A CLASSIFICATION OF TWENTY-NINE CONTEMPORARY PHILIPPINE LANGUAGES

Jessie Grace U. Rubrico
University of the Philippines
April 2001

1. Introduction

Historical linguistics is the branch of linguistics that focuses on the interconnections between different languages in the world and/or their historical development. Historical linguists investigate how languages evolve and change through time, how multiple "offspring" languages can arise from one past "parent" language, and how cultural contact between speakers of different languages can influence language development and evolution (Lehmann,1992). Some forty years ago, Douglas Chretien of the University of California at Berkeley, attempted to classify twenty-one of the then-contemporary Philippine languages on the basis of shared vocabulary Chretien (1961). Extracting data from a 1953 Composite Vocabulary of Philippine Languages published by the Institute of National Language and applying statistical tools on these, Chretien carefully built up a case for an argument that the Philippine languages could be categorized as belonging to groups or "families" of languages that he termed the "Luzon Sequence," the "Macro-Bisayan Group," and the "Mindanao-Sulu Group." He also identified certain "transition" languages like Bikol and Hiligaynon that lay between the groups, and less-precise labels such as "climax" and "marginal" languages. Curiously, he had Tagalog stand apart as a language, not quite sure what to do with it but arguing against the findings of both Conklin and Dyen. Finally, he concluded that his method of "non-genetic classification" brought out the commonality of the disparate Philippine languages as a continuing process worth investigating further.

Twenty years later after Chretien's work, a seminal work of reconstruction of Proto-Philippine phonemes and a dictionary of proto-morphemes was accomplished by a Filipino linguist (Paz, 1981). Defending her use of the comparative method which was propositioned by Dyen, Paz posits that the phonemes and morphemes she arrived at are proto-forms belonging to a parent language(s) in the obscure past of Philippine history, records of which were extant prior to the 16th century. Their manifestations in what are *now* (circa 1981) contemporary Philippine languages is directly traceable to this ancestor language known as PP. These proto-forms reconstructed by Paz, particularly the morphemes, would be an ideal "jump-off" point in validating Chretien's classification since there is a gap between them of something like twenty years within which the languages were allowed to evolve or develop. In other words, one has a scholarly curiosity on what *actually* happened to Chretien's language groupings in the light of Paz' findings.

2. Problem, Scope, and Limitations

Recalling Chretien's earlier work on language classification, he statistically arrived at a basic list of 1,904 morphemes describing every-day Philippine life, and with their distributions over the 21 languages contemporary in his time (the mid-1950s). On the other hand, Paz' work provides us with a valuable dictionary of proto-morphemes containing 396 cognate sets. Each set in turns contains, on the average, about 8 to 13 equivalent morphemes spread in some 9 to 11 languages. Paz likewise chose only 29 Philippine languages out of the more than "300 languages and dialects" (quoting Constantino, 1968), as her representative list, given the close similarities and resemblances of these languages to each other.

The present paper is an indicative study of whether the morphemes derived from 29 contemporary Philippine languages (circa 1981) would result to a *new* grouping, or, subgroupings of Philippine languages, on the basis of shared innovations. It is indicative in the sense that only a 12 per cent sample size, or 46 cognate sets out of a total of 396 sets, are used, resulting to an array of 599 morphemes distributed over these 29 languages. The selection, however, of these 46 cognate sets is by random sampling, as will be described more fully below. It is also indicative in that the writer does not make any definite assertions other than to suggest patterns of language aggrupations in the light of new data (viz., Paz, 1981).

One asks the question: do the language groupings of Chretien still hold today when they were first proposed 40 years ago? If yes, then Chretien's hypothesis is probably correct. If no, what are the possible new aggrupations? This paper will attempt to provide answers to these questions.

