Towards a Definitive Philippine Wordlist: The Qualitative Use of Vocabulary in Identifying and Classifying Languages

R. David Zorc

R. DAVID ZORC

TOWARDS A DEFINITIVE PHILIPPINE
WORDLIST--THE QUALITATIVE USE OF
VOCABULARY IN IDENTIFYING AND
CLASSIFYING LANGUAGES

1. OBJECTIVES. Most linguists
doing research on Philippine
languages have been in agreement
in including the glosses of the
Swadesh or Gudschinsky lists in
their fieldwork, if not for
actual lexicostatistical or
glottochronological purposes,
then to obtain matching sets for
comparative work. The number of
linguists who have gone consider-
ably beyond those 200 or so basic
vocabulary items has increased in
recent years, thereby increasing
the repertoire of forms that com-
parativists have to work with.
With so many researchers now in
the field, and with a growing
concern over the linguistic
situation in Borneo and Celebes
(many of the languages of which
are close in type to languages of
the Philippines), the time is ripe
for a serious attempt at drawing
up a definitive, enlarged, cul-
turally-oriented list in order to
insure a larger body of semanti-
cally matched data for continued
comparative work on the one hand,
and in order to begin the recon-
struction of Philippine culture
history on the other.

    In traditional terms, a word is
composed of form (a cluster of

phonological features) and meaning (a cluster of se-
mantic units). The gamut of lexical research deals
with each and both aspects: in fieldwork glosses
are given (usually translations in a contact lan-
guage) in order to elicit forms that accurately
translate those meanings. The development of any
list must rely on consideration of the glosses that
should be used, the forms expected to be elicited by
those glosses, and the status of those forms (such
as: Will form X identify a subgroup? Will form Y
probably contain a good example of *R, or *j, or *e?
Will form Z be instructive or interesting in the
reconstruction of Philippine cultural history?)

It is the purpose of this paper to examine exist-
ing wordlists, to explore the various criteria that
have been employed in their formation, and to formu-
late principles for the development of a single
comprehensive Philippine questionnaire, which could
then be adapted to the needs of each researcher.[1]

Unfortunately, the most intriguing notion of all,
that of ranking a long list of (1000) glosses that
are relevant to Philippine life and culture beyond
the ranking done by Dyen for the Swadesh list (see
2.4 below), has not yet been undertaken due to the
lack of time and data necessary for such a compre-
hensive statistical analysis. An approximation in
this direction is attempted in Section 5, working
only on a broader base of word groups (rather than an
actual list ranked from 1 through N), drawn up to
show possible levels of retentiveness from Proto-
Philippine or from various posited Philippine meso-
languages.

2. EXISTING WORDLISTS. Many wordlists currently
exist, drawn up by researchers, and tailored to the
specific needs of each linguist, depending on the
objective of the survey intended. In recent research
on Philippine languages, such objectives have been:
the gathering of a simple wordlist, the beginnings of
a dictionary-project, a dialect survey, the subgroup-
ing of newly discovered speech types, the making of a
dialect geography or linguistic atlas, the discovery
of correspondences with the established phonemes of a
posited protolanguage, the making of an etymological
dictionary of immediate or remote mesolanguages or
protolanguages, and so on. Is it possible that a

single list can be developed so that it can be
adapted or employed for any or all of these func-
tions? If so, what should the size, the content,
and the ordering of such a list be based on?

The most well-known are the lexicostatistical
lists of Swadesh (100 and 200 glosses) and
Gudschinsky (215 glosses). These have been drawn
up on the premise of containing and obtaining lan-
guage universal meanings that are non-cultural in
character. Fieldworkers have used these lists in
every conceivable configuration: arranged alpha-
betically in the linguist's native tongue, or in a
contact language or trade language; arranged se-
mantically by groups (such as body parts, pronouns,
natural phenomena, etc.); and arranged according to
the proposed etyma for each gloss. Laycock (1970)
reports a structural grouping, i.e., the arrange-
ment of items by word classes (noun, verb, adjective),
among some lists for use in the New Guinea area, but
no such arrangement has yet been noted among lists
for use in the Philippines.

Among lists drawn up specifically for languages
of the Philippine type there are three kinds of
ordering.

    2.1. *Alphabetically arranged lists.*

        Conklin (1951) [2278 entries, 107 pp.]
        Fox, *et al.* (1965) [197 entries]
        Grimes (1972) [c. 2070 entries]
        Institute of National Language (1953)
           [1110 entries, 139 pp.]
        McFarland, Curtis (1972) [400 entries]
        Postma (n.d.) [483 entries]
        Ray (1911) [211 main entries + 28 functors]
        Reid (1971) [372 entries + pronouns, deictics]
        SIL (1962) [304 entries + pronouns, deictics]
        Zorc (1968) [350 entries]
        Zorc (1971*a* [500 basic entries, with cross
           references]

    2.2. *Semantically arranged lists.*

        Dyen (1973) [45 entries]
        Esser-Holle list (1931) [1047 entries +
           36 sentences; 131 pp.]
        Ferrell (1969) [c. 336 entries]

```
                    Prentice (1969) [250 entries]
                    Tri-Institutional Pacific Program (1954)
                        [920 entries; 611 lexical + 309
                        grammatical]
                    Tsuchida, Shigeru (1962) [513 main entries
                        + sentences; 27 pp.]
                    Tsuchida, Shigeru (1968) [2331 main entries
                        + sentences; 94 pp.]
```

The above lists are each grouped around semantic
concepts as determined by the authors. The Tsuchida
lists have a code whereby more important items are
encircled in the case of a more rushed or less in-
tensive survey.

   2.3. *Etymologically oriented lists.* In some
cases, fieldworkers have found it important to have
a list of etyma to be traced in each individual sur-
vey. In this case, the form takes a certain priority
over meaning, with the justification that one wishes
to reconstruct the reflexes of the various proto-
phonemes, or to test the validity or expand the then-
known list of reconstructions.

```
                    Dyen-McFarland (1970) [755 entries; 31 pp.]
                    Llamzon (n.d.) [127 forms + 50 sentences
                        + 73 numerals]
                    Llamzon (1971) [111 lexical items + 40
                        sentences + 140 functors]
                    Zorc (1971b) [3773 entries; 113 pp.
                        + 152 pp. English glossary]
```

   There are two lists which I have been unable to
classify under any of these three headings because
no order is apparent in the presentation of the
entries.

```
                    Pittman and associates (1953) [58 entries/
                        morphemes]
                    SILUND (Summer Institute of Linguistics/
                        University of North Dakota--Work Papers)
                        (1959) [144 entries]
```

   2.4. *A ranked list.* There is only one wordlist
of which I am aware, that has an underlying notion
of ranking. Dyen's adaptation of the Swadesh 200,
viz., 196 glosses, have been ranked for the proba-
bility of retention of each form in each meaning.
This ranked list is discussed in two articles:

Dyen (1964) [196 entries]
Dyen, *et al.* (1967) [196 entries]

In this latter article, Dyen, James, and Cole present a new theory:

> They (Swadesh, Lees, Dyen)...make the assumption that the retention (or persistence) rate of words is the same for each meaning of the lexicostatistical list. This assumption is badly in need of modification in order to produce a realistic model. We now permit different retention rates for words in different meanings. Cognation between words in a meaning of a lower retention rate scores more strongly for closeness of relationship than cognation between words in a meaning of a higher retention rate. [150]

This kind of wordlist is quite important in determining the kind of questionnaire one will wish to work with. If, for example, one wants to determine the phonological correspondences of a language with another (or with a protolanguage), then one would stick primarily to the top of the list (i.e., the higher ranked items); if one wants to undertake subgrouping, he would attach more significance to those agreements towards the bottom of the list.

3. SITUATIONS THAT REQUIRE ADJUSTMENTS TO ONE'S LISTS. Many linguists have found that they were prepared, and yet unprepared, for various situations in the field. That is, they have had set goals and objectives, and have been armed with a job-specific questionnaire, yet they found it necessary to re-adjust or re-orient their objectives under different circumstances than those expected.

A look at some actual experiences in fieldwork and collation will be helpful. These situations prompted the initial research that went into this paper. In discussing them with other linguists, I found that I was not alone in the desire to have a list or even several lists organized in such a way that it facilitated research, depending on one's objectives, or on situations as they arose.

3.1. *Situation one: Limiting data to qualitative vocabulary.* In evaluating scores that I obtained from

a hand-count of the Swadesh 100 list over many dia-
lect and language pairs, I was impressed by several
alignments which seemed in error. One of those was
that of Kagayanen, the language spoken on Cagayan
Island, between Negros and Palawan. Although it
shows no significantly high score with any Philip-
pine language, it has scores in the neighborhood of
60 percent with Kuyonon, Kinaray-a, and Aklanon
(which are members of the western branch of the
Bisayan family) and with Binukid and other members
of the Manobo family (such as Ilianen). High scores
with the Bisayan cluster, particularly with Kuyonon,
led Dyen to classify Kagayanen in the Tagalic Hesion,
coordinate with Bisayan, Mamanwa, and Tagalog.
[1965:29; Dyen has since grouped Kag and other lan-
guages of the Manobo groups together into a single
family (personal communication).] Since Manobo and
Bisayan are two separate language families within
the Sulic (Southern Philippine) Hesion, it is un-
likely that we have a linking member in Kag. Further-
more, the scores of Kag do not parallel the Bis
groupings. Thus, even from the alignment of lexico-
statistical percentages there is reason to suspect
the scores of Kag with the other languages.

If one looks more critically at the content of the
100 word list for Kag (see Table 1) there are 17
forms that are more readily traced to the Manobo
family, and 13 forms that appear to be Bisayan. Of
these, only four are found throughout the Bisayan
group; five are more typically West Bisayan, and four
are more typically South Bisayan. We get no informa-
tion from 58 of the forms because they are found
scattered throughout the Philippines, while 12 are
problematic in that they represent innovations in
form or meaning within Kag itself. It is important
to note that of the 100 forms then, only 30 are of
significance in giving information about the genetic
relationship of Kag to a Philippine language group.
If we look at those 30 forms, evidence for membership
within either Bis or Man should become apparent.
The 17 Kag forms that can generally be traced to
Manobo, or, more specifically, that cannot be related
to any Bis dialect are found in Table 2. On the
other hand, the 13 Kag forms that do not appear in
Manobo languages or dialects, and which appear to be
Bisayan are found in Table 3.

