



## Deseret Language and Linguistic Society Symposium

---

Volume 27 | Issue 1

Article 7

---

3-23-2001

### Lexicostatistics Applied to the Historical Development of Three Languages of the Philippines

Hans Nelson

Follow this and additional works at: <https://scholarsarchive.byu.edu/dlls>

---

#### BYU ScholarsArchive Citation

Nelson, Hans (2001) "Lexicostatistics Applied to the Historical Development of Three Languages of the Philippines," *Deseret Language and Linguistic Society Symposium*: Vol. 27 : Iss. 1 , Article 7.

Available at: <https://scholarsarchive.byu.edu/dlls/vol27/iss1/7>

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Deseret Language and Linguistic Society Symposium by an authorized editor of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

# Lexicostatistics Applied to the Historical Development of Three Languages of the Philippines

Hans Nelson

For years, historical linguists have attempted to classify language family relationships using a variety of methods. One such method is lexicostatistics. "Lexicostatistics . . . is a technique that allows us to determine the degree of relationship between two languages by comparing the vocabularies of the languages and determining the degree of similarity between them" (Crowley 1998, 171). Lexicostatistics is used to determine "(1) time depth (glottochronology), (2) subgrouping, and (3) genetic relationship" (Anttila 1972, 397). The vocabulary used for such comparisons is taken from the Swadesh list of basic vocabulary. This list of 100 English words, developed by Morris Swadesh, is an attempt to formulate a list of lexical items that resist cultural influences and therefore are not easily affected by neighboring languages. Thus, when comparing two languages it is assumed that the words on the Swadesh list have not changed significantly from their original form (Campbell 2000, 177).

Campbell and others challenge this aspect of lexicostatistic methodology, saying that there is no culture-free basic vocabulary in a language. But even though there may not be an impenetrable list of culture-free vocabulary, there are lexical items in a vocabulary that are less likely to have been borrowed from other languages (Anttila 1972, 397). The lexicostatistic approach can, in fact, be useful for subgrouping large language families that are in close geographic proximity with relatively limited lexical data avail-

able for analysis. Thus the lexicostatistic approach is useful for an analysis of the languages among the islanders of the Philippines, an area only about the size of New Mexico, with roughly eighty-five to one hundred known languages. The approach could be effective for at least roughly categorizing and subgrouping these variants.

This paper will use lexicostatistics to look at the historical relations among Tagalog, Ilokano, and Bikolano, three languages of the Philippines, with regard to the wave theory. More specifically, this paper explores whether Tagalog and Ilokano are the most closely related of the three languages being compared. This analysis uses Swadesh's list of 100 basic words. The rate of retention, the rate of cross-linguistic loss, and the dates of divergence among the three languages are also calculated. Additionally, the effectiveness of glottochronology is examined as it pertains to effectively finding relationships within the Austronesian family, more specifically the major dialects in use on the Philippine island of Luzon.

## BACKGROUND

Before proceeding, it will be useful to consider a few notions in greater detail. We will first consider the contrast in the kind of questions in which lexicostatistics and glottochronology are most usefully applied. Glottochronology uses word comparisons between languages for calculating dates of divergence. In doing so, it uses Swadesh's

list of vocabulary. As previously noted, the Swadesh word list is assumed to be a culture-free, or at least culturally resistant, word list that can help determine dates of divergence more accurately. While the vocabulary is resistant to change, it does slowly change, and it does so at a constant rate that gives some idea of how long particular languages have been diverging from each other. This is analogous to Carbon C dating techniques. Because  $^{14}\text{C}$  is radioactive, it has a constant rate of decay, or half-life. The ratio of  $^{14}\text{C}$  to  $^{12}\text{C}$  in a particular object can determine its age. Likewise, glottochronological methods are typically used to show and explore genetic linguistic relationships among languages, meaning that once their relation is determined, a system is set up in relation to time as opposed to geography.

Glottochronologists try to avoid including loanwords in their work. Loanwords, especially from distant, unrelated regions, contaminate this constant rate of retention. "It is very important that you exclude copied (or borrowed) vocabulary . . . as these can make two languages appear to be more closely related to each other than they really are" (Crowley 1998, 175).

In contrast to glottochronology, lexicostatistics is, as Anttila states, "a wider field of statistics in the service of historical vocabulary studies" (1972, 396). We have noted that languages can be described not only in terms of their genetic relation to one another, but also in terms of their geographic relation, or what Comrie calls *areal* relations (1989, 11). The idea of comparing languages by geographic region rather than time depth is termed "wave theory." As noted by one scholar: "When no particular linguistic innovation can be given chronological priority, subgrouping results in a brush-like tree without depth (one node)" (Anttila 1972, 304). And it is in the geographic dimension where lexicostatistic work can be useful. In both lexicostatis-

tics and glottochronological studies, borrowing from languages that are geographically distant (and thus generally linguistically distant or unrelated) diminishes accuracy in determining the relationship between two geographically adjacent languages. However borrowing from neighboring languages is expected and not harmful to determining their degree of relation.

Lexicostatistics proves to be an effective initial grouping strategy for large language families, though when applied glottochronologically, its calculations for dates of divergences admittedly seem arbitrary, without historical evidence to support the figures. Lexicostatistics also proves less a delineator of genetically dated family relations than it does of geographic family relations. Its main function is to determine the degree to which items in more than one lexicon are related. This method will determine the closeness of the three languages we are considering.