Given Chretien's mass of data (23,000 words) and the length of time (one year) completing his study under a full scholarship grant, Chretien essentially did a good job of statistically categorizing the languages of the Philippines and same will be utilized here. Paz' dictionary of proto-morphemes is an invaluable source for this present study. Due to time and other constraints, the writer limited her own data to a small sample. She simply used random sampling, to wit: Go over the proto-morpheme dictionary and select every eight cognate sets. In this manner, she was able to get a sample size of 46 cognate sets, or, 12% of the total. By inspection, she determined whether the selected sets were descriptions of everyday life in the Philippines (example:'abága' shoulder and 'láksut' jump) and if doubtful, replace this set with an adjacent set that was. The next step was to array the individual 599 morphemes found in the cognate sets according to language, resulting in a frequency distribution table found in Tables I, I-b, and II, following pages.

3. Tabulating the Data. Initially, a smaller sample of 26 cognate sets was assembled by the method described above and its 410 morphemes arrayed as shown in Table I-b (below). The purpose of this array was to benchmark the emerging pattern (if any) when the sample got bigger, as in Table I. The first tabulation is labeled "Table I-b" and the next "Table I" to show the primacy of Table I over Table I-b as basis for interpretation. Once the data from 26 cognate sets were arrayed by frequency distribution to the respective Philippine languages, individual

percentages to total /language were computed in order to reveal the pattern. Next, the tabulation was continued, this time to include a bigger sample of 46 cognate sets. The results are shown in Table I. Last, since *within* the cognate sets Paz already distributed the morphemes among sub-groups of languages (example: 'sáluk' bamboo dipper common to both Tagalog and Bukidnon), it was a simple matter for this writer to array said morphemes according to language groups identified in Paz' dictionary.

Table I
Indicative Distribution of Morphemes Over 29 Contemporary Philippine Languages
(Base: 46 Cognate Sample Sets)

No.	Contemporary Philippine Language	Abbreviation	No. of Morphemes	Per Cent
1	Tagalog	Tag.	44	7.5 %
2	Sebuano	Seb.	35	6.5
3	Naga Bikol	Nag.	33	6.0
4	Virac	Vir.	31	5.5
5	Iloko	Ilk.	31	5.5
6	Aklanon	Akl.	31	5.5
7	Pangasinan	Png.	28	5.0
8	Kapangpangan	Kap.	25	4.0
9	Waray	War.	25	4.0
10	Tausog	Tau.	26	4.0
11	Agutaynon	Agt.	20	3.5
12	Bukidnon	Buk.	19	3.5
13	Tagbanwa	Tbw.	19	3.5
14	Bahi	Bah.	18	3.0
15	Iba	Iba	18	3.0
16	Itawis	Itw.	18	3.0
17	Kamalignon	Kam.	18	3.0
18	Maranao	Mar.	18	3.0
19	Yakan	Yak.	16	2.5
20	Subanen	Sub.	16	2.5
21	Isinay	Isi.	16	2.5
22	Ilongot	Igt.	14	2.0
23	Ibanag	Ibg.	13	2.0
24	Bontoc	Bon.	13	2.0
25	Buhid	Buh.	13	2.0
26	Itbayat	Itb.	13	2.0
27	Blaan	Bla.	10	1.5
28	Bagobo	Bag.	9	1.5
29	Kalinga	Kal.	9	1.5
	Totals		**599**	**100.0%**

Table I-b
Sample Distribution of 363 Morphemes by Philippine Languages
(Base: 26 Cognate Sample Sets)