TABLE 1. Kagayanen 100 Word List

all	tanán	Bs	man/male	ma:ma	Mb
ashes	qabú	PH	many	ta:maq	--
belly	gettek	Mb	meat	sapúq	--
big	bakéd	--	moon	bu:ʟan	PH
bird	yu:pan	PH	mountain	bu:kid	PH
bite	kagát	PH	mouth	baqbaq	PH
black	mi:tem	PH	name	nga:ran	PH
blood	langessa	Mb	neck	liqég	PH
body	la:wa	Mb	new	bagqu	PH
bone	bekkeg	SBs	night	kilém	Mb
breast	su:su	PH	nose	qirúng	PH
burn	su:nug	PH	not	di:liq	SBs
cloud(rain)	qitém	--	one	qisya	--
cold	tignaw	SBs	person	qittaw	Mb
come/arrive	qabút	Bs	rain	qurán	PH
die	-patáy	PH	red	min:ug	--
dog	qa:yam	PH	road/trail	da:ʟan	PH
drink	qinúm	PH	root	gamút	PH
dry	-ma:ra	PH	round	bilúg	PH
ear	tali:nga	PH	sand	pantad	Mb
earth	basák	Mb	say/said	qambaʟ	WBs
eat	ka:qan	Mb	see	ki:taq	PH
egg	tallug	PH	seed-rice	bi:niq	PH
eye	matá	PH	sit	pungkuq	WBs
fat	tambek	Bs	skin	langgit	Mb
feather	buʟbuʟ	PH	sleep	tunu:ga	Mb
fingernail	su:ʟu	Mb	small	sisét	--
fire	qapúy	PH	smoke	qasú	PH
fish (n)	siddaq	PH	stand	tindeg	PH
fly (v)	layúg	PH	star	bituqún	PH
foot	bati:qis	PH	stone	batú	PH
full	pennuq	PH	sun	qadlaw	PH
give	qa:tag	SBs	swim	luuy	--
good at	miyád	WBs	tail	qi:kug	PH
green/unripe	qiláw	PH	this	tiní	PH
hair	buuk	PH	that	sanyaq	Mb
hand	li:ma	PH	thou	ka:un	Mb
head	qu:ʟu	PH	tongue	di:laq	PH
hear	ka-ma:tiq	Bs	tooth	ngi:pen	PH
heart	tagipusu:qun	WBs	tree/wood	ka:uy	PH
horn	sungáy	PH	two	darwa	PH
I	yakén ~ qa	Mb	walk	panáw	PH
kill	patáy	PH	warm/hot	qi:nit	PH
knee	bu:qul	Mb	water	wa:ig	Mb
know-fact	na:man	WBs	we (excl)	kamí	PH

leaf	da:un	PH	what?	qarán	--
lie down	neggaq	--	white	putíq	PH
liver	qatáy	PH	who?	kinú	--
long	langkaw	--	woman	ba:y	PH
louse	kutú ~ tumá	PH	yellow	duɫaw	PH

Bs = General Bisayan form; SBs = South Bisayan; WBs = West
Bisayan; Mb = a Manobo form; PH = a Philippine form of wider
occurrence than just Manobo or Bisayan.

TABLE 2. Kagayanen Forms Relatable to Manobo

	KAGAYANEN	EXPECTED, IF BISAYAN
'belly'	gettek	*tiyán
'blood'	langessa	*dugúq
'body'	la:wa	*la:was
'earth'	basák	*lugtaq, *lu:paq
'eat'	ka:qan	*ka:qen
'fingernail'	su:ɫu	*kukú, *kulú
'I'	qa	*qakú
'knee'	bu:qul	*tu:(h)ud
'man'	ma:ma	*lala:ki
'night'	ki:lem	*gab(i)qi, *delém
'person'	qittaw	*ta:wu
'sand'	pantad	*barás, *bu(h)a:ngin
'skin'	langgit	*pa:nit
'sleep'	tunu:ga	*tu:rug
'that (near)'	sanyaq	*qináq, *dan, *qiyán, *yaqún
'thou'	ka:un	*qikáw
'water'	wa:ig	*tu:big

On the basis of the 30 forms gleaned from just
the 100 word list, we can conclude that Kag is a
Manobo, rather than a Bisayan language. The follow-
ing are some of the reasons:

1. The quality of the Manobo innovations which
are found in Kag is rather convincing: *langesa
'blood', *getek 'belly', *qa 'I' (enclitic form),

TABLE 3.   *Kagayanen Forms Relatable to Bisayan*

	GENERAL BISAYAN	EXPECTED MANOBO FORM
'all	tanán	*langun, *tibeq, *qelin
'arrive'	qabút	*quma, *dungguk
'hear'	ka-ma:tiq	*dineg, *paliman
'fat'	tambek	*lambuq
	SOUTH BISAYAN	
'bone'	bekkeg	*tulqan (N.B. MPH *bekeg 'fishbone')
'cold'	tignaw	*gennaw (cf. Kag ginnaw 'chilled')
'give	qa:tag	*beggay
'not (so)'	di:liq	*kennaq
	WEST BISAYAN	
'good at'	miyád	*(qu)piya
'heart'	tagipusu:qun	*pusung
'know (facts)'	na:man	*sabut, *taqu, (+?)
'say'	qambaɬ	*ka:gi
'sit'	pungkuq	*pinuqu, (+?)

*buqel 'knee', *-kilem 'night', *langgit 'skin',
*ku:na 'thou', *-yaq second person deictic.

2.   The contrastive evidence that, while forms like
*sulu 'fingernail', *wahiR 'water', *la:wa 'body',
*ma-qa:ma 'man', *qetau 'person' are more widespread
than just the Manobo subgroup, they are found through-
out the Manobo subgroup, but are not found in any Bis
dialect, nor even in the wider circle of Tagalic
languages, to which Bis belongs.

3.   The uneven distribution of forms from two
different Bis subgroups suggests two different
periods of contact, rather than the continuation of
a single genetic descendant.   No Bis dialect shows
such a distribution, since each dialect of a Bis sub-
group agrees in reflecting a form common to its own

subgroup in the meanings cited. That is to say, all
WBs dialects (Akl, Kin, Kuy, etc.) reflect *tulqan
'bone', *ma-ramíg 'cold', *taqú 'to give', and *qindiq
'will not' vs *bekén 'not so', while all SBs dialects
(Sur, Jaun, But) reflect *ma-dayáw 'good', *kasingka:
sing 'heart', *hibarú/*qingat 'to know', *laqúng 'to
say', and *(1,q)ingkud 'to sit' (compare these forms
with those given in Table 3).

Confirming these conclusions are the following
points from outside the domain of the 100 word list.

4. We know that the Manobos as a whole are not
seafarers, nor have they been for some time now. On
the other hand, we know that the Bisayans are and
have been seafarers, traders, fishermen. It is a
simpler solution to explain the Manobo elements on
Cagayan Island as the retentions of an original Manobo
immigrant population and the Bisayan elements as of
secondary introduction.

5. Further study of other forms, particularly the
pronouns *din 'his/her', *nay 'ours', *dan 'theirs',
*kiyu 'ye', *kay 'we' (excl.-enclitic), *kaw 'ye'
(enclitic) illustrate the Manobo substratum of
Kagayanen. So it is with other lexical innovations
that so far appear only in other Manobo languages:
Kag qindis, Man *qindes 'to defecate', Kag ma-dyuq,
Man *ma-dyuq 'far', Kag la:suq, Man *lasuq 'penis',
Kag n-ła:qu, Man *laqu 'thirsty', Kag qindi, Man
hendeqi 'where?', Kag qansaq, Wbm qinsaq 'to ask',
Kag qumaw, Man *qumaw 'to call', Kag, Bkd, Dbw lamqed
'to swallow', Kag błengngan, Ata, Tig *qabelengan
'throat'.

Thus, our attention is drawn to the possibilities
and problems of ranking lexical evidence such that
more information is obtained, and of excluding other
evidence as inappropriate for our purposes. The con-
sideration of select lexical elements to the exclusion
of others proves helpful, if not significant, in the
subgrouping or classification of a speech type. In the
case of Kag, we have other information supporting the
conclusions arrived at through the use of lexical evi-
dence. Wherever possible, one must make use of
geographical, ethnographical, archaeological, or other
information. (See Sapir (1916), "Time Perspective in
Aboriginal American Culture: A Study in Method.")

3.2. *Situation two: The need for specific lists
in the field.* In a dialect or language survey when
one wants a first approximation where a speech type
fits into the already known language picture, one
will want to avoid both universally distributed and
overly-specific forms. Once one has an idea of the
language placement, then he can get dialect-specific
forms in order to judge the points of contrast from
other dialects, and he can get more universal forms
in order to compute the reflexes more accurately for
each protophoneme.

With a carefully constructed list oriented to
innovations in the various Philippine mesolanguages,
one can determine after a few hours work if continued
research will be productive or repetitive. It may
prove advantageous to close up shop and move on to a
different area.

Sometimes the time element can be even more of a
consideration. On numerous occasions, I found myself
waiting for some means of transportation--on a de-
serted stretch of road, at a rural airport, near a
provincial pier. I would have an uncertain amount of
time on my hands before the vehicle arrived. In-
variably someone walked up or by who became of lin-
guistic interest (if not importance), such as a small
group of Tadyawans (outside of Victoria, Mindoro) who
were on their way to town loaded with bundles of
rattan for sale, or, at the Cagayan de Oro airport,
a speaker of a dialect of Surigao that retained the
original four-vowel system. I would fumble for my
notebook and pen, but then found myself fumbling for
what to ask. The elicitation of my alphabetically
arranged 500 word list and my 250 sentence grammatical
questionnaire was out of the question within the next
few minutes (or even hours) before my transportation
arrived. There arose the need of having a list ar-
ranged, not in alphabetical or even semantic order,
but by specialized criteria indicating probabilities
of genetic affiliation, subgroup membership, inter-
esting phonological reflexes, and the like. Even a
short list like the Swadesh 100 has items of low
informational content in that the probability of
predicting what the form would be was rather high,
such as [mata], [di:laq], [ka:huy], [duwa], and the
like.[2]

4. DEVELOPING A DECISIVE PHILIPPINE WORDLIST. Limit-
ations on time or content have led to the consider-
ation of limiting or marking items for elicitation or
research. Words that have a high predictability
factor, because they have a greater probability of
being retained, may be separated from those that tend
to be very diverse, because there is a high probabil-
ity of replacement or innovation. If we elicit a form
cognate with *bu:lan 'moon', *pa:naw 'to walk; to
leave', or *penúq 'full', we learn only that the lan-
guage is Malayo-Polynesian. Forms that reflect in-
novations of lower order protolanguages are ranked
higher in significance because they give qualitative
evidence of membership in some subgroup, such as
Bisayan *damgu 'to dream' or *gegma 'to love', Manobo
*langesa 'blood', Samalan *sángum 'night', Ifugao-
Pangasinan *(q)eléng 'nose', Hanunoic *panggasan
'star', Kalamian *tinanguni 'body', and the like.

Since we are proposing a list dealing with lan-
guages of the Philippine type (which include some
languages of Borneo and Celebes), widespread forms
from earlier protolanguages give only one kind of
information; forms from various mesolanguages give
another kind of information; and dialect-particular
forms give still another. It is important to keep
these kinds of forms apart, at least by marking them
in one way or another, so that we can glean the kind
of information we want when we need it.

Before going directly to a consideration of rating
glosses and forms according to various criteria, it
is important to have an overall view of the kinds of
meanings and forms we will want to be dealing with.
Based on these observations any researcher must add
to or subtract from the wordlist presented in this
paper according to his needs.

4.1. *Overall criteria for including a gloss or
meaning in a list*. The following criteria are sug-
gested as important in the consideration of whether
or not one wants to include a meaning in a wordlist
for the field or in comparative studies at home.

1. The meaning is language universal (such as
those proposed by Swadesh, Gudschinsky, *et al.*).

2. The meaning, and hence the form sought, are in
the basic vocabulary of the target language, e.g.,
'banana', 'cooking pot', 'coconut', etc.

3. The meaning is one that undergoes only infrequent replacements of form, and thus the forms elicited may prove important in the subgrouping of members of various mesolanguages, e.g., 'water' (*Danúm, *tu:biR, *wa:hiR, *sapáq, etc.).

4. The meaning most often elicits a retentive form, which (if cognate) will be of phonological interest or importance, such as determining the reflexes of *D, *e, *j, *q, *R, and so on, such as *pu:sej 'navel', *qatéy 'liver', *Ramút 'root', *hiD(e)Ráq 'to lie down', etc.

5. The meaning (or the form expected) is of historical interest, such as dating the introduction and spread of rice agriculture, weaving, metalwork, etc.

6. The meaning (or the form) is of cultural interest, such as the importance of cockfighting as a sport (as evidenced by a proliferation of terms in this area), the types and use of medical or medicinal objects (dry-cupping, herbs, and other cures), etc.

7. The meaning is on someone else's list or is part of an established survey (such as Ray (1911), INL (1953), Ferrell (1969), Llamzon (1969), Reid (1971)), and is thus available for further comparative work. This paper includes all forms from the above-mentioned comparative lists.