It will be useful to further discuss the languages under consideration. Tagalog and Ilokano were among the three languages selected for comparison due to obvious geographic proximity. They coexist with native speakers of each language separated by only tens of miles. After spending time in the Luzon region, and as a fluent speaker of Tagalog, I became vividly aware of the many different languages in such a small area. As one travels the countryside it becomes evident to the astute listener that these languages may well share a branch of the same linguistic tree. Given the geographic proximity of Tagalog and Ilokano, I became curious as to their origins and their relationship with one another. Let it be stated that the three languages chosen, as well as many or even most other languages spoken in the Philippines, are not merely dialects of each other, but truly languages in their own right. At times the languages share similarities in grammar, syntax, and lexicon, yet they remain fully

separated languages and should be treated as such. Originally I had considered using Kapampangan as the third comparison language in this study, but because of limited data resources and time constraints, it was not feasible to do so. However an acceptable replacement, Bikolano, was found. While its geographic area of usage lies slightly south of the area originally envisioned, it is a suitable replacement because it is still geographically adjacent to Tagalog.

As background for what follows, it is necessary to explain how Spanish, Indonesian, and Sanskrit have influenced the three languages under consideration. In 1521, Ferdinand Magellan claimed the land of the Philippines for Spain, whose imperial rule lasted until the United States of America gained possession of the islands after victories in the Spanish-American War of 1898–1901. During the almost four hundred years prior to this, the Spanish had dominated Philippino culture and influenced the language. As to the origins of the Indonesian language influence in the Philippines, Francisco states that Indian influence in the Philippines could be dated to between the tenth and twelfth centuries A.D. on the basis of the linguistic evidences shown in the earliest Old Malay and Old Javanese inscriptions which have been discovered there (1965, xiii). Sanskrit influence entered the Islands by way of the Tamil of the South Indian (Dravidian) culture, which had been in contact with the islands hundreds of years before Indian influences reached the area (Makarenko 1992, 65). Thus, Tagalog, the national language of the Philippines, has been influenced significantly through contact with Sanskrit, Indonesian, and Spanish. Because of these influences, many of Tagalog's original lexical properties have been replaced by loanwords. Not only has Tagalog been influenced, but most other languages in the Philippines have also experienced somewhat parallel modifications. In order for

the vocabulary of Tagalog or other Philippine languages to be compared in a historical context with other languages in close geographical proximity, the vocabularies of the languages must be considered in their preinfluenced state to the degree possible.

## METHODOLOGY

The first step in this study was to remove any borrowed words from languages that are geographically distant. As mentioned earlier, words borrowed from distant languages tend to skew results in determining the degree of relation between two languages. However, if one is seeking to show a geographic relation between languages, then some inclusion of loanwords from geographically neighboring languages is helpful. The next step was to perform a lexicostatistical comparison of the languages. Once that was completed, then the dates of divergence were determined through glottochronological comparisons.

Any distant lexical borrowing involving the Swadesh list of basic vocabulary presents a problem in calculating rates of language divergence and comparing the degree of similarity between the languages under consideration. A study was undertaken by Zorc using the Swadesh-200 word list (a variation based on the original 100-word list) in which he determined the proto-Tagalic stress of vocabulary, yet he failed to remove distantly geographic borrowed words from his list (1972, 13). Therefore his findings in regard to these borrowed words are not credible. In my study, those words that are known to be borrowed from Spanish, Sanskrit, and Bahasa Indonesian were removed.

After identifying a word borrowed from one of these languages, my initial response was an attempt to replace the word with an older form or synonym in the particular language (see Appendix 2). If no word replacement could be found,

then the word was simply dropped from the list. If a particular borrowed word appeared to be more influenced by Philippine languages than by some other language, the word was left in the revised list without modification. Through this process, the original list of 100 words (as shown in Appendix 1) was whittled down to 89 words (see Appendix 3). In Appendix 2, the chart is color coded as follows: white letters on black for Indonesian, black letters on light grey for Sanskrit, and white letters on medium grey for Spanish. Bracketed words are those original borrowed words in the languages from which they were taken. For example, in Appendix 2 the Tagalog, Ilokano, and Bikolano word for *seed* (No. 24) is printed on light grey background; this shows that in all three cases it is borrowed from the Sanskrit language. Next to the Tagalog word *binhi* (No. 24) the Sanskrit word *hiji* from which the Tagalog word is borrowed appears in brackets. Since the Ilokano and the Bikolano words are also printed on light grey and no bracketed word appears next to them, they are also borrowed from the same Sanskrit word shown in the Tagalog column.

In the subsequent discussion of research procedure and findings, the following abbreviations will be utilized. *Tagalog*, *Ilokano*, *Bikolano*, *Spanish*, *Indonesian*, and *Sanskrit* have been abbreviated to *Tag.*, *Ilk.*, *Bik.*, *Spa.*, *Indo.*, and *Sak.* respectively. Of the 100 words forming the original Tag. word list (Appendix 2), there were 4 loanwords from *Spa.*, 13 loanwords from *Indo.*, and 7 loanwords from *Sak* (counting only one of the two words listed for No. 71). Thus, 24 percent of Tag.'s core vocabulary (the 100 original words) is composed of loanwords. In *Ilk.* there were 3 loanwords from *Spa.*, 12 from *Indo.*, and 7 from *Sak*. Therefore, 22 percent of *Ilk.*'s basic vocabulary is composed of loanwords. In *Bik.*, there existed only 1 loanword from *Spa.*, 12 from *Indo.*, and 4 from *Sak.*, totaling 17 percent

of the core vocabulary of the language as borrowed. From this perspective, there appears to be little doubt as to the relative influence from *Spa.*, *Indo.*, and *Sak.* on the core vocabulary of these three Philippine languages. Regardless of total core influence, Indonesian clearly dominates in influence.

In the standard Swadesh word list, words No. 1, *I*, and No. 2, *you*, were found to closely resemble the *Indo.* *aku* and *kau* in all three languages (see Appendices 2 and 3). Yet they resemble and appear to be influenced more by the closer western neighbors of Austronesian than by the more distant Malay East. Thus, they were retained in the list and not treated as loanwords.

With regards to No. 3, *we*, only the inclusive form for each language remained in the list because of the obvious borrowing of the exclusive form in all three languages. The (inclusive) *Bik.* word *kita* resembles the *Indo.* word *kita* yet is retained for comparison of the entire entry row. All three languages must be compared, not just two of the languages. The *Bik.* word No. 4, *ini*, seems to be borrowed from the *Indo.* *ini*. However it was not removed from the list as a borrowed form because it filled the *Bik.* No. 4 slot necessary for there to be a three language comparison of the word. If No. 4, *ini*, were removed from the *Bik.* column, it would mean that both words from *Tag.* and *Ilk.* would have no third word from *Bik.* for comparison. If this were done, the entire entry row would need to be removed. Appendix 5 explains how the other borrowed words listed in Appendix 2 were evaluated, resulting in the revised Swadesh list provided in Appendix 3.

Appendix 3 shows the revised Swadesh list of words that reflects all word removals and replacements for all three languages based on the evaluations in Appendix 5. Words which are shaded light grey and given an asterisk are those words that are still in the list despite

replacement or modification in some fashion from the original Swadesh list in Appendix 1. Words or whole lines which are shaded dark grey are those words or list numbers that have been completely removed from the list of words considered for comparison.

## GLOTTOCHRONOLOGICAL COMPARISON

In order to enable a glottochronological study, I made a comparison of the remaining 89 basic vocabulary word-list items to discover any cognate forms (see Appendix 3). The cognate sets are marked *A*, *B*, or *C*, depending upon the number of different forms of the word in need of representation (see Appendix 4), (Crowley 1998, 176). All Tag. word forms are marked with the letter *A* to show a single word form. If a second form is found in Ilk. unlike the first *A* form, then the letter mark *B* is assigned to denote a new form and so forth with the letter *C* showing an even different form than that of *A* or *B*. If all forms of a particular word in all three languages are cognates, they were all marked with the letter *A*. This process of marking the cognate sets continues for each Swadesh list number until all 89 words are reviewed for their relationship to each other.

## LIMITED LEXICAL DATA

Biko. presented a new problem in the application of glottochronology. Because of its limited lexicon, some words could not be obtained for the basic vocabulary

list, so a justifiable means had to be developed for giving these a fair trial *in absentia*. After much consideration and deliberation, a method was decided upon. There was a total of 14.6 percent of the lexical data missing. These missing words are represented on a background of light grey (see Appendix 4). A system was devised whereby the missing word would receive a letter based upon a percentage corresponding to the ratio order of the others. For example, the number of times in a given combination that the first two words were *A* and the last was *A* was 59.5 percent of the time, and the number of times in a given combination that the first two words were *A* and the last was *B* was 40.5 percent of the total. So, it then follows that if there were 4 of the 13 missing items in which the first two words were marked *A*, then approximately 2 of those statistically should be *A*, and 2 should be given the letter *B*. By this means the 13 missing items were reconciled. Table 1 shows the number breakdown (see also Appendix 4).

## COGNATE PERCENTAGES

With that dilemma resolved, the cognate percentage figures were available for calculation. The cognate percentage calculations shown in Table 2 clearly demonstrate that 46 of the 89, that is, about 52 percent, of the core vocabulary words in Tagalog and Ilokano are cognates. Tagalog and Bikolano also share about 52 percent of their basic vocabulary, and Ilokano and Bikolano share

**Table 1. Combinations**

SL#	Tagalog	Ilokano	Bikolano	%Occurrence	#Missing	Distribution
1	A	A	A	59.5%	4 AA	2A
2	A	A	B	40.5%		2B
3	A	B	A	44.1%	9 AB	4A
4	A	B	B	0.0%		0B
5	A	B	C	55.9%		5C
				<b>Total Missing</b>	<b>13</b>	<b>13</b>

**Table 2. Cognate percentages**

Tagalog			Tagalog		
46 of 89	Ilokano		51.69%	Ilokano	
46 of 89	27 of 89	Bikolano	51.69%	30.34%	Bikolano

about 30 percent of their vocabulary as cognates.

The means for determining cognates, as suggested by Crowley (1998, 178), was primarily based upon the inspection method wherein "intelligent guesswork" was applied in determining whether or not the two forms were cognates. Also, some systematic sound correspondences were used among the languages to determine true cognates. For instance, Blust points out the relationship of the *r* and *g* in the Southern Luzon languages of the Philippines (1991, 73–129). Other sound changes worth mentioning were *d* and *r* along with the standard *l* and *r* changes. The phonological aspect of the comparison undertaken in this study will not be discussed in any further depth, for it was not the primary focus in determining cognate relationships.

## DATES OF DIVERGENCE

As a next step, the glottochronology formula was invoked to calculate the dates of divergence among the languages. The figure of 86 percent was used as the established *r*, or constant change factor, in the mathematical formula to work out the time depth, or the period of separation, of two languages;  $t = \log C / 2 \log r$ ; as given by Campbell (2000, 179). The following findings emerged: (1) Tag. and Ilk. diverged about 2,200 years ago. (2) Tag. and Bik. diverged about 2,200 years ago. (3) Ilk. and Bik. diverged about 4,000 years ago. These findings strongly imply that Tag. and Ilk. are languages of a common family, as are Tag. and Bik. It seems that Bik. diverged from Tag. at about the same time Ilk. diverged from Tag. Both are related more to Tag. than to each other.

## CONCLUSION

Prior to the research, it had been theorized by the investigator that Tagalog would have a greater cognate relationship with Ilokano than it would with Bikolano. This assumption did not hold true. Tagalog seems to have an equal relationship with both of the languages. When the percentages are examined, it appears that the ratio of common core vocabulary is the same for Bik. and Tag. as it is for Ilk. and Tag. The similarity of the relationships between Ilk. and Tag. and Bik. and Tag. could result from the specific list of comparison words chosen or from the partial subjectivity in the determination of cognates. The results are surprising, considering that Bikolano's area of usage is farther south and therefore it is less of a geographic neighbor to Tagalog than is Ilokano.

Concluding anything about subgrouping causality at this point would be presumptuous and not consistent with the limited evidence. However, the notion that Tagalog is a parent to the two other languages appears justified. The fact that it shares 50 percent of its cognates with both of these languages is an interesting finding and a stimulus for further research. Further studies are envisioned which will attempt comparisons of additional languages that the island of Luzon has to offer. These studies will build on the data gathered during the research just completed. A limited review of the literature reveals little existing scholarly work of this type using languages of the Philippines. As to the effectiveness of using glottochronological methodology within the Philippine language cluster, definitive conclusions are premature. Insufficient studies of a simi-

lar type exist with which these results can be compared.

This research seems to demonstrate that glottochronology could be effective in showing core cognate percentages of languages and in turn giving historical linguists a starting point for determining language groupings among families. Relative to Campbell's concerns about the Swadesh list of basic vocabulary, it must be admitted that the 100-word list is not a totally culture-free, unchanging list. However, it can be effectively argued that by removing the borrowed words from the original list, one can form an essentially culture-free word list sufficient to undertake exploratory research. The most important contribution of this study may be the platform and stimulus it provides for additional research applications and modifications of the methodology in the future.

## REFERENCES

- Anttila, Raimo. 1972. *An introduction to historical and comparative linguistics*. New York: Macmillan.
- Blust, Robert. 1991. The greater central Philippines hypothesis. *Oceanic Linguistics* 30 (2): 73–129.
- Campbell, Lyle. 2000. *Historical linguistics: An introduction*. Cambridge: MIT Press.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*. 2nd ed. Chicago: The University of Chicago Press.
- Constantino, Ernesto. 1971. *Ilokano dictionary*. Honolulu, Hawaii: University of Hawaii Press.
- Crowley, Terry. 1998. *An introduction to historical linguistics*. 2nd ed. Auckland, New Zealand: Oxford University Press.
- English, Leo James. 1997. *English-Tagalog dictionary*. 23rd ed. Metro Manila, Philippines: National Book Store, Inc.
- English, Leo James. 1996. *Tagalog-English dictionary*. 12th ed. Metro Manila, Philippines: National Book Store, Inc.
- Francisco, Juan R. 1965. *Indian influences in the Philippines*. Manila, Philippines: Benipayo Printing Co., Inc.
- Grace, George W. 1964. The linguistic evidence. *Current Anthropology* 5 (5): 361–68.
- Llamzon, Teodoro A. 1978. *Handbook of Philippine language groups*. Quezon City, Philippines: The Ateneo de Manila University Press.
- Makarenko, V. A. 1992. South Indian influence on Philippine languages. *Philippine Journal of Linguistics* 23 (1–2): 65–77.
- Rubino, Carl R. Galvez. 1998. *Ilocano*. New York: Hippocrene Books.
- Silverio, Julio F. 1980. *New English-Pilipino-Bicolano dictionary*. Caloocan City, Philippines: National Book Store, Inc.
- Wolff, John U. 1971. *Beginning Indonesian: Part one*. Ithaca, New York: Cornell University.
- Zorc, R. David. 1972. Current and proto Tagalic stress. *Philippine Journal of Linguistics* 3 (1): 43–57.



**Appendix I****Swadesh 100-Word List of Basic Vocabulary**

<b>SL#</b>	<b>English</b>	<b>Tagalog</b>	<b>Ilokano</b>	<b>Bikolano</b>
1	I	ako	siak	ako
2	you (sig.)	ka	sika	ika
3	we (incl.excl)	tayo, kami	datayo, dakami	kita, kata
4	this	ito, iri	daytoy	ini
5	that	iyon, iyon	dayta	idto
6	what	ano	ania	ano
7	who	sino	sino	isay
8	no	hindi, wala	saan, awan	dai, buku
9	all	lahat	amin	gabos
10	many	marami	naruary	orog
11	one	isa	maysa	saro
12	two	dalawa	dua	duwa
13	big	malaki, dakila	dakkel	dakula
14	long	mahaba	atiddog	maigot
15	small	maliit	bassit	saditsadit
16	woman	babae	babai	babaye
17	man	lalaki	lalaki	lalaki
18	person	tao	tao	tawo
19	fish	isda	sida	sira
20	bird	ibon	billit	gamgam
21	dog	aso	aso	ido
22	louse	lisa	lis-a	lisa
23	tree	punong kahoy	kayo	poon
24	seed	binhi	bukel	butud
25	leaf	dahon	bulong	saka
26	root	ugat	ramut	pungo
27	bark	banakal	ukis ti kayo	upak
28	skin	balat	kudil	
29	flesh	balat	lasag	
30	blood	dugo	dara	dago
31	bone	buto	tulang	butud
32	egg	itlog	itlog	sogok
33	grease	grasa	manteka	taba
34	horn	sungay	sara	
35	tail	buntot	ipus	ikog
36	feather	balahibo	dutdot	
37	hair	buhok	buok	buhok
38	head	ulo	ulo	ulo
39	ear	tainga	lapayag	talinga
40	eye	mata	mata	mata
41	nose	ilong	agong	dongo
42	mouth	bibig	ngiwat	ngoso
43	tooth	ngipin	ngipen	ngipon
44	tongue	dila	dila	dila
45	claw	kuko	kuko	kawit
46	foot	paa	saka	bitis
47	knee	tuhod	tumeng	tohod
48	hand	kamay	ima	kamot
49	belly	tiyan	tian	tikab

50	neck	leeg	tengnged	liog
51	breast	suso	suso	daghan
52	heart	puso	puso	
53	liver	atay	dalem	
54	drink	uminom	uminum	
55	eat	kumain	mangan	kumaon
56	bite	kumagat	kagat	kagat
57	see	makita	kita	hilingon
58	hear	making	mangngeg	
59	know	alam	am-ammo	maraman
60	sleep	tumulog	turog	turog
61	die	mamatay	matay	
62	kill	patayin	patayen	
63	swim	lumangoy	aglangoy	
64	fly	lumipad	agtayab	lumpat
65	walk	lumakad	magna	lakaw
66	come	lumipat	umay	dolokon
67	lie	humiga	agidda	kabuwaan
68	sit	umupo	agtugaw	
69	stand	tumayo	agtakder	
70	give	ibigay	ited	bugay
71	say	magsalita, magwika	sarita	tataramon
72	sun	araw	init	aldaw
73	moon	buwan	bulan	bulan
74	star	bituin	bitwen	bituon
75	water	tubig	danum	
76	rain	ulan	tudo	uran
77	stone	bato	bato	bako
78	sand	buhangin	darat	baybay
79	earth	mundo	daga	kinaban
80	cloud	ulap	ulep	ambon
81	smoke	usok, aso	asuk	aso
82	fire	apoy	apuy	kalayo
83	ash	abo	dapo	abo
84	burn	sunog	uramen	pasuon
85	path	daan	dana	agihan
86	mountain	bundok	bantay	bukid
87	red	pula	nalabaga	pula
88	green	berde	berde	
89	yellow	amarilyo	amarilio	amarilio
90	white	maputi	puraw	
91	black	maitim	nangisit	itom
92	night	gabi	rabii	banggi
93	hot	mainit	napudot	
94	cold	malamig	lamek	takig
95	full	puno	napno	pano
96	good	mabuti	naimbag	magayon
97	new	bago	baro	bago
98	round	mabilog	natimbukel	bilog
99	dry	tuyo	namaga	alang-alang
100	name	pangalan	nagan	ngaran

## Appendix 2

## List of Borrowed Words - Color Coded

SL#	English	Tagalog	Ilokano	Bikolano
1	I	ako [aku]	siak	ako
2	you (sig.)	ka [kau]	sika	ika
3	we (incl., excl.)	tayo, kami [kami]	datayo, dakami	kita, kata [kita]
4	this	ito, iri [ini]	daytoy	ini [ini]
5	that	iyan, iyon	dayta	idto
6	what	ano	ania	ano
7	who	sino	sino	isay
8	no	hindi, wala	saan, awan	dai, buku
9	all	lahat	amin	gabos
10	many	marami	naruay	orog
11	one	isa	maysa	saro
12	two	dalawa	dua [dua]	duwa
13	big	malaki, dakila	dakkel	dakula
14	long	mahaba	atiddog	maigot
15	small	maliit	bassit	saditsadit [sedikit]
16	woman	babae	babai	babaye
17	man	lalaki [lalaki]	lalaki	lalaki
18	person	tao	tao	tawo
19	fish	isda	sida	sira
20	bird	ibon	billit	gamgam
21	dog	aso	aso	ido
22	louse	lisa [liksa]	lis-a	lisa
23	tree	punong kahoy	kayo	poon [pohon]
24	seed	binhi [hiji]	bukel	butud
25	leaf	dahon	bulong	saka [sakha]
26	root	ugat	ramut	pungo
27	bark	banakal	ukis ti kayo	upak
28	skin	balat	kudil	
29	flesh	balat	lasag	
30	blood	dugo	dara	dago
31	bone	buto	tulang	butud
32	egg	itlog	itlog	sogok
33	grease	grasa [grasa]	manteka [manteca]	taba
34	horn	sungay	sara [sara]	
35	tail	buntot	ipus	ikog
36	feather	balahibo	dutdot	
37	hair	buhok	buok	buhok
38	head	ulo	ulo	ulo
39	ear	tainga	lapayag	talinga
40	eye	mata [mata]	mata	mata
41	nose	ilong	agong	dongo
42	mouth	bibig	ngiwat	ngoso
43	tooth	ngipin	ngipen	ngipon
44	tongue	dila [lidha]	dila	dila
45	claw	kuko	kuko	kawit
46	foot	paa [pada]	saka [sakha]	bitis
47	knee	tuhod	tumeng	tohod
48	hand	kamay	ima [lima]	kamot
49	belly	tiyan	tian	tikab

50	neck	leeg	tengnged	liog
51	breast	suso	suso	daghan
52	heart	puso	puso	
53	liver	atay [hati]	dalem	
54	drink	uminom [minum]	uminum	
55	eat	kumain	mangan	kumaon
56	bite	kumagat	kagat	kagat
57	see	makita	kita	hilingon
58	hear	makinig	mangngeg	
59	know	alam	am-ammo	maraman
60	sleep	tumulog	turog	turog
61	die	mamatay [mati]	matay	
62	kill	patayin	patayen	
63	swim	lumangoy	aglangoy	
64	fly	lumipad	agtayab	lumpat
65	walk	lumakad	magna	lakaw
66	come	lumipat	umay	dolokon
67	lie	humiga	agidda	kabuwaan
68	sit	umupo	agtugaw	
69	stand	tumayo	agtakder	
70	give	ibigay	ited	bugay
71	say	magsalita, magwika	sarita	tataramon
		[carita] [vaka]		
72	sun	araw	init	aldaw
73	moon	buwan	bulan	bulan
74	star	bituin	bitwen	bituon
75	water	tubig	danum	
76	rain	ulan	tudo	uran
77	stone	bato [batu]	bato	bako
78	sand	buhangin	darat	baybay
79	earth	munido [mundo]	daga	kinaban
80	cloud	ulap	ulep	ambon
81	smoke	usok, aso	asuk	aso
82	fire	apoy [api]	apuy	kalayo
83	ash	abo [abu]	dapo	abo
84	burn	sunog	uramen	pasuon
85	path	daan	dana	agihan
86	mountain	bundok	bantay	bukid
87	red	pula	nalabaga	pula
88	green	berde [verde]	berde	
89	yellow	amarilyo [amarillo]	amarillo	amarillo
90	white	maputi [puti]	puraw	
91	black	maitim	nangisit	itom
92	night	gabi	rabii	banggi
93	hot	mainit	napudot	
94	cold	malamig	lamek	takig
95	full	puno [penuh]	napno	pano
96	good	mabuti [bhuti]	naimbag	magayon
97	new	bago	baro	bago
98	round	mabilog	natimbukel	bilog
99	dry	tuyo	namaga	alang-alang
100	name	pangalan	nagan	ngaran

Spanish
Sanskrit
Indonesian
Original Item

## Appendix 3

## Revised Swadesh List (reflects removals, replacements, and substitutions):

SL#	English	Tagalog	Ilokano	Bikolano
1	I	ako	siak	ako
2	you (sig.)	ka	sika	ika
3	we (incl,excl)	tayo*	datayo*	kita*
4	this	ito*	daytoy	ini
5	that	iyang*	dayta	idto
6	what	ano	ania	ano
7	who	sino	sino	isay
8	no	hindi*	awan*	dai*
9	all	lahat	amin	gabos
10	many	marami	naruay	orog
11	one	isa	maysa	saro
12	two	dalawa	aduwa*	duwa
13	big	dakila*	dakkel	dakula
14	long	mahaba	atiddog	maigot
15	small	maliit	bassit	saday*
16	woman	babae	babai	babaye
17	man	lalaki	lalaki	lalaki
18	person	tao	tao	tawo
19	fish	isda	sida	sira
20	bird	ibon	billit	gamgam
21	dog	aso	aso	ido
22	louse	kuto*	kuto*	kuto*
23	tree	punong kahoy	kayo	poon
24	seed	buto*	bukel	butud
25	leaf	dahon	bulong	saka
26	root	ugat	ramut	pungo
27	bark	banakal	ukis ti kayo	upak
28	skin	balat	kudil	
29	flesh	balat	lasag	
30	blood	dugo	dara	dago
31	bone	buto	tulang	butud
32	egg	itlog	itlog	sogok
33	grease	taba*	taba*	taba
34	horn	sungay	sara	
35	tail	buntot	ipus	ikog
36	feather	balahibo	dutdot	
37	hair	buhok	buok	buhok
38	head	ulo	ulo	ulo
39	ear	tainga	lapayag	talinga
40	eye	mata	mata	mata
41	nose	ilong	agong	dongo
42	mouth	bibig	ngiwat	ngoso
43	tooth	ngipin	ngipen	ngipon
44	tongue	dila	dila	dila
45	claw	kuko	kuko	kawit
46	foot	paa	saka	bitis
47	knee	tuhod	tumeng	tohod
48	hand	kamay	ima	kamot
49	belly	tiyan	tian	tikab

50	neck	leeg	tengnged	liog
51	breast	suso	suso	daghan
52	heart	puso	puso	
53	liver	atay	dalem	
54	drink	uminom	uminum	
55	eat	kumain	mangan	kumaon
56	bite	kumagat	kagat	kagat
57	see	makita	kita	hilingon
58	hear	makinig	mangngeg	
59	know	alam	am-ammo	maraman
60	sleep	tumulog	turog	turog
61	die	yumao*	matay	
62	kill	patayin	patayen	
63	swim	lumangoy	aglangoy	
64	fly	lumipad	agtayab	lumpat
65	walk	lumakad	magna	lakaw
66	come	lumipat	umay	dolokon
67	lie	humiga	agidda	kabuwaan
68	sit	umupo	agtugaw	
69	stand	tumayo	agtakder	
70	give	ibigay	ited	bugay
71	say	magsabi*	kuna*	tataramon
72	sun	araw	init	aldaw
73	moon	buwan	bulan	bulan
74	star	bituin	bitwen	bituon
75	water	tubig	danum	
76	rain	ulan	tudo	uran
77	stone	bato	bato	bako
78	sand	buhangin	darat	baybay
79	earth	daigdig*	daga	kinaban
80	cloud	ulap	ulep	ambon
81	smoke	usok, aso	asuk	aso
82	fire	apoy	apuy	kalayo
83	ash	abo	dapo	abo
84	burn	sunog	uramen	pasuon
85	path	daan	dana	agihan
86	mountain	bundok	bantay	bukid
87	red	pula	nalabaga	pula
88	green	lunti*	nalangto*	
89	yellow	dilaw*	kiaw*	darag*
90	white	maputi	puraw	
91	black	maitim	nangisit	itom
92	night	gabi	rabii	banggi
93	hot	mainit	napudot	
94	cold	malamig	lamek	takig
95	full	puno	napno	pano
96	good	maganda*	naimbag	magayon
97	new	bago	baro	bago
98	round	mabilog	natimbukel	bilog
99	dry	tuyo	namaga	alang-alang
100	name	pangalan	nagan	ngaran

Altered Item\*

Removed Item

**Appendix 4****Cognate List:**

<b>SL#</b>	<b>English</b>	<b>Tagalog</b>	<b>Ilokano</b>	<b>Bikolano</b>
1	I	A	A	A
2	you (sig.)	A	A	A
3	we (incl.)	A	A	B
4	this	A	B	C
5	that	A	B	C
6	what	A	A	A
7	who	A	A	B
8	no	A	B	C
9	all	A	B	C
10	many	A	A	B
11	one	A	A	B
12	two	A	A	A
13	big	A	A	A
14	long	A	B	C
15	small	A	B	C
16	woman	A	A	A
18	person	A	A	A
19	fish	A	A	B
20	bird	A	B	C
21	dog	A	A	B
22	louse	A	A	A
24	seed	A	B	A
25	leaf	A	B	A
26	root	A	B	C
27	bark	A	B	C
28	skin	A	B	A
29	flesh	A	B	A
30	blood	A	A	A
31	bone	A	B	A
32	egg	A	A	B
33	grease	A	A	A
35	tail	A	B	C
36	feather	A	B	A
37	hair	A	A	A
38	head	A	A	A
39	ear	A	B	A
40	eye	A	A	A
41	nose	A	A	A
42	mouth	A	B	C
43	tooth	A	A	A
45	claw	A	A	B
47	knee	A	B	A
48	hand	A	B	A
49	belly	A	A	B
50	neck	A	B	A
51	breast	A	A	B
52	heart	A	A	A
55	eat	A	B	A
56	bite	A	A	A

57	see	A	A	B
58	hear	A	A	A
59	know	A	A	A
60	sleep	A	A	A
62	kill	A	A	B
63	swim	A	A	B
64	fly	A	B	A
65	walk	A	B	A
66	come	A	B	C
67	lie	A	A	B
68	sit	A	B	C
69	stand	A	B	C
70	give	A	B	A
71	say	A	B	C
72	sun	A	B	A
73	moon	A	A	A
74	star	A	A	A
75	water	A	B	C
76	rain	A	B	A
77	stone	A	A	A
78	sand	A	B	C
79	earth	A	B	C
80	cloud	A	A	B
81	smoke	A	A	A
82	fire	A	A	B
84	burn	A	B	C
85	path	A	A	A
86	mountain	A	B	C
87	red	A	B	A
88	green	A	B	C
89	yellow	A	A	B
91	black	A	B	A
92	night	A	A	B
93	hot	A	B	C
94	cold	A	A	B
96	good	A	B	C
97	new	A	A	A
98	round	A	B	A
99	dry	A	B	C
100	name	A	A	A

Total 89

Calculated Letter



## APPENDIX 5

With list word No. 4, the Indo. word *ini* was removed in Tag. because a more suitable replacement was found using the word *ito*. List word No. 8, *no*, meaning negative, has been used in all three languages, and the *no*, meaning "have not any," or "none," or "without" has been removed. Standard word No. 12, the word for "two," shows contact with India; however the word has been derived from the Austronesian *duwa* (cf. Francisco 1965, 44). The Ilk. form has been changed to reflect a dialect pronunciation of the word expressing a more suitable Austronesian sound.

In the Tag. list word No. 13, *malaki*, has been removed, for it resembles too closely the word for "man" in Indo., and thus is likely borrowed. Of the two words originally listed as possible Tag. words with the meaning appropriate for No. 13, the more Tag. word *dakila* was left. Word No. 15, the Bik. *saditsadit*, which resembled the Indo. *sedikit*, was removed and replaced with Bik. *saday*, also meaning "smallness," leaving its original meaning intact. Word No. 17, *lalaki*, in all three languages has been totally removed from the list in all three languages because it was borrowed from the Indo. word *lalaki*. The standard list word No. 22 in all three languages has been replaced with the older more mature meaning of the word for "louse," *kuto*, versus the borrowed young immature meaning of "louse," *lisa*, taken from Sak. *liksa*.

Word No. 23 has been removed from the list in all three languages because the Bik. *poon* is borrowed from the Indo. *pohon* and the Tag. *punong kahoy*, meaning "tree of wood," with the root *puno* meaning "tree," was also borrowed. This leaves only the Ilk. *kayo* left in the standard list that is not borrowed. Having no other language with which to compare the Ilk. form, standard list No. 23 was removed. Tag. No. 24, *binhi*, borrowed from Sak. *hiji*, was replaced with another

form, *buto*, meaning "bone" and "seed." Bik. word No. 25, *saka*, borrowed from Sak. *sakha*, has been removed as no suitable replacement was found. Tag. word No. 33, *grasa*, borrowed from Spa. *grasa*, has been replaced by the Tag. *taba*, meaning "fat" in both Tag. and Ilk. The Ilk. word *manteka*, borrowed from *manteca*, meaning "butter," has been replaced by *taba*.

Standard list word No. 34 has been totally removed from the list in all three languages because of the borrowing of the Ilk. *sara* from Sak. *sara*, meaning "horn." This left only one word in Tag. and none in the Ilk. or the Bik. list. With nothing remaining with which to compare the Tag. word, No. 34 was removed in all three languages. Standard word No. 44 has been totally removed in all three languages because *dila* is borrowed from the Sak. *lidha*. The entry row for standard word No. 46 has been removed since both the Tag. and Ilk. words are borrowed from Sak. Standard word No. 48, Ilk. *ima*, possibly borrowed from Sak. *lima*, meaning "five," has been left even though it could be a borrowed word (one possible argument for it as a Sak. borrowing is that it could be that the Sak. word meaning "five" became a representation of the word for "hand," which consists of five fingers).

The entry row for word No. 53 has been removed from the list because the Tag. *atay* is borrowed from Indo. *hati*, thus leaving only one word for comparison. Word No. 54 is borrowed, in both Tag. and Ilk. from Indo. *minum*, and the entry row has therefore been removed from the list. Tag. word No. 61, *mamatay*, is borrowed from Indo. *mati* and was replaced with *yumao*, meaning to "pass away," yet no replacement could be found for the Ilk. form, which was also borrowed from Indo. This left insufficient data for comparing the three languages. Because of this, No. 61 in both the Tag. and Ilk. languages was removed.

Tag. No. 71 *magsalita* and *magwika* were borrowed from Sak. *carita* and *vaka*. The Ilk. *sarita* was also borrowed from Sak. They have been replaced with Tag. *magsabi* and Ilk. *kuna*. Standard list word No. 77 appears to be borrowed from Indo. *batu*, yet it was not removed in all three languages for the same rationale as applied in the case of word No. 12. Tag. word No. 79, *mundo*, was borrowed from Spa. *mundo* and was replaced with Tag. *daigdig*, an older Tag. word also meaning "world" or "earth." The Tag. *apoy* and Ilk. *apuy* words for No. 82 appear similar to the Indo. *api*, yet they will not be considered to be truly borrowed from the Indo. language. In all three languages, No. 83 is removed from the list because all forms are borrowed from the Indo. *abu*.

Tag. word No. 88, *berde*, is borrowed from Spa. *verde*, meaning "green" and was replaced with an older Tag. form *lunti*. The Ilk. *berde*, borrowed from Spa., was replaced with *nalangto*, also an older synonym for the word. All three languages use word No. 89 as *amarilyo* or *amarilio*. Both are borrowed from Spa. *amarillo*, meaning "light grey." They were replaced with the Tag. *dilaw*, the Ilk. *kiaw*, and the Bik. *darag*. No. 90 was removed from the list in all three languages. The Tag. *maputi* is believed to be borrowed from Sak. *pudi*, meaning "honorable, pure, virgin" and was thus replaced. This left only one word for comparison. No. 95 has been removed from the list in all three languages because it is borrowed from Indo. *penuh*, meaning "full." Tag. No. 96, *mabuti*, is borrowed from Sak. *bhuti* and was therefore replaced with the Tag. word *maganda*, which means "beautiful."