No.	Contemporary Philippine Language	Abbrevia-tion	No. of Morphemes	Per Cent
1	Tagalog	Tag.	27	7.0 %
2	Sebuano	Seb.	21	5.0
3	Naga Bikol	Nag.	18	4.0
4	Virac	Vir.	19	5.0
5	Iloko	Ilk.	18	4.0
6	Aklanon	Akl.	22	5.5
7	Pangasinan	Png.	20	5.0
8	Kapangpangan	Kap.	18	4.0
9	Waray	War.	19	5.0
10	Tausog	Tau.	18	4.0
11	Agutaynon	Agt.	13	3.0
12	Bukidnon	Buk.	13	3.0
13	Tagbanwa	Tbw.	13	3.0
14	Bahi	Bah.	12	3.0
15	Iba	Iba	16	4.0
16	Itawis	Itw.	11	2.5
17	Kamalignon	Kam.	9	2.5
18	Maranao	Mar.	12	3.0
19	Yakan	Yak.	14	3.0
20	Subanen	Sub.	12	3.0
21	Isinay	Isi.	11	2.5
22	Ilongot	Igt.	6	1.5
23	Ibanag	Ibg.	9	2.5
24	Bontoc	Bon.	12	3.0
25	Buhid	Buh.	10	2.5
26	Itbayat	Itb.	8	2.5
27	Blaan	Bla.	6	1.5
28	Bagobo	Bag.	13	1.5
29	Kalinga	Kal.	10	2.5
	Totals		410	100.0%

- Values expressed in italics in above Table I-b are samples as yet and are not yet indicative of the possible classification of Philippine languages.

Table II is the result of such arraying whereby like morphemes are lumped according to groups of languages where they are found, in combinations of two languages, three languages, and so on all the way to twenty-six languages.

Table II
Distribution of Shared Morphemes by Number of Philippine Languages

Number of Philippine Languages	Distribution of Morphemes	Per Cent to Total	Cumulative Percentage
1	-	-	
2	84	14.0%	14.0%
3	63	11.0	25.0
4	40	7.0	32.0
5	33	5.5	37.5
6	30	5.0	42.5
7	28	4.6	47.1
8	16	2.7	49.8
9	41	6.8	56.6
10	48	8.0	64.6
11	64	10.7	75.3
12	9	1.2	76.5
13	11	1.5	78.0
14	13	2.2	80.2
16	12	2.0	82.2
18	20	3.3	85.5
20	20	3.3	88.8
22	21	3.5	92.3
26	46	7.7	100.0%
Totals	**599**	**100.0%**	

* The gaps in the no. of morphemes shared by a number of Phil. languages (15, 17, 19, 21 to 25, and 27 to 29 above) could indicate discrete data and is the result of random sampling procedures followed.

It should be emphasized at this point that, a, say, two-language combination refers to *any* two languages, not a particular pair of languages such as was done by Chretien. This observation will be dealt with in the next section.

4. Interpreting the Data

A number of observations is apparent or can be deduced from these tables. Let us take Table I-b first, as follows:

1. The numerical values are expressed in italics because they are sample values and are not yet indicative of the possible classification of Philippine languages.

2. Nevertheless, the values show a possible *trend* or tendency on the direction of the distribution of morphemes by language. Please refer to the apparent *dominant* positions of the top 10 major languages occupying the top slots (Tag, Seb, Naga Bikol, Virac, Iloko, Aklanon, Pangasinan, Kapangpangan [or Kapampangan], Waray, and Tausug) ranging from 7.0% to 4.0% of the base total of 410 morphemes. The rest of the "minor" languages, from Agutaynon (3%) to Bagobo (1.5%), occupy the lower positions of the table and have a diminished importance vis-à-vis morphemes shared by the dominant languages.

Why has not this writer followed Chretien's *correspondence pair* in comparing languages? Why did she not compare a language to another in pairs? Aside from its extreme tediousness (there are 406 paired combinations = [29 x 28]/2) not within the scope of this paper, the more important reason is : *it is not necessary.* Why so? Even a cursory inspection of Table I-b (later on validated by Table I) reveals that, unlike Chretien's findings in 1957 whereby the languages are more heterogeneous among clusters, there is now (in 1981) a marked tendency of these languages to be more *homogenous* and to cluster around the so-called dominant or major languages.

Table I clearly indicates the clustering tendency. With a bigger sample size of cognate sets, Tagalog has now 7.5% (from 7%) of the morphemes arrayed, Sebuano 6.5% (from 5%), Naga Bikol 6% (from 4%), and so on. The minor languages have their percentages remaining essentially the same. Given that, this paper is now ready to state the following:

1. There is a *definite* indication on the probable distribution of the 29 languages, the major ones increasing their dominance over the minor ones. These are shown in the big plusses, or, in remaining the same in their respective percentages. Compare the Tag-Seb-Nag-Vir-Ilk-Akl-Png-Kap-War group with, say, the Ibg-Bon-Buh-Itb group.

2. Improved commerce, communication, transportation, and urban living are probably the reasons for the shift from multi-lingualism in the different Philippine regions to language commonality limited to a few major ones like Tag. and Seb.

The third table, Table II, is even more indicative. It shows the manner by which the 599 morphemes are distributed among the language combinations. For instance, 14% of all morphemes studied are shared by only *two* languages, while almost 8% are shared by 26 languages combinations. As expected, there is no universal morpheme shared by all 29. This

much is apparent from the dictionary of Paz. What are these language combinations? The chart below was constructed on the basis of how the morphemes appeared in frequency, thus:

Lang. Comb.	Languages
2	- Tag, Seb
3	- Tag,Seb,Nag
4	- Tag,Seb,Nag,Vir
5	- Tag,Seb,Nag,Vir,Ilk
6	- Tag,Seb,Nag,Vir,Ilk,Akl
7	- Tag,Seb,Nag,Vir,Ilk,Akl,Png (or Kap, or War)
8, 9, or 10	- Tag,Seb,Nag,Vir,Ilk,Akl,Png or Kap or War, Tau

8

Beyond 10 language combinations, what language combination follows become merely speculative because of the writer's limited data. Nevertheless, even by inspection of the raw data (not shown here), it is quite apparent Tagalog, Sebuano, Naga, Virac, Iloko, Aklanon, Waray, Pangasinan, Kapampangan, and Tausug share the most number of morphemes.

Cumulative percentages of morphemes shared to total by number of languages are also computed in the last column of Table II. What do these cumulative figures mean? It simply means that, as the number of language combinations increase, a larger and larger number of morphemes are also being shared , *but up to a certain point only.* To illustrate, two languages (Tag and Seb) share the most morphemes among any two languages (14%). Going from two to eight languages, the cumulative percentage is now 49.8%. In other words, these eight language combinations cumulatively have *half* (50%) the morphemes studied, leaving the rest of the other 21 languages sharing the rest. Comparing this to the findings of Chretien where he said "…clearly the languages are greatly differentiated one from the other.." this writer on the contrary finds much uniformity in the dominant languages. The point at which the languages begin to differentiate from each other is when one starts to include the exclusive morphemes of the minor languages (say, Itb) to the whole group. Nevertheless, these morphemes are not as numerous as those of the dominant groups.

5. Conclusions : Suggested Language Groupings and Implications

What implications are there in the findings discussed above?   In the first place, the language groupings propositioned by Chretien, the so-called "Luzon Sequence," "Macro-Bisayan," and "Mindanao-Sulu" so rigorously arrived at by *k*-values computations *can no longer hold water*. The mathematical values of higher or lower coefficients computed by Chretien and which would link related languages or discriminate them, were, by his own admission "… not entirely amenable to the procedure outlined above.." that is why he had to introduce Hiligaynon, Kinaray-a, and Kuyonon as aberrant languages not belonging to any group under his methodology. Likewise for Tagalog, which he left hanging and uncategorized.