4.2. *General criteria for excluding a gloss or meaning from a list.* The following criteria are suggested as precautionary in determining whether or not a form or meaning should be excluded from a wordlist.

1. The form elicited is a widespread borrowing, even if the concept or function is a native one. Glosses such as 'green', 'trunk (box)', 'fence', 'but', 'more than', 'until' have frequently been replaced by Spanish forms. However, such replacements may be interesting if not important in comparative work, and alone do not serve as definitive criteria for exclusion. The use of Spanish gusto in many areas for the concept of 'want, like' is both widespread and significant, even if there is a local equivalent. The question to what degree hispanicization has taken place throughout the Philippines is an important one, and the degree to which a researcher

wants to become involved in this question should
determine whether or not he will include such
glosses.

2. The form is culturally irrelevant, or at
least bound to another culture outside of the lan-
guage group being surveyed. This may apply to the
Philippines as a whole, where there is a recon-
structible 'pillow', but not 'pillowcase' or 'bed';
or it may apply to select areas, such that a city
dweller may not be aware of betel-chewing terminol-
ogy, a mountain dweller of a fisherman's tools of
trade, or a remote tribe of words such as 'high
tide', 'low tide', or 'deep sea'. A list,or at
least the items on a list, must be adjusted to the
sophistication and awareness of each informant, and
of each informant's community and environment.

3. The meaning is apt to be confusing to either
the researcher or the informant and is likely to
lead to an error in elicitation. Thus, 'lie' may
elicit 'to tell an untruth' or 'to recline', 'green'
may be a term related to texture or ripeness rather
than color. Of course, careful explanation of a
gloss can remedy this problem, such as 'bark (of
tree)' vs 'bark (of dog)'.

The role of the elicitation language is also
important in this regard. If a single term is sought
which is not contained in the elicitation language,
the informant may translate rather than give the
monomorphemic equivalent. Thus, in the case of
Tagalog, one may use alisin am butó or himáy 'to re-
move bones (from fowl or fish)' and, depending on
which is used, receive a cognate of CPH *hiN-bakeR
or *kuha bakeR in return.

4. The meaning is likely to cover a class of
meanings for which there is no single generic term
in the target language. Since we want full cognate
sets in our comparative work it is better to exclude
such a meaning rather than introduce a clearcut
probability of error (i.e., the form elicited would
not actually be non-cognate, but just poorly matched).
Such is the case when the meaning is a product of
generalization in English or other Western languages
that lump together whole classes of concepts that are
found in multiple forms and meanings in the Philip-
pines. Thus 'hit' must be specified as to the type,

the form, and the object of hitting, i.e., to hit a
person with one's fist, to hit a person with an ob-
ject (stick vs whip vs ax, etc.), to hit a thing with
an object, to hit a thing with one's fist, to slap
across the face, to hit the top of the head, to slap
across the ear, to hit an object with a downward
motion, to hit an object with a sideward motion, etc.
Likewise, 'shake' would have to be specified, as in
to shake or throttle a person, to shake a tree in
order to get fruit down, to shake oneself as a wet
dog shaking off water, to shake a container to see if
there is liquid inside, to shake a container to see
if there is something solid inside, to see if there
is something loose inside, etc. See 'to hang' in
Group 7.

5. The meaning may be part of a pair of meanings,
one of which may be omitted where there is bound to
be an overlap of forms within the Philippine group
under research. Thus, the forms for 'die' and 'kill'
are often the same, or related as stative and causa-
tive counterparts. However, this criterion must be
applied with caution. Reid mentions that an original
SIL draft "eliminated from the Swadesh list . . . one
member of frequently occurring doublets, such as
husband/wife, foot/leg, meat (flesh)/fruit, wood/
stick, sun/day." (1971:viii) In several cases, this
omission was infelicitous, since many languages do
have different forms for 'sun' as opposed to 'day
(time)', 'meat' vs 'fruit', 'foot' as opposed to
'leg'. Rather than eliminating such glosses alto-
gether, it would be most advantageous to list suspect
pairs together, and then omit them during research on
each language encountered where duplication does occur,
marking the duplication with ditto marks or an x. It
is in this case that semantic organization of a list
is far superior to an alphabetical arrangement of the
glosses.

4.3. *Types of vocabulary worth searching for*. The
following considerations concerning the lexicon must
be weighed when choosing meanings to elicit particular
forms.

1. *Universal Forms*, which occur throughout all
Philippine groups. Such forms usually have one or
two characteristics that make them a desirable goal.
(a) There is one basic reconstructible proto form,
with regular reflexes throughout the subgroups, such

as *la:ña 'vegetable/coconut oil', *di:laq 'tongue'.
(b) The form is well-defined in that it has one
clear-cut meaning assigned to it. A good example is
*buqa:ya 'crocodile'. There are areas in the Philip-
pines where no crocodiles can be found, yet the form
is passed on in legends, tales, and fables, so that
the word has survived as a definite part of Philip-
pine basic vocabulary. While the referent is
environmentally restricted, the form is well-defined.

2. *Widespread Forms*, which tend to diversify due
to semantic changes, such as *pa:qa 'thigh; leg;
foot', *betíØis 'calf; leg; foot', *bibíR 'mouth;
lips; jaw', and so on. Some of these forms may have
been loosely defined in the proto language, or be-
came so in subsequent mesolanguages, such as *bulbul,
which is most widespread in the form 'feather', since
it is often found in the frozen form *himulbul (< *hiN
+bulbul) 'to remove feathers'; but this form has
passed on rather freely to the meanings 'body hair',
'fur', 'underarm hair', 'pubic hair', '(head) hair',
and so on.

3. *Subgroup Particular Forms* which usually inno-
vation or replacement of a standard and earlier form,
the new form surviving in subsequent stages of daugh-
ter languages. The replacement of $Da:Raq 'blood' by
*DuRúq 'sap' is a classic instance of this among Meso-
Philippine languages, or by *lang()sa 'gory, having
the smell of fish or blood' among most Manobo lan-
guages.
(a) Such replacements can be semantic or lexical,
although to some degree both are involved in every
change. For example, among some languages of the
southern Philippines, *hi:nang means 'to work, make,
do'; in some Bikol dialects this form (presumably
related) means 'to sweat'. Although this is a seman-
tic shift in those Bikol dialects, it obviously is a
lexical replacement as well in that *hi:nang must
have replaced a then-existing word for 'sweat'. The
interplay of these two factors is noteworthy in that
replacement often involves two innovations or even a
chain of innovations: the loss of a form, the intro-
duction of another form, competing forms, etc. Among
South Bisayan dialects, the otherwise widespread
Central Philippine *dakmel 'thick' has been lost, re-
placed by *bagáq, which originally meant 'swollen; an
abscess or boil'; in turn, an apparent innovation
*he:bag now fills the semantic space of *bagáq,
meaning 'swollen; boil, abscess'.

(b)   There are some meanings that appear well-defined
(at least from our Western point of view), but which
elicit a plethora of forms defining very small dia-
lect areas, such as forms for 'sweat', 'acrid (flavor
of unripe banana or of betel nut)'. The elicitation
of such forms proves very rewarding in surveys
dealing with the diversity of dialect chains.
(c)   Other forms that serve to define mesogroups
have frozen descriptive morphemes which point to an
earlier period when the word may have been coined,
but which has been retained since then, such as the
many words for 'rainbow', 'shooting star', 'pupil
(of the eye)', as well as the risky but intriguing
field of names for various species of insects,
flowers, fish, shellfish, and the like.
(d)   Some forms have seemingly mobile semantic
features that are difficult to encapsulate in an
English gloss, and the researcher must be aware of
the history of the form in order to elicit it. That
is to say, a form becomes the end of one's research,
and the semantic changes throughout the Philippines
are an important study in themselves. For example,
*tuktuk has at least the following glosses 'top of
head', 'forehead', 'peak of mountain', 'summit of
mountain', 'top of [anything]'; *[qØ]u:nung 'to be
loyal to', 'a cause, reason', 'to die with', 'to
avenge', 'to report on', 'to guard', 'to watch over'.
While there is a semantic thread connecting all of
these meanings to a single form, there is no way one
can be assured of eliciting the form without asking
for it (or its expected reflex) and then ascertaining
its meaning within the target language. It thus be-
comes necessary to have some kind of etymological
checklist incorporated with one's questionnaire, such
as listing: *qira:ya 'inland, upriver/upstream,
upwards, in the hinterlands' and *qila(w)úd 'seawards,
downriver/downstream, downwards, towards town'.

4.4. *Some notes on basic vocabulary in the
Philippines*. The whole notion of just what is basic
vocabulary is open to question, challenge, and dis-
cussion. The following is meant to be a discussion
of some of those points that have struck me, and I
welcome both criticism of these points as well as the
addition of others that I have overlooked.

4.4.1. High text frequency. Perhaps the most
obvious form of basic vocabulary is that which is not
directly considered vocabulary in the sense of

contentives or lexical items, but rather functors or
grammar-based forms. If one were to let a tape re-
corder run for any length of time in an area where
speech is most likely to occur, and then transcribe
and collate the recording, chances are that very
common lexical items, such as 'eat', 'sleep', 'eye',
'tongue', 'full', and so on may not occur more than
a few times, if at all. But in the Philippines (as
elsewhere) the text would be replete with pronouns,
deictics, conjunctions, negatives, interjections,
and the like. In the Philippines, one would also
find case marking particles (at least on personal
nouns), and a large set of discourse particles, like
Tagalog palá, kayáq, namán, sa:na, ay, qe, mu:na,
na, pa, qe:wan, ba, etc. The occurrence of such
forms has in actual case studies far outranked the
occurrence of even the simplest vocabulary items.
Such core items rank high in the list of disiderata.
Due to problems of elicitation, it is obvious that
in a rushed survey, many of them, such as more
subtle quotative or attitudinal particles, cannot be
obtained, but the more solid equivalents of Tagalog
na 'already; now', pa 'still, yet', din/rin 'also,
too', la:mang 'only, just', mu:na 'first', sa:na
'hopefully' can easily be elicited through sentences
giving an appropriate context.

4.4.2. Common knowledge. Words known to any
native speaker at an average speech level should
form the core of basic vocabulary among contentives.
A corollary on the side of the researcher would be
words which are easiest to elicit with accuracy.
There are problems that beset either the informant
or the linguist which make certain words poor can-
didates. Some of these difficulties may be:

1. distinction in language level, which is more
prevalent in the Indonesian area than in the Philip-
pines, but does occur in the Philippines in select
areas of vocabulary and hence must be taken into
account.[3]

2. the gender of a form as opposed to the sex of
a speaker. It is a well-known fact that there are
male and female intonation patterns in the Philippines,
and that an utterance pronounced one way may be
effeminate for a man or tomboyish for a woman. This
distinction runs into at least the vulgar vocabulary
of many dialects, and may be more widespread.

3. the age of the informant and the relative age-level of the referent. The most clearcut example I have of this are three words for 'vagina' in Masbate, depending on whether it is that of a child [putáy], a mature girl [pudá], or an adult [birát].

4. the prestige or status of a form, whether or not it is regulated by a social taboo, and so on. A form that comes closer in sound to a prestigious trade language may be given rather than the local equivalent. Often informants have given me a Tagalog or Cebuano form which I used for elicitation as true of their own language, while it is at best a form known in the language community, but not the native word which is actually used among speakers in their conversations.

I take the simplest starting point to be forms that can be elicited monolingually with accuracy, such as parts of the body, or the immediate environment (house, leaf, rock, water, earth). After this would come those forms one could elicit with accuracy through a contact language, where each has a one-to-one semantic correspondence. Here one must proceed with caution in that the understanding of the meaning of the form in the contact language is presumed. All kinds of error occur in this regard. For example, in eliciting the word for 'acrid (flavor of unripe banana)' one may use the Tagalog pakla, but the informant may have heard or the fieldworker may have pronounced the Tagalog baklaq 'effeminate'. One way to circumvent this problem is to elicit forms through sentences, such as 'the banana is acrid', in which case it is helpful for the fieldworker to review the forms he desires and combine them in such a way as to be economical. In the above case, one would also obtain the form for 'banana' and it would not be necessary to reelicit it. The combination of many forms into a few well-phrased sentences is a field tactic that saves time, on the one hand, and helps limit error (through misunderstanding), on the other.

4.4.3. Well-defined. A term that is well-defined is usually part of the most unmarked core vocabulary of a language. Such forms are widespread throughout the Philippines and are an important kernel in comparative work. Some good examples are the low numbers and the better-defined parts of the body. Although many plant and animal names are subject to spread and borrowing, some are amazingly stable and

reflect a single etymon, such as *lengá 'sesame',
*lequ:ya 'ginger', *ba:wang 'garlic', and others.

4.4.4. Retentive. Many forms that may or may
not strike us as basic vocabulary have been retained
from an early protolanguage. Some of the factors
that seem to be involved are the following.

1. The frequency of incidence of the object or
referent. If something is used (*ku:Den 'cooking
pot'), seen (*bu:lan 'moon', *bitu:qen 'star'), or
referred to (*buqa:ya 'crocodile') often enough in
the history of a language, it is more apt to be re-
tained intact in both shape and meaning. But there
are also classic exceptions to this, such as Hanunoo
panggasan 'star' and Buhid magdánun 'moon', Palawano
benwa 'house', and Kuyonon tingway 'needle'.

2. The less marked member of semantic pairs tends
to be preserved, while the more marked member is
frequently subject to innovation and change. That
is, what is *schwer* (to use the German term with its
wide range of meaning) tends to persist longer in
shape and meaning. Thus there are more etyma for
'light(weight)' than for 'heavy' throughout the
archipelago; conversely, there are fewer etyma for
'sour' than for 'sweet', fewer for 'dark' than for
'bright'. To some degree this is parallel to the
phenomenon in English where one word is less marked
in phrases like:

　　　How [tall/short] is he?
　　　How [heavy/light] is it?
　　　How [old/young] are you?
　　　Is it [dark/bright] in there?

The fact that the less marked forms tend to be
preserved longer is no doubt bound up with matters
of frequency and predictability, which we need not
go into here. It is suggested however that such
pairs be put together in a list. Besides facilita-
ting elicitation, a number of comparative features
may be brought to light that would otherwise not be
noticed in a different ordering.

5. THE GROUPING OF A PRELIMINARY LIST. Since not
enough time or data are available for an accurate
statistical ranking of Philippine-oriented glosses,

a preliminary list of over 600 meanings has been
rated according to the following procedure. First,
I divided the Philippine area into subgroups. Where
I have been made aware of genetic relatedness, I made
the subgroups on such information. [The classifica-
tions of Thomas and Healey (1962), of Dyen (1965),
and of Fox, Sibley, and Eggan (1965); my own work in
connection with my dissertation on Bisayan and other
languages of the central Philippine area.] However,
the main intent of this subdivision is not a genetic
classification, but rather a working base for iden-
tifying the distribution of forms, particularly
cognate sets. The first subdivision is a split of
the Philippines into two main groups (North =
Cordilleran, and South = Sulic). Each of these
main subdivisions is assigned four further sub-
groupings (as noted in Table 4). Languages such as
Ivatan (Itbayaten, Yami), Tiruray, Bilaan, Samal
(Abaknon, Jama Mapun, Sibutu), and Sangir-Sangil, I
leave as outlying test languages. It must be under-
stood, however, that a better picture of the actual
subgrouping of Philippine languages will ultimately
lead to an accurate ranking (not just a grouping) of
the glosses and forms discussed here. For the time-
being, the subdivision employed is adequate enough
to account for the number of cognate sets for each
meaning proposed hereunder.

Second, I looked through several data sources
[such as Reid (1971), the Composite Vocabulary (INL.
1953), Panganiban's Tesauro (1972), and my own data
compiled for 60 different Philippine speech types]
and looked for agreement in cognate sets. I ignored
uniques and forms of very limited distribution so
long as they were not numerous, and were not found
in more than two or three of the proposed subgroups.
In an actual ranking, uniques cannot be so ignored.

If five or less cognate sets could be established
throughout the Philippines (including the outlying
test languages), I ranked them from one to five de-
pending on the number of widely distributed etyma
per gloss. These forms are given in the tables for
groups one through five. If one etymon was apparent
for all groups in a single meaning, it was assigned
to group one. If one basic etymon prevailed, but
there were competing etyma well-distributed in the
major subgroups, the form was classified on the
basis of the total number of etyma. For example,

ILOKANO -Ilokano -Itawis -Ibanag -Itneg (Tingguian)	IFUGAO -Ifugao -Kalinga -Balangaw -Bontok -Kankanay
PANGASINAN -Pangasinan -Ibaloi -Kallahan -Ilongot	NEGRITO -Atta -Agta -Gaddang -Yogad
NORTHERN PHILIPPINE GROUPS	
SOUTHERN PHILIPPINE GROUPS	
MESO-PHILIPPINE -Pampango -North Mangyan -South Mangyan -Palawan -Kalamian	CENTRAL PHILIPPINE -Tagalog -Bikol -Bisayan -Mansakic -Subanon

-Manobo Group
 	
-Muslim Group
  [Maranao]
	-Borneo
	
	-Mongondow
	  (Celebes)

since *matá 'eye' is in virtually every Philippine
language (with the only known exception being Iraya
margang) it is assigned to group one. Although
*hiDeRaq 'to lie down' can be found in at least one
language of each subgroup, there are so many com-
peting forms and uniques throughout the entire
Philippine linguistic area that the gloss was
assigned to group six, on the basis that it is
potentially valuable in eliciting forms defining
smaller subgroups (such as *kuláng, found in

Tausug, Butuan, Kamayo, and Mansaka, *lubúg found in
Kalamian, etc.).

This procedure was followed if two, three, four,
or five etyma were found, whether throughout the
Philippines or in the various groups. Thus, while
*Daténg 'to arrive' is found in members of both the
Northern and Southern groups, there are at least four
other reconstructions found in mesogroups, such as
*qabút (in Kalamian and Bisayan), *quma (in Manobo),
*saNpet (scattered in both Northern and Southern
languages), and so on, leading to the classification
of 'to arrive' in group five.

Groups one and two are proposed for determining
the reflexes of the various proto phonemes. Groups
three, four, and five are proposed as lists for
determining membership in the various Philippine
mesolanguage groups.

Group six was reserved for glosses that were
likely to reveal significant information about sub-
grouping a language, if the gloss was not already
listed in groups one through five. In particular,
group six contains a number of functors (pronouns,
deictics, markers, negatives, and particles) that
form a key part of the morphology of a Philippine
language.

Group seven contains those forms which have not
been assigned membership in prior groups. Further
study is necessary to assign them more accurately,
since some items may actually belong to groups one
through five. Group seven is meant to contain all
remaining glosses that are important in a complete
survey of a language. Thus, if a form occurs in
any of the major sources (Swadesh 100 or 200,
Gudschinsky 215, Reid's Minor Languages, Ray's
Borneo Languages, etc.), and it has not yet been
found to belong in an earlier group, it should
occur in group seven.[4] I excluded forms from the
Swadesh list, such as 'freeze', 'ice', and 'snow',
following the procedure of Dyen (1965), but also ex-
cluded three forms used by Dyen, namely 'stick',
'some', and 'with', because each of these presents
difficulties in elicitation.

Group eight is envisioned as a list of items that
are culturally relevant, or historically important.

Words dealing with broad categories such as *agricul-ture, cockfighting, house-building, barter and trade, weaving, clothing, fishing, rice-culture, betel-culture*, that are not found in earlier groups would be put here in the hope that we will be able to make significant steps in reconstructing early Philippine culture history.

Group nine is envisioned as a list for thesaurus-expansion. Thus all of the *flora* and *fauna* in an area, all the species of *fish* or *sealife*, the names of all the *winds* or *phases of the moon* would be in-cluded herewith.

GROUP ONE

1. alcoholic-drink [Arb] *Øa:rak
2. bitter *paqít
3. brain *[hØ]uték
4. branch *saŋá
5. chicken *manúk
6. crocodile *buqa:ya
7. to die/kill *patáy/*matáy
8. to drink *Øinúm
9. to eat *(paN)ka:Øen
10. eight *walú
11. elbow *si:ku
12. embers *ba:Ra(h)
13. excrement *ta:qi
14. eye *matáØ
15. fathom *Depá
16. five *limá
17. flatulence *Øetút
18. four *Øepát
19. full *penúq
20. gall/bile *qap(e)ju(h)
21. ginger *lequ:ya
22. garlic/leek/onion *ba:waŋ
23. to get-up/rise *ba:ŋun
24. goat *kaN[b,d]iŋ
25. gray-hair *qu:ban
26. to grind/mill *gi:liŋ
27. to grow/sprout *tu:buq
28. head *qu:lu
29. hundred *Ratús
30. I *-ak(ú)
31. leech--land *(qa)lima:tek
32. leech--water *lintaq

33. lime *qa:puR
34. liver *qatéy
35. louse *ku:tu
36. monsoon-wind *haba:Rat
37. moon *bu:lan
38. moss *lu:mut
39. name *(ŋ)a:jan
40. navel *pu:sej
41. nine *siyá[m,w]
42. nit *liseh] áq
43. oil [coc./veg.] *la:ña
44. outrigger *ka:tiR
45. to pay/buy *ba:yaD
46. person *ta:Øuh
47. pestle *haqlu
48. pole-bamboo *tekén
49. pus *n[a,e]:naq
50. rayfish *pa:Ri
51. to return (home)
    *[Ø]u(:)liq
52. rice-husked *beRás
53. rice-plant *pa:jey
54. rice-seed *benhiq
55. ridge(beam) *bubuŋ(an)
56. road/trail *Da:lan
57. roof (thatch) *qatép
58. sail (boat) *la:yaR
59. sesame *leŋá(h)
60. seven *pitú
61. to sip-noisily *hi:Rup
62. six *Øeném
63. stairs/ladder *haR(e)Dan

64. stone *batúØ
65. tear(drop) *lu:heq
66. ten [unit] *-pu:luq
67. thou *(Øi)ka(Øu)
68. thousand *Ri:bu
69. three *telú
70. thy *-mu
71. tongue *di:laq
72. tree/wood *ka:huy/*ka:yuh

73. trunk/base
    *pu:nuq/*pu:qun
74. two *Dewha
75. umbrella *pa:yuŋ
76. vein (blood) *quRát
77. we [exclusive] *-kamí
78. we [inclusive] *(ki)tá
79. widow(er) *ba:lu
80. yam *qu:bi

GROUP TWO

Note. Although the etyma listed below are widely distributed throughout the major Philippine subgroups, many uniques and forms of limited distribution are often found in some areas. These cases are marked by the symbol ++.

1. armpit *ki:li[-2], *qi:Dek
2. bait *paØen, *[]umpan ++
3. betel-chew/quid *mamaq-en,
   *tilad ++
4. to chase-away [animal] *bu:Raw,
   *(t)a:buR
5. comb--fine tooth [for removing
   lice] *su:jud, *sa[g]ú
6. to cook [in general] *lu:tuq,
   *[Øh]apuy
7. to cover [as jar] *takép,
   *tak(a)lub
8. debt *qu:taŋ ++
9. deep *da:lem ++
10. to delouse *hiN(k)utuØ, *sukay
11. to drip/leak from (roof)
    *tu:Duq, *teDteD
12. ear *tali:ŋa, *tulí
13. field/swidden *qumah ++
14. to itch *katél ++
15. lake *Danáw, *lebéŋ
16. loincloth *baháR ++
17. mortar *lesúŋ, *bayu-an
18. mote-in-eye *pu:liŋ ++
19. my *-ku, *-ken
20. needle *(ke)Da:Rum ++
21. orphan *[]uli:la, *[]i:lu
22. particles of food stuck be-
    tween teeth *tiŋáh,
    *[]iŋát ++

23. pig *ba:buy ++
24. pot--cooking *ku:Den,
    *ba:ŋaq
25. to pound-rice *bayúØ,
    *lebék
26. rafters *kasáw, *kilú
27. ricestraw *DaRa:mi ++
28. right(side) *(ka)wanan,
    *(kaliN)tuØu
29. ring *siŋsiŋ ++
30. root *Ra:mut, *dalid
31. salt *qasín, *timús
32. sky *la:ŋit ++
33. sole (of foot) *dapan,
    dapadapa ++
34. soup *sabáw ++
35. span [c. 8 inches]
    *da:ŋaw, *da:ŋan
36. spirit/anito *qani:tu ++
37. spouse/wife *qasa:wa ++
38. to steal *ta:kaw ++
39. still/yet *pa- ++
40. sugarcane *tebúh, *qunás
41. tail *Øi:kuR, *[]i:pus
42. termite *Øa:nay ++
43. thin [object] *ni:pis,
    *yapit
44. unripe/green *hiláw,
    *[][ae]ta

45. vapor/steam *(qali)seŋáw ++
46. to vomit *sukaØ, *qu:taq
47. to weave [cloth] *habel,
     *t[ei]num
48. to weave [mat] *la:ja,
     *Øañam
49. to winnow, *tahép ++
50. ye *kamú, *kayú

GROUP THREE

1. to adhere/stick-to *dekét
2. areca-nut *bu:ŋa, *buwaq
3. ash(es) *qabúh
4. to blow *heyúp
5. charcoal *[]ujiŋ
6. to choose *pi:liq
7. coconut [gen] *niyúR
8. day(time) *qaljaw
9. earthquake *li:nuR
10. egg *qit(e)luR
11. eyebrow *ki:Day
12. far *-Dayúq
13. father *[]ama[q,h]
14. heavy *beRqat
15. house *baláy
16. leaf *Da:hun
17. lungs *ba:Raq
18. neck *li:qeR
19. new *baqRu
20. nose *q[V]júŋ
21. palm (of hand) *pa:laj
22. to plant *taném
23. raft *Rakit,
     *[Øq]a[r]u[r]
24. rattan *quwey
25. rib(s) *Ru:suk
26. to ride *sakáy
27. star *bitu:qen/*bitewqen
28. torch *sulúq
29. wall *diŋdiŋ
30. what-you-may-call-it
     *kuØa(n)

GROUP FOUR

1. afraid *ta:kut, *haldek
2. to answer *tebaR, *sabát
3. bat [animal] *kabe[g],
     *p[ae]niki(q)
4. black *qitém, *ŋi:tit
5. carabao *ka[r,R](a)baw,
     *qenuwaŋ
6. child *Øanák, *[Øq]uŋaq,
     *[]ebiŋ
7. cotton *[kg]ápes
8. fat (adj) *tabéq, *tambek
9. fence *[]a:laD, *la(m)bat
10. floor *saléR, *daqtaR
11. fruit *bu:ŋa
12. to give *beRey
13. hair [head] *buhék
14. how many? *pijá(h)
15. husband *ba:nah
16. to know (person) *kila:la
17. lightning *kilát
18. man/male *-la:ki
19. mother *[]ina[qh]
20. penis *qu:tin, *bu:tuq
21. rain *quDán
22. sea/ocean *Da:Rat
23. shadow *Øali:nu
24. shoulder *qaba:Ra
25. to suck *sepsep
26. tooth (ŋ)i:pen
27. turbid *lebéR
28. water *Danúm, *tu:biR,
     *wahiR
29. white *putíq
30. to win/defeat *daØe[g]

31. winnowing-basket *ni:Ru
32. woman/female *+ba:Øi

GROUP FIVE

1. alive/to live = exist *bu:hay,
   *bi[h]áR, *quya[g]
2. to arrive *Datén, *qabút,
   *sampet
3. arrow *panáq; see Tg palasóq,
   túnod
4. to bathe *Di:Rus, *Di:Ruq
5. banana [gen vs spec]
   *sa:[g]iŋ, *baRát, *punti
6. belly *t[e,i]yán
7. blanket *qules, *habel,
   *ku:mut
8. breast *su:su
9. to bury/inter *lebéŋ,
   *taném
10. canoe *baŋkaq, *ba[r]útu,
    *ba[r]aŋay, *[]abaŋ
11. comb [general, not fine-
    tooth for lice] *sagaysay,
    *suwat, *sulday
12. to count *bi:laŋ, *heyáp
13. crow *[q]uwák, *ga:yaŋ
14. deaf *beŋél, *beŋég, *tuleŋ
15. to dream *taR(a)qi:nep
16. to dry in sun *beláj
17. eyelashes *pi[D]ek, *kimat
18. eggplant *ta[r]u[m,ŋ]
19. fire *[Øh]apúy, *kala:yu
20. to float *letáw
21. flower *bu:Dak, *-sabuŋ
22. fly [insect] *la:ŋaw, *lalej
23. frog *tukak, *+bak,
    *p(al)akáq
24. full/satisfied *besúR, *biyaR

25. to hear *DeŋéR, *kineR
26. heart *pu:suq
27. to hide/conceal *ta:Ruq
28. to kick *si:paq,
    *sik(y)ad
29. knee *tu:hed, *lu:luj
30. left(side) *wiRi,
    *walá(h), *gibaŋ
31. medicine *[]a:gas,
    *tambal, *bulúŋ
32. pain/ache *sakít
33. to sew *tahíq, *+Da:qit
34. shark *qi:hu, *ba[g]ís,
    *patíŋ
35. sharp *taDém
36. slave *qeDi:pen,
    *suRuq-en
37. smoke *qasú, *[]asuk,
    *[]ebel, *[]anus
38. sour *qalsem
39. string *lu:bid, *pi:siq
40. thigh *pa:qa, *[]ulpu,
    *payaŋ
41. thorn *Du:Ri, *tenék,
    *sebit, *si[]it
42. urine *[]iheq,
    *qisbu/*siqbu
43. to wait *helát,
    *taRáD; Tg hintáy,
    antáy
44. wing *pakpak, *payak,
    *[]elad, *panid

GROUP SIX

1. above/over Tg sa itaás
2. across Tg sa kabiláq
3. and [NP conj] Tg at
4. ashamed *heyaq, *ba[q]ín

5. at [locative mkr] Tg sa
6. bad [in health]
   Tg samáq
7. bamboo [sp] *kawa:yan,
   *buluq

8. bird Tg íbon
9. to bite *kaRát
10. blind *butáh
11. blood *Da:Raq, *DuRúq
12. body Tg katawán
13. bone *tuqla[n,ŋ], Tg butó
14. bow [arrow] *bu:suR
15. to breathe *Rinha:wa,
    Tg hiŋá
16. to buy *belíh
17. to carry/bring *DaDá[hØ]
18. cheek *pisŋi
19. chest/bust *debdeb, *Deghan
20. clear [not turbid] Tg línaw
21. come-here! Tg halí-ka
22. companion Tg kasáma
23. dark *De(m)Dém
24. dew *hamquR
25. difficult Tg hírap
26. dog *[Ø]a:su
27. down/below Tg sa ibabáq
28. earlier *ka[q,n]i:na
29. eel *[]í[g]at, *kasi:li,
    *dalít
30. to fall *hu:luR
31. feather [gen] *baDahi:bu
32. to fetch-water *(sa)qegeb
33. fingernail *kukú(h)
34. first [particle] Tg mu:na
35. fish [n] *qis(e)Daq, *Øikan
36. to fly *lepád, *layug, *sayap
37. to go-to Tg puntá
38. guts/intestines *bitu:ka,
    *tina:qi
39. hand Tg kamáy
40. he/she *siyá
41. heel Tg sákoŋ
42. here [near me/us] Tg díto
43. his/her *-ña
44. how? [manner] Tg paanó
45. how much? [price] Tg magkáno
46. hungry *Rutem
47. to hunt *[q]anup
48. if/whenever Tg kuŋ
49. in(side) Tg sa loób
50. inland *(qi)Daya
51. later-on Tg mamayáq
52. to laugh *(ka)ta:wa
53. leg Tg paá; *sekí

54. to lie-down *hiD(e)Raq
55. light/bright Tg liwánag
56. light(weight) *Reqán
57. lip Tg lábiq
58. to live/dwell Tg tirá
59. many Tg marámi
60. [markers: common noun]
    Tg aŋ, naŋ
61. [markers: personal] Tg
    si, ni, kay...
62. [marker/linker] Tg
    na, -ŋ
63. mouth *baqbaq; Tg
    bibíg
64. near(by) *ma-Dapit
65. night(time) *RabiØi
66. none [existential]
    *waDáq
67. not: don't! [imperative]
    Tg huwág
68. not: not so [predicative]
    *bekén
69. not: will-not [verbal]
    Tg hindíq
70. now/already [completive]
    *na
71. oar/paddle *beRsay,
    *[gk]a[hq]ud
72. one [in series] Tg isá
73. our [exclusive] *[]a:men
74. our [inclusive] *[]a:ten
75. to put Tg *i-lagáy;
    CPH *betáŋ
76. to repeat Tg úlit
77. riverbank *paŋpaŋ
78. rope *ta:li(q)
79. seaward *(qi)lawéd
80. to see *ki:taq
81. straight *tuqlid,
    *taldeŋ
82. sun *qaljaw, *sega(h)
83. that [near you] Tg
    iyán
84. that [yonder] Tg iyón
85. their *niDa
86. there [near you]
    Tg diyán
87. there [yonder] Tg doón

88. there is [existential] *may
89. they *siDá
90. thick *kapál, *da(k)mel
91. this [near me/us] Tg itó
92. thunder Tg kulóg
93. today/now Tg nayón
94. tomorrow Tg búkas
95. under Tg sa ilálim;
      *(qi)Da:lem
96. to understand Tg unáwaq,
      intindi
97. up/top-of Tg sa ibábaw
98. what? *[]anú, *[]u:nu
99. when? [past vs. future?]
100. where? [whence? whither?]
101. who? *si[q]nu
102. whose? Tg kaníno
103. why? Tg bákit
104. worm: earthworm *qu:lej
105. worm: intestinal
      Tg buláti
106. year Tg taón
107. yes Tg oo, ohoq;
      *huØe(n)
108. yesterday *kaha:pun
109. young [person] *ba:taq
110. your [plural] *i(n)yu

GROUP SEVEN

1. abaca *la:nut, Tg abaká
2. abdomen *pusqun
3. acrid Tg paklá
4. afternoon *ha:pun
5. all Tg lahát
6. angry Tg galít
7. animal *ha:yep
8. ankle *bukubukú
9. anus *+but (*lubut, *butbut)
10. ant [gen; sp.] Tg langám,
      gúyam
11. to ask (request) Tg hiníq
12. to ask (question) Tg tanóng
13. back (anatomical) Tg likód
14. back (place) Tg hulí,
      likurán
15. bark (of tree) Tg balát
      nang káhoy
16. beard Tg balbas, *bungut
17. because Tg dahil
18. to begin Tg simuláq
19. to belch Tg digháy
20. betel-leaf *bu:yuq
21. between Tg sa pagítan nang...
22. big *Dakel(aq)
23. to blame *ba:sul
24. to boil [intr.] kulóq
25. boil/abscess *peRsa
26. bolo/knife Tg gúlok; *sundang
27. born Tg ipinangának
28. to borrow *heDám
29. brave Tg ta:pang
30. to break (bone) *ba:Diq
31. breakfast *pama:haw
32. bridge *teytey(an),
      *tuláy
33. to bring/escort *hatéD
34. broom [hard vs. soft]
      Tg walís
35. bunch (of fruit) *bu:liR
36. to burn-down *su:nuR
37. butterfly *[qk]alibangban
38. buttocks Tg puwít
39. calf (of leg) *betiØis
40. to call *ta:waR
41. camote/sweet-potato
42. to carry/transport
      *ha:kut
43. centipede *qal[e,u]hi:pan
44. chin [not 'jaw'] Tg
      ba:baq
45. to climb-up *panahik
46. cloud/raincloud Tg qu:lap
47. cloud (white) *paña:nu[j]
48. cockfight *sa:bung,
      *bu:lang
49. cockroach *[Ø]i:pes
50. coconut--young *beteng,
      Tg bu:ko
51. coconut--old *lahing

52. cold (to touch) Tg lamíg
53. to command/send on errand
    *su:Ruq
54. to cough *[Ø]ebúh
55. to crush-lice *te[d]és
56. to cry/weep *ta:ŋis
57. cup (native) = bowl *maŋkuk,
    *piŋgan
58. to cut/slice (meat) *hi:waq
59. deer *qu(R)sá
60. to defecate SPH*[]inta:lun
61. to dig *kalih, *kutkut
62. dirty Tg marumí
63. to drag *[gR]u:yuD
64. to drift with current
    *qañud
65. to drizzle Tg qumambón
66. to drown *lemes,
    *l[u,e]n[u,e]D
67. drunk/intoxicated
68. dry *majá; Tg tuyóq
69. dull (blade) Tg puról
70. dust Tg qalikabók
71. earth (soil) Tg lu:paq
72. earwax *-tulí
73. east Tg sila:ŋan
74. easy [not difficult] Tg
    dalíq, luwág
75. face Tg mukháq
76. famous *bantu[g]
77. fast [to run-] Tg mabilís
78. few Tg kaquntiq
79. to fight *[]a:way
80. finger Tg dali:riq
81. to finish (off) *ta:pus
82. flood *baháq
83. to flow Tg qa:gos
84. fog/mist Tg qu:lop
85. to follow *su:nud
86. foot *sekí, Tg paqá
87. forehead Tg noqó
88. forget *lipat, *limut
89. front (of) *qaDep
90. grass Tg damó
91. good/well/fine Tg mabu:ti
92. grandparent/grandchild Tg apó
93. green (color) Tg luntíq
94. to guard *bantay

95. hair--body Tg balahi:bu
96. hair--pubic Tg bulból
97. half *teŋáq
98. to hang--on peg *saqbit,
    *kabqit
99. to hang--by rope
    *bi:tay, *bi:tin
100. to hang--over [fence,
    rope] Tg sampáy
101. happy Tg masayá
102. hard [substance] Tg
    matigás
103. headcold/mucus *sipqun
104. to hiccough Tg sinók
105. to hit/box *suntuk
106. to hit the mark
    *ta:maq, *[]iRúq
107. hoarse *paRáw
108. to hold [general term]
    Tg háwak
109. horn [of animal]
    *su:ŋay
110. hot (to touch) *qi:nit
111. husk of rice *qepá
112. industrious Tg sípag
113. in-law *bi[r]ás
114. in-law: brother *bayáw
115. in-law: child Tg
    manúgaŋ
116. in-law: co-parent
    *bala:Øi
117. in-law: parent Tg
    biyanán
118. in-law: sister *hi:paR
119. jaw [not 'chin'] Tg
    pañá, *selán
120. to know (how) Tg
    marúnoŋ
121. to know (fact) Tg alam
122. ladle [gen. utensil]
    *luwá[g]
123. ladle [coconut shell]
    *sanduk
124. lame *piláy
125. to learn *[Ø]a:dal,
    *tu[q]en
126. to lie/deceive Tg
    sinuŋálin

127. lonely Tg maŋláw
128. long [object] *h(al)abaq
129. long [time] Tg tagál
130. to lose [something] Tg nawaláq
131. to love [someone] Tg íbig
132. maiden/young lady *DaDa:Ra
133. mat [sleeping] *baní[g]
134. meat/flesh Tg lamán
135. middle (of) Tg sa gitnáq
136. milk *[g]a:tas
137. molar-tooth *baRqaŋ
138. monkey [gen; spec]
139. morning *[Ø]a:gah
140. mosquito *ñamúk
141. mountain *bu:ki[dj]
142. to move [not be still] Tg galáw
143. mud Tg pútik
144. nape of neck Tg bátok
145. narrow [NB Tg kítid vs. kípot]
146. to need Tg kailáŋan
147. nest [of bird] *pu:[g]ad
148. nest [of beast] *DeR(u)mun
149. niece/nephew *pag-um-anak-en
150. nipa [palm *ni:paq; frond *pawed]
151. nipple/teat *[]utéŋ
152. noon *qu[R]tu
153. numb *banhej, *binhej
154. octopus *[k,p]uRi:ta
155. old [person] *[R]u:Daŋ
156. old [thing] *da[]an, *dati
157. to open-up [package] *bukás
158. to order/leave-behind *bi:lin
159. other/different *[Ø]ibáh
160. parent Tg magúlaŋ
161. to pass-by Tg daán, CPH *[]a:gi
162. pillar [of house] *haDi:Ri
163. pillow *qul(u:)nan
164. placenta *qinul(u:)nan
165. to play Tg laróq
166. to point (to/at) *tulduq
167. pregnant Tg buntís
168. to pull Tg híla
169. to push *tu:lud, *tu:lak
170. quiet/silent Tg tahímik
171. rainbow Tg bahaghâriq

172. rat/mouse Tg dagáq
173. red *pu[l]á
174. to regret/repent *selsel
175. to remember *tandaq, Tg ala-ála
176. to revenge/repay *ba:les
177. rice--cooked *kaq(e)nen
178. right/correct Tg totoó
179. ringworm/herpes *buqni
180. ripe *hinú[g]
181. river Tg ílog; *subáq, *sa:luR
182. rotten [as tree] Tg gapók
183. rotten [as egg] Tg bugók
184. round *bilu[g]
185. to rub/massage *hi:lut, Tg hágod
186. to run Tg takbó; SPH *dala:gan
187. sad Tg luŋkót
188. saliva *la:way
189. saltless *tabqaŋ
190. salty *[]a:Dat
191. sand *buha:ŋin, *ba[j]ás
192. sap/resin *sa:leŋ
193. to say [He said...] Tg nag-sábi
194. scab Tg laŋíb
195. scar *pi(gk)lat
196. to scratch [an itch] Tg kamót
197. seed [of fruit] Tg butó
198. to sell Tg ipagbili; SPH *baligyaq
199. shade [of tree] Tg lílim
200. shallow [water] *ma-ba:baw
201. to shave/scrape *kiskis
202. sheath(e) *ta[g]éb; Tg kalúban
203. shelter/underhouse *si:Duŋ

204. shore [of sea] *baybay
205. short [person] *pandak
206. short [thing] Tg ikliq vs. mabábaq
207. to shout Tg sigáw
208. shrimp *quDáŋ
209. sibling [relation] Tg mag-kapatíd
210. sibling--elder Tg kúya vs. áti
211. sibling--younger *maŋ(u)hed
212. to sing Tg áwit
213. to sit Tg upóq; SPH *[l,q]iŋkud
214. skin [animal vs. fruit] Tg balát
215. skinny [person] Tg payát, *ni:waŋ
216. to sleep *tu:DuR
217. to sleep-with *hu:lid, *duDúg
218. small [object/child] Tg maliít
219. to smell/sniff Tg umamóy
220. smell--bad odor *ba:huq
221. smell--fragrant *baŋlu
222. smooth Tg makínis
223. snake *ha:las
224. soft Tg lambót
225. to speak [language] Tg salitáq
226. spider Tg gagambá
227. to spit Tg luráq
228. to split [wood] Tg Sibák
229. spring [of water] Tg bukál, bátis
230. to squeeze *peRáq
231. to stab Tg saksák
232. to stand *tindeR
233. storm *baR(i)yuh
234. story Tg bída, kuwento
235. to string-together [beads] *tu:huR
236. strong [person] Tg malakás; *keseR
237. summit/top *tuktuk

238. surprise(d) Tg na-gúlat
239. to swallow Tg lunók, lulón; *telén
240. sweat Tg páwis
241. sweet *tamqis, *hamis
242. to swell *bu:kel
243. to swim *laŋúy
244. to take *ku:haq
245. to take/snatch *[ ]a:gaw
246. tall/high *ta[ ]as
247. taro/root-crop Tg gábi
248. to taste Tg tikím
249. to tell *su:gid
250. to think *[ ]i:sip
251. thirsty *qu:haw
252. throat Tg lalamúnan
253. to throw [as ball] Tg hágis
254. to throw-away Tg tápon
255. to tie/bind-up Tg táliq, gápos
256. tired Tg pagód; *kapuy
257. to turn/revolve Tg umíkot
258. turtle [land] Tg pagóŋ; *baququ
259. turtle/tortoise *pawi:kan
260. twenty Tg dalawampúq
261. to use *[g]a:mit
262. vagina *pu:ki
263. vegetables Tg gúlay
264. waist Tg baywáŋ
265. to walk *lakád, *lakáw, *pa:naw
266. to wash/rinse-off *hu:Ras
267. to wash-clothes Tg labá
268. to wash-face *hiraqmus
269. to wash-hands *henáw
270. waterfall Tg talón; MPH *besáy
271. waterjar Tg tapáyan, kalambáq
272. wave [of ocean] *qa:lun, *humbak
273. weak Tg mahínaq
274. to wean [baby] Tg wálay, áwat
275. to wear [clothes] *suqlut
276. west Tg kanlúran
277. wet *basáq

278. to whet *ha:saq
279. wide Tg lápad vs. luwáŋ
280. wind *ha:ŋin
281. wind [south] *ti:muR
282. wind [seasonal] *qami:han
283. to wipe *pa:hid
284. woods/forest Tg gúbat
285. to work/make/do Tg gawáq
286. wound/injury Tg súgat
287. to write *su[Rr]at
288. wrong/error *saláq
289. to yawn Tg higáb, hikáb
290. yellow *deláw, *diláw

5.1. NOTES ON ORDERING OF A LIST. Thus far we have groups of glosses, rated for the most part by the number of etyma we find throughout the Philippines, and for the likelihood of revealing important information of one kind or another. The lists presented here are not intended as a ready-made questionnaire for actual fieldwork, although for some kinds of work they may be useful as they stand. For example, if one was undertaking a language subgrouping, he might best start with group six, and work backwards through groups five, four, three, two, and one; then upwards from seven.

However, for convenience here, these groups have been presented in alphabetical order. It would be more convenient in research to have the glosses arranged according to a more logical order, perhaps the most desirable being within semantic ranges. For example, since the numerals occur mostly in group one, it would be best to elicit them in sequence, whenever it was thought best to get them-- they are not logically our first choice for every survey. Nor does one want to stick rigidly to these groups. Even if most antonyms do not occur within the same group, it is desirable from the point of view of elicitation and collation to juxtapose them on a list (viz., 'bitter-sour-sweet', 'heavy-light', 'black-white-red-yellow-green', etc.). Likewise, morphological elements, like the nominative pronouns, deictic pronouns, oblique (genitive) pronouns, etc., are best elicited in sequence. Thus, a finished list, ready for a field project, would best contain semantically grouped sets, with illustrative sentences in many cases to set the gloss in context, and to economize by eliciting related forms as well, such as 'He split the wood', 'The banana is acrid', 'There are two houses over there', etc.

The following are examples of questionnaire types,
one or more of which will be useful to a researcher
if prepared before fieldwork actually begins:

1. A list alphabetized according to glosses in
English or the researcher's native language as a
quick cross-reference to etyma, to other related
glosses, and perhaps also to semantic codes and to
the number of known etyma.

2. A list organized semantically for ease and
naturalness in elicitation, including (the above-
mentioned) illustrative sentences for key or
ambiguous items.

3. A grammatical questionnaire of sentences pre-
arranged to elicit full paradigms: nominative (topic)
pronouns, enclitic genitive pronouns, dative pronouns,
pronominal deictics (demonstratives), locatives,
temporals, case-marking particles (for common and
personal nouns), negatives, verb-aspect, tense, and
voice-/focus-marking affixes or particles, adjective
comparisons, adverbial and conjunctive uses, etc.

4. A list of etyma, arranged alphabetically for
easy reference.

5. Lists of problem reflexes for comparative
work, such as the initial, intervocalic, and final
occurrences of *R, *D, *y, *h, *j (non-initial), *q,
etc., or the appearance of *e in the ultima, penult,
and antepenult, etc.

5.2. TWO GLOSSES AS AN EXAMPLE OF THE AIM OF ONE
KIND OF LIST. The glosses 'blood' and 'water' are
drawn from groups six and four respectively. The
distribution of the forms translating these glosses
gives a significant amount of information, as can be
seen in Tables 5, 6, and 7. Even where a single
etymon is reflected, the particular reflexes may be
helpful in establishing further subgroups. Thus, the
sound shift *R → y is seen in the first languages
listed (Yami [5]/Itbayaten through Alangan). The loss
of final *-q separates the Cordilleran languages
from the others. These Cordilleran languages can be
further subdivided by the sound shift *R → r, or l,
or g. Innovations for 'water' separate the Ifugao
group, and the Alangan-Iraya group from the others;

innovations for 'blood' set the Kallahan, Ilongot,
and Kankanay languages apart from each other and all
of the others. (See Table 5.)

TABLE 5:   *'blood' = *Da:Raq, 'water' = *Danɨm*

in Northern Philippine languages

Yami	ralaq*		ranɨm		
Itbayaten	rayaq		ranum		
Ivatan	rayaq		danom*		
Sambal	dáyaq		lanɨm*		
Pampango	da:yaq		danɨm		
Iraya	dáyaq				sápaq
Alangan	dáyaq				sápaq
Ilokano	dára		danɨm		
Itneg	dála		danɨm		
Balangao	dála		dɨnɨm		
Bontok	dála		danóm		
Ifugao/Byn	da:la				le:teng
Ifugao/Amg	dála				liting
Ifugao/Btd	da:la		danum ∿		leteng
Kallahan/Kyp	dala ∿ kuhiyaw		danum		
Kallahan/Kly		kuheyaw	danum `		
Pangasinan	dalá		danɨm		
Ibaloi	cala		canom		
Agta	dɛgɛ		dɛnum		
Atta	da:ga		danum		
Isneg(Reid)	da:ga		danum		
Isneg(Vanb)	dáxa		danɨm		
Ibanag	dága		danɨm		
Itawis	dága		danɨm		
Dumagat/Cas	digéq		dinóm		
Gaddang	qada		danum		
Ilongot		matgɨm	dɨ:num		
Kankanay		basá	danóm		

Languages that are genetically remote from the
other languages of the central and southern Philip-
pines, such as Samalan and Sangir, reflect the
etyma *Da:Raq 'blood' and *wa:hiR 'water'. (See
Table 6.)

TABLE 6:   'blood' = *Da:Raq, 'water' = *wahiR
           in Southern Philippine languages

[Malay]	darah	air	
Abaknon	lahaq	bwahiq	
Samal	lahaq	bohèq	
JamaMapun	lahaq	boheq	
Sibutu	lahaq	boheq	
Banggi	raah	beig	
Bajau	raha	bohe	
Tiruray	daraq	wayeg*	
Kadazan	zaa	vaig	
RungusDus	raha	vaiqig	
Sangir	daha		akeq
Tobatu	dahaq		akè
Bantik	daha		akè
Sangil	dàra		qàkeq

In the central and southern Philippine area, a
number of innovations are noted. These innovations
set off: the Meso-Philippine languages (*DuRúq and
*tu:biR), the Palawan languages (*DuRúq and *Dánum;
including Hanunoo[6]), the Kalamian languages (*tageq
< PPH *getaq, and *wa[]ik), Muslim languages of
Mindanao (*DuRuq and *[]iR), Northern Manobo (*langesa
and *wahiR), Southern Manobo (*dipanug and *wah:iR),
and Bilic (*Ritaq and *[w]a[]εR). Gorontalo/Bunda
and Bentenan/Sahasara are included to show the further
distribution of both *DuRuq and *leng[]sa 'blood'.
A few other languages (Buhid, Tadyawan, and Batak)
exemplify the variations and difficulties that can
be encountered on even these two glosses. (See
Table 7.)

    Although this example is rather simplistic, it
illustrates the kind of qualitative information one
may want to obtain from the first glosses on a
specialized wordlist, aimed at subgrouping newly
discovered speech varieties.

TABLE 7:   'blood' and 'water' in the Central and Southern Philippines

	*DuRuq 'blood'		*Danum	*tu:biR	*wahiR	'water'
Tagalog	dugʉq			tu:big		
Bikol/Naga	dugʉq			tu:big		
Bikol/Iriga	rugʉq			tubig		
Bisayan/Ceb	dugʉq			tu:big		
Bisayan/Kin	dugʉq			tu:big		
Tausug	dugʉq			tubig		
Kamayo	dugʉq			tu:big		
Mansaka	duguq			tubig		
Kalagan	duguq			tubig		
Mamanwa	duguq					sapaq < *sapaq
Subanon/Sin	duguq			tubig		
Subanon/Soc	duguq			tubig		
Mongondow	duguq			tubig		
Ponosakan	duhuq			tiwig		
Aborlan	dʉguq		danʉm			
Palawano	dʉguq		danʉm			
Hanunoo	dugʉq		danʉm			
Buhid		fʉlut	danʉm			
Tadyawan		pilit				lɨbɨng < *lebeng
Batak/Pal		tagɨk	danʉm			
Kalamian	duhuq*∿tahɨk				waiq	
Nor.Tagbanwa	duguq*∿tagɨk				waiq	
Agutaynen	duguq				waiq	
Maranao	rogoq				qig	
Magindanao	luguq				qig	
Gorontalo	duhu					taluhu / talugu < *saluR
Bunda	dugu					
Bentenan		malunsa				akè (cf.Sangir,
Sahasara/San		malensa				akèq Sangil, Tobatu)
Dibabawon		langɨsa			wɨhig	
Binukid		langɨsa			wahig	
West.Bkd.Man.		lɨngɨsa			wahig	
Ilianen		lɨngɨsa			wayig	
Tigwa		langɨsa			wɨig	
Ata		langosa			woig	
Kagayanen		langɨssa			waig	

	*DuRuq 'blood'	*Danum	*tu:biR	*wahiR 'water'
Kal-Cotabato		dɨpanug		wayɨg
Sarangani		dipanog		wayɛg
Bagobo		dipanog		weeg
Tagabawa		dipanog		waig
Tasaday		dafanug		wayeg
Bilaan/Kor	litɨq			qeqel
Bilaan/Sar	litɨq			yeqel
Tagabili	litoq			qel
Tiruray	gitoq*			wayeg*

5.3. SELECTING QUALITATIVE (LESS RETENTIVE) GLOSSES.

In the initial stages of research on this topic, I selected ten glosses in an attempt to subgroup the languages of northern Luzon for which I had available information. Selection was made on the basis of cognate sets that had a limited distribution (no more than four etyma) among Cordilleran languages, although such distribution did not necessarily have to be limited to Cordilleran languages. The glosses (with tentative reconstructions) listed in Table 8 were chosen.

TABLE 8: *Ten glosses and their distribution in Northern Luzon*

banana	*dup(e)t, *báRat
body	*b(e)gi, *()adél, *bakDang, *()áwak
cheek	*padingil, *tamíl, *()apíng
chest	*palagpag, *pagákew, *balukeng, *badang
companion	*kabulun, *()íba, *ka-du()a, *(kg)a(q)it
deaf	*bengég, *túleng, *púkit
difficult	*lígat
eel	*dalit, *()igat, *kíw(e)t
heavy	*damet, *dagsen, *beRqat
heel	*túmang, múkud

On the basis of a comparison with this brief list on
the Cordilleran languages found in Reid (1971) and
on three others for which I had information, I com-
puted the scores found in Table 9. Now, I do not
propose that such a small list has validity in sub-
grouping the languages of northern Luzon,[7] or any
languages of the Philippines. I do wish to show how
qualitative items, selected with care, can show us
something of the relationship of languages. If, for
example, I had chosen any ten of the following

*TABLE 9: Scores of Northern Luzon languages on
the basis of 10 less-retentive glosses*

Agta

  7 Atta

  7 6 Gaddang

  5 5 6 Isneg -- Yogad (6)

  4 3 5 3 │ Itneg [Tinguian] -- Ilokano (9)

  1 1 1 3 │ 7 Kalinga

  2 2 2 2 3 3 │ Ifugao [3 dialects]

  1 1 1 2 4 │ 6 7 Balangao

  1 1 2 1 4 5 4 │ 6 Bontok

  1 2 2 2 4 5 │ 6 7 8 Kankanay

  1 2 1 2 3 3 4 4 4 │ 6 Ibaloi--
                                            Pangasinan (4)

  1 1 1 1 3 4 5 │ 6 8 8 7 Kallahan
                                           [Keleyqiq]

Ungrouped by these criteria: Casiguran Dumagat, Ilongot.

glosses, I would be obscuring the interrelationships
among the languages, because, with few exceptions,
most of the Northern languages reflect the same
etymon for each gloss: 'all' *(ng)aqmin, 'bitter'
*paqít, 'bone' *tuqlang, 'brain' *[]útek, 'carabao'

*[qe]nuáng, 'chicken' *manúk, 'to choose' *píli(q),
'die' *n-atáy, 'dog [q]ásu, 'to drink' *[]inúm, 'to
eat' *maN+(k)án, 'elbow' *síku, 'eye' *matá, 'fat'
*tabá(q), 'father' *[]áma, 'fire' *[]apúy, 'four'
*[]epát, 'full' *na+penú(q), 'head' *[]úlu, 'heart'
*púsu(q), 'louse' *kútu, 'lime' *[]ápuR, 'man/male'
*-láki, 'moon' *búlan, etc.

If we emphasize the agreement of cognate sets
that are more likely to change or undergo innovation,
rather than the agreement of cognate sets that are
more probably retentive, we will have qualitative
criteria for subgrouping. A list made up of such
glosses would score more strongly than any other
list made up of contentives.

6. STATISTICAL EVIDENCE FOR THE VALIDITY OF THE
SEVEN PRIMARY GROUPS. A preliminary version of the
seven groups was subjected to statistical analysis
by Curtis McFarland on his available data on Bikol
dialects. From his computations, we get some
statistical verification of the relative accuracy
of each successive group in giving glosses that are
less universal and less retentive. The following
are the figures:

GROUP	NUMBER OF GLOSSES USED	NUMBER OF GLOSSES WITH ONLY ONE COGNATE SET	AVERAGE NUMBER OF FORMS PER MEANING	PERCENTAGE OF GLOSSES WITH ONLY ONE COGNATE SET
1	41	36	1.1	87%
2	15	11	1.6	73%
3	22	15	1.5	68%
4	17	13	1.3	76%
5	20	13	1.5	65%
6	25	12	2.2	48%
7	117	31	3.8	26%

There is no difference in the overall results of a
comparison of groups three, four, and five. Since
the groups were made on the basis of the distribu-
tion of cognate sets in eight major subgroups

throughout the entire Philippines, one would not ex-
pect to find much diversity of forms in a small sub-
group like Bikol, when the entire Philippines itself
has only three to five reconstructible etyma. In a
group with a higher order of diversity (like Manobo
or Igorot) one would expect a slow gradation of
the figures, from a high percentage of glosses with
only one cognate set for group one, down through a
very low percentage for group seven. Conversely,
one would expect a constantly increasing figure for
the average number of forms per meaning generated
by groups one through seven. Such statistical
evaluation should prove to be the best criticism of
the proposed groups. In the above data for Bikol,
there is the expected high percentage for group
one, and the significantly low percentages for
groups six and seven.

7. CONCLUDING REMARKS. This study has concerned
itself with the principles behind drawing up a
Philippine wordlist. Several groups of glosses
have been rated for the likelihood of eliciting
innovations as opposed to retentions. Seven basic
groups have been formed; the first five are based
on the number of cognate sets found throughout the
Philippines, and on the probability of a gloss
eliciting a qualitatively useful form. No matter
what the goals of fieldwork may be, a job-specific
list is generally more productive than a long list,
be it arranged alphabetically or semantically. A
list can be selectively arranged from the glosses
discussed here in such a way so as to give the
highest informational value, depending on one's
survey.

    A future definitive Philippine wordlist will
depend on: (1) the refinement of the principles
discussed herein, (2) a more accurate genetic
classification of known Philippine languages, (3)
statistical verification of the seven basic groups,
and (4) the cooperation of fieldworkers and linguists
in critically assessing the glosses and forms pro-
posed here, and in their contributing additional
candidates to supplement or replace those presented
here.

    I have not discussed either the dangers or the
benefits of using vocabulary for various purposes.

While it is true that lexical items are the simplest
to elicit, nothing can replace the careful recording
and transcription of actual speech. Nor can con-
tentives give the qualitative evidence found in the
agreement of functors.

A principal advantage of vocabulary study is that
the lexicon is an open system, while the syntactic
and morphological elements are part of a more closed
system. If one is dealing with great time depths
that have obliterated most of the agreements in
functors that may have existed between two speech
types, then vocabulary becomes the only means left
for comparative work. Such evidence is not easily
obtained or ranked. A very large corpus of data
will be necessary, with careful collation, culling
out possible borrowings, and searching as thoroughly
as possible to ascertain that putative innovations
are not just mutual retentions from an early meso-
language. Thus, the openness of lexicon is both an
advantage and a pitfall. Considering further that
in-depth vocabulary study is time-consuming and
that nevertheless the rewards can be few, the reader
should be advised that the enthusiasm that I express
for this kind of study is tempered by a personal
understanding that footnotes of caution should be
present throughout this paper.[8]

NEW HAVEN

NOTES

I am grateful to my wife, Nellie, for all of the typing she has done, and for her understanding.

This revision of the paper owes much to discussions with Benedict, Grace, Laycock, Milner, Prentice, Reid, and Tchekhoff during the course of the conference. I wish to express my gratitude to all of them for their advice and encouragement. The additional 134 items in the present wordlist are largely the result of the time and efforts of Shigeru Tsuchida, who arranged the original list of 500 items semantically, so that the many gaps became apparent.

[1]Several discussions at the conference underscored the need for similar enlarged lists in other Austronesian areas. Much of this Philippine-oriented list will undoubtedly have appli- cation, provided editing and addition are done by the various area specialists or researchers. Laycock (1970) has compiled and rated a list of 357 entries, complete with a full com- mentary on the provenance, the difficulty in eliciting, and a technique for eliciting each gloss.

[2]I am indebted to Charles Walton of the S.I.L. for first suggesting this intriguing idea to me. He suggested that words like *matá and *di:laq told us almost nothing about a speech variety, except that it was Philippine. My initial reaction was to reject his statement, but the thought it provoked led to the drafting of this paper.

[3]In working with Tausug I had difficulty in eliciting the sentence 'The bird landed on the branch' because there apparently were different words for 'to land, perch' depending on the "status" of the bird. My informant asked me if it was "a royal bird or an ordinary bird that landed." The trans- lational equivalents in English were similar in style to the difference between 'The bird alighted on the branch' as opposed to 'The bird plopped down on the branch'. In the case of Tausug, however, a real speech level distinction seemed to be operating in that (in the instance) there was no generic term for 'to land'.

[4]Omission would be due to an oversight on my part, unless it has been discussed as problematic in this paper, such as 'freeze', 'ice', 'stick', 'some', 'with', 'shake', etc. The Tsuchida revision of my original list has corrected a large number of omissions in my first list.

[5]The -l- in Yami ralaq represents a secondary change based on an analogy with inherited PPH *y → Yami l, e.g., IVT *sayap 'to fly', Yami semalap, Ivt, Itb sumayap; IVT *(h)ayam 'to

walk', Yami qumalam, Ivt qumayam, Itb humayam, PPH *kawáyan (bamboo) → Yami kawalan.

[6]See my paper for this conference, "Internal and external relationships of the Mangyan languages," wherein I propose the possibility of a grouping of Hanunoic (South Mangyan) with Palawanic (particularly Aborlan Tagbanwa).

[7]This subgrouping is apparently not that far from that which does exist among the Cordilleran languages. Conklin (in a personal communication) told me that the most serious errors are: the obscuring of a closer relationship of Kalinga with Ifugao, and of Ilongot with the Pangasinan/Ibaloi group; the over-emphasis of closeness between Kankanay and the Pangasinan/Ibaloi group. Borrowing has doubtless introduced other errors as well. Any approximation to the real situation underscores the need for selecting out less retentive meanings and using them (but in much greater quantity) in making a more or less definitive measure of interrelatedness.

[8]In particular, one must be chary in assigning any weight to numbers of agreements, which (scores) reflect at best a random search, and certainly not a real percentage of agreements between lexicons. Thus, to say that Tagalog and Bisayan share 21 innovations as opposed to a sharing of 62 innovations by Bisayan and Bikol is to give figures relative to one's research, not to the actual language relationship.

REFERENCES

Blanco Woordenlijst. 1931. Esser's list, Holle, ed. Blanco woordenlijst uitgegeven op last der Regeering van Ned.-Indië ten Behoeve van Taalvorschers in den Ned-Indischen Archipel. Batavia, Landsdrukkerij.

Conklin, Harold C. 1951. Lexical check list for Philippine languages. (Mimeo./Xerox.)

Dyen, Isidore. 1964. On the validity of comparative lexicostatistics. In: Proceedings of the 9th International Congress of Linguists, ed. by Horace G. Lunt, pp. 238-252. The Hague, Mouton.

Dyen, Isidore. 1965. A lexicostatistical classification of the Austronesian languages. Memoir 19, Research Center in Anthropology, Folklore, Linguistics. Baltimore, Waverly Press Inc.

_____, A. T. James, and J.W.L. Cole. 1967. Language divergence and estimated word retention rate. Language 43:150-171.

_____ and Curtis McFarland. 1970. Proto-Austronesian etyma constituting an Austronesian cognate finder list. Yale. (Mimeographed).

_____. 1973. Some cultural characteristics of the Proto-Austronesians. Paper presented at the XXIXth Congress of Orientalists, July 16-22, Paris, France.

Elkins, Richard E., ed. 1962. Word list. Summer Institute of Linguistics.

Ferrell, Raleigh. 1969. Taiwan aboriginal groups: Problems in cultural and linguistic classification. Institute of Ethnology, Academia Sinica Monograph No. 17. Nankang, Taipei, Taiwan.

Fox, Robert B., Willis E. Sibley, and Fred Eggan. 1965. A preliminary glottochronology for northern Luzon. Asian Studies 3(1):103-113.

Grimes, Joseph E. 1972. Alphabetized list of English glosses to entries in the Comparative Word List Data Base. In: The Comparative Word List Data Base Manual of Procedures. University of Oklahoma, Informative Science Series, Monograph No. 7. (Norman, Oklahoma 73069.)

Gudschinsky, Sarah C. 1956. The ABC's of lexico-statistics (glottochronology). Word 12:175-210.

INL (Institute of National Language). 1953. A composite vocabulary of Philippine languages. Manila, Bureau of Printing.

Laycock, Donald C. 1970. Eliciting basic vocabulary in New Guinea. In: Pacific Linguistic Studies in Honour of Arthur Capell, ed. by S. A. Wurm. Series C-No. 13. The Australian National University. Sydney, A. H. and A. W. Reed.

Llamzon, Teodoro A. n.d. Language questionnaire. Manila. (Mimeo.)

_____. 1971. Language questionnaire I, for handbook of Philippine language groups. Manila. (Mimeo.)

McFarland, Curtis. 1972. A comparative vocabulary of the Bikol dialects. Yale. (Mimeo.)

Pittman, Richard S., and Associates. 1953. Notes on the dialect geography of the Philippines. 3rd ed. Summer Institute of Linguistics. University of North Dakota.

Postma, Antoon. n.d. Test-vocabulary for Philippine languages. Mansalay, Mindoro. (Mimeo.)

Prentice, D. J. 1969. A word list for use in Borneo. Papers in Borneo Linguistics, No. 1. Pacific Linguistics, Series A, No. 20, pp. 1-8.

Ray, Sidney. 1911. The languages of Borneo. The Sarawak Museum Journal, Vol. 1, No. 1.

Reid, Lawrence A. 1971. Philippine minor languages. Word lists and phonologies. Oceanic Linguistics Special Publication No. 8. Honolulu, University of Hawaii Press.

Samarin, William J. 1967. Field linguistics. New York, Holt, Rinehart and Winston.

Summer Institute of Linguistics. University of North Dakota (SILUND). 1959. Comparative Malayo-Polynesian-English wordlist.

Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts. Proceedings of the American Philosophical Society 96:452-463.

_____. 1955. Towards greater accuracy in lexicostatistic dating. International Journal of American Linguistics 21(2):121-137.

Tri-Institutional Pacific Program. 1954. Linguistic questionnaire. Department of Anthropology, Yale University, New Haven, Conn.

Tsuchida, Shigeru. 1962. Questionnaire for
   Philippine basic vocabulary I.

_____. 1968. Linguistic questionnaire for
   Philippine languages. Tokyo. (Mimeo.)

Zorc, R. David. 1968. Information sheet and
   linguistic check list for Philippine languages.
   Kalibo, Aklan. (Mimeo.)

_____. 1971*a*. Proto-Austronesian worksheet.
   Cornell Univ., Ithaca, Dept. of SEA Studies.
   (Duplicated.)

_____. 1971*b*. Proto-Philippine finder list.
   Cornell. (Xerox.)

_____. 1971*c*. Proto-Philippine worksheet/
   grammatical questionnaire. Cornell. (Mimeo.)

You have printed the following article:

Towards a Definitive Philippine Wordlist: The Qualitative Use of Vocabulary in Identifying and Classifying Languages
R. David Zorc
Oceanic Linguistics, Vol. 13, No. 1/2, Papers of the First International Conference on Comparative Austronesian Linguistics, 1974: Proto-Austronesian and Western Austronesian. (Summer - Winter, 1974), pp. 409-455.
Stable URL:
http://links.jstor.org/sici?sici=0029-8115%28197422%2F24%2913%3A1%2F2%3C409%3ATADPWT%3E2.0.CO%3B2-G

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

References

Language Divergence and Estimated Word Retention Rate
Isidore Dyen; A. T. James; J. W. L. Cole
Language, Vol. 43, No. 1. (Mar., 1967), pp. 150-171.
Stable URL:
http://links.jstor.org/sici?sici=0097-8507%28196703%2943%3A1%3C150%3ALDAEWