (ang form)'	16666	16259.61	61	noon 'then, when (past)'	2040	1878.41
12	mga (plural marker)	15080	14704.75	62	talaga 'really'	1973	1863.87
13	pa 'still, yet, even'	11599	11426.24	63	nása 'be at'	1944	1857.71
14	ay (topicalization marker)	12509	11292.08	64	lahat 'all'	1959	1830.82
15	kanya 'he/she (sa form)'	11303	10800.80	65	saríli 'self, own'	1883	1819.87
16	mo 'you (singular ng form)'	10768	10507.96	66	táyo 'we (including addressee -- ang form)'	1920	1779.89
17	ko 'I (ng form)'	11122	10371.72			1983	1706.93
18	lang 'only, just'	8785	8411.88	67	babáe 'woman; female'	1983	1706.93
19	ako 'I (ang form)'	8885	8343.03	68	pala (adverb expressing new knowledge)	1674	1608.44
20	iyon 'that (far -- ang form)'	8423	8088.93			1679	1580.54
21	may 'have; there is'	8317	8051.85	69	bakâ 'maybe'	1679	1580.54
22	ito 'this (ang form)'	8057	7790.90	70	man 'even'	1696	1576.54
23	naman (adverb showing contrast)			71	loob 'inside; heart, mind'	1624	1559.61
		8122	7634.16	72	báhay 'house'	1671	1487.51
24	din 'also'	7293	7203.16	73	iyán 'that (near -- ang form)'	1608	1485.56
25	kung 'if, when'	7276	7136.11	74	áyaw 'not want, not like'	1525	1467.14
26	nang 'when (past)'	7008	6758.10	75	daw 'it is said'	1578	1457.11
27	ka 'you (singular enclitic ang form)'			76	o 'or'	1533	1411.10
		6957	6744.14	77	mata 'eye'	1524	1397.01
28	isa 'one, a'	6782	6481.88	78	táo 'person'	1514	1393.15
29	walâ 'have no; there isn't'	6356	6256.43	79	hábang 'while'	1468	1369.02
30	nito 'this (ng form)'	5864	5488.35	80	kayo 'you (plural -- ang form)'	1444	1352.68
31	ba (interrogative adverb)	5514	5327.81	81	tanong 'question'	1561	1352.14
32	silá 'they (ang form)'	4574	4353.93	82	hò (adverb of respect)	1748	1350.88
33	kay (sa article for name)	4457	4208.12	83	anak 'child (offspring)'	1549	1300.67
34	ngâ 'really'	4270	4135.33	84	kasi 'because'	1435	1282.80
35	péro 'but'	3967	3601.55	85	ninyo 'you (plural -- ng form)'	1401	1279.38
36	nila 'they (ng form)'	3930	3599.94	86	mangyári 'happen'	1377	1277.73
37	ano 'what'	3649	3564.03	87	huwag 'don't'	1332	1267.56
38	sabíhin 'say'	3132	2999.61	88	kailangan 'need'	1284	1222.54
39	pára(-ng) 'like, as if'	3100	2867.15	89	ganoon 'like that (far)'	1294	1218.98
40	alam 'known'	2879	2733.42	90	nátin 'we (including addressee -- ng form)'	1342	1204.18
41	kayâ 'so, therefore'	2793	2707.93			1287	1193.64
42	dáhil 'because'	2824	2611.39	91	marámi 'much, many'	1287	1193.64
43	gusto 'like, want'	2748	2610.13	92	kami 'we (excluding addressee -- ang form)'	1297	1148.74
44	ngayon 'now'	2668	2575.90			1211	1141.95
45	dító 'here'	2681	2554.16	93	mukhá 'face'	1211	1141.95
46	kanila 'they (sa form)'	2550	2379.97	94	búhay 'life'	1191	1124.04
47	bákit 'why'	2449	2355.95	95	bágo 'before'	1151	1115.20
48	sábi 'said'	2601	2281.34	96	hanggang 'until'	1164	1097.10
49	makíta 'see'	2357	2239.61	97	múna 'first'	1156	1083.35
50	iyó 'you (singular -- sa form)'	2339	2230.98	98	lálò 'especially; even more'	1119	1074.05
				99	kita (= ko + ka)	1150	1070.69
				100	pára (sa) 'for'	1196	1052.45

The preceding document was presented at the Tenth International Conference on Austronesian Linguistics (10-ICAL). To properly reference this work, please use the following format:

<LastName>, <FirstName>. 2006. <PaperTitle>. Paper presented at Tenth International Conference on Austronesian Linguistics. 17-20 January 2006. Puerto Princesa City, Palawan, Philippines.
<http://www.sil.org/asia/philippines/ical/papers.html>

For other papers that were presented at 10-ICAL, please visit <http://www.sil.org/asia/philippines/ical/papers.html>.