Fortunately for the case for Philippine historical linguistics, new data have come to us by way of the pioneering work of Paz on Philippine proto-forms. Granted that the underlying assumptions on language classification are correct, namely:

     a. Common retention from prototype language;
     b. Common innovation subsequent to proto-language;
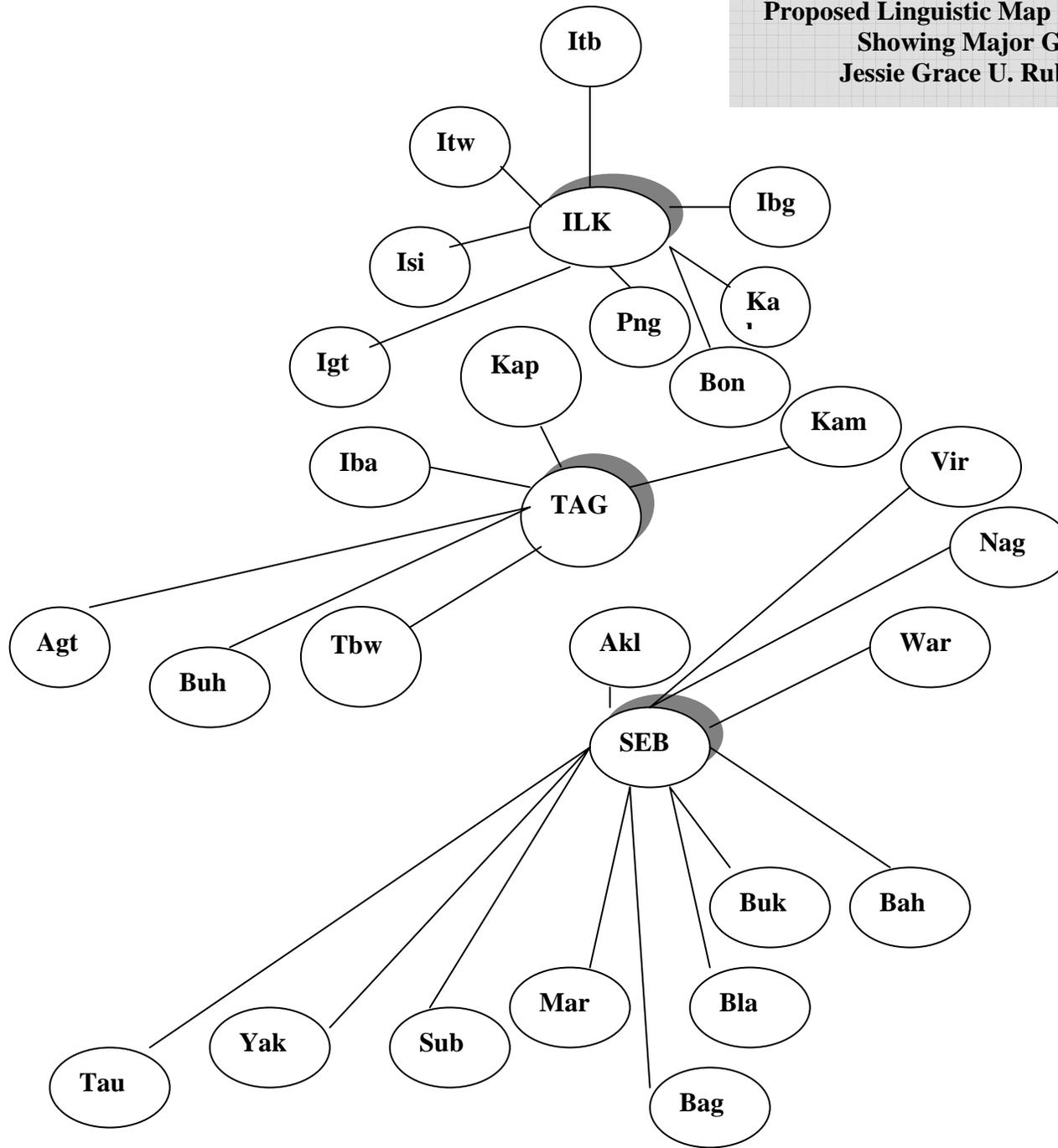   c. Borrowings from one language to another and vice-versa;
     d. Convergence

we may surmise that all of these factors contributed, in varying degrees, to the development of new language aggrupations contemporary to our times. Hence, this writer would like to suggest the following new groupings or sub-groupings as supported by new data, to wit:
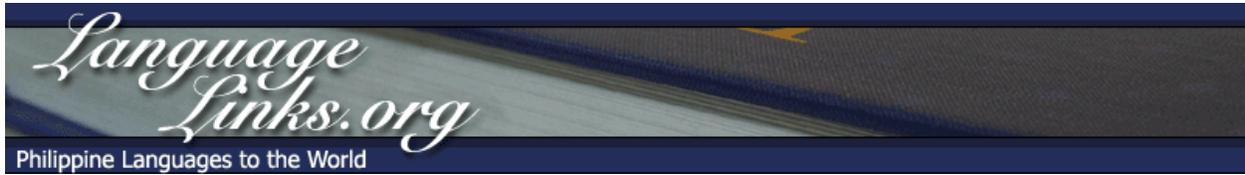
A. The *Greater Ilokano* Group - the possible "Ilokanization" of Itb, Itb, Isi, Kal, Bon, Igt and Png into one language known as "Greater Ilokano;"

B. The *Greater Tagalog* Group - the possible "tagalization" of Kap, Agt, Buh, Iba, Tbw, and Kam into one language known as "Greater Tagalog;"

C. The *Greater Sebuano* Group - the possible "sebuanization" of Nag, Vir, Akl, Mar, Buk, Bah, Bla, Bag, Sub, Yak, and Tau into one language known as "Greater Sebuano."

And beyond that is an even greater convergence. Already, there are indications that this seeming convergence of the Philippine family of languages into one language known as *Filipino* is fast becoming a reality. Combinatory words such as "ugnayan ng pahinungud," "bahay kalinaw," and borrowed words like "salbij," "anawnser," and "isyu" have become common lexical terms (Rubrico, 1998).

Chretien had, as an initial step, indicated a family of Philippine languages closely-related to each other by geography and intercultural contacts. Paz has posited that these languages have in fact developed from an ancient proto-Philippine language. By marrying these two concepts together, this paper suggests an evolving series of language sub-groupings tentatively labeled as "Greater Ilokano," "Greater Tagalog," and "Greater Sebuano." Please refer to the "Proposed Linguistic Map of the Philippines Showing Major Groupings" (circa 2001) below. To what extent these sub-groupings have evolved in present times will require a more thorough research outside the scope of this paper. Nevertheless, the diachronic implications of this development are sundry and manifold.

A Classification of Twenty-nine Contemporary Philippine Languages/*Jessie Grace U Rubrico*

**Proposed Linguistic Map of the Philippines
Showing Major Groupings
Jessie Grace U. Rubrico, 2001**

Itb

Itw

ILK

Ibg

Isi

Ka

Png

Igt

Kap

Bon

Iba

Kam

Vir

TAG

Nag

Agt

Tbw

Akl

War

Buh

SEB

Buk

Bah

Mar

Bla

Yak

Sub

Tau

Bag

References

Chretien, Douglas. 1961. "A Classification of Twenty-One Philippine Languages." A paper presented at the Tenth Pacific Science Congress of the Pacific Science Association, August 21-September 6, 1961, at the University of Hawaii under the joint sponsorship of the National Academy of Sciences, Bernice Pauahi Bishop Museum, and the University. of Hawaii.

Lehmann,Winfred A. 1992[1962]. *Historical Linguistics: An Introduction.* 3rd rev. ed.   London & New york: Routledge.

Paz, Consuelo J. 1981. "A Reconstruction of Proto-Philippine Phonemes and Morphemes," *The Archive: A Journal Devoted to the Study of Philippine Languages and Dialects 3,* Ernesto Constantino, et.al. (eds.) . Published by the Cecilio Lopez Archives and Philippine Linguistic Circle:U. P. Diliman, Quezon City.

Rubrico, Jessie Grace U. "The Metamorphosis of Filipino as National Language." Retrieved 21 January 2001 from the author's website at:
 www.languagelinks.org/onlinepapers/fil_met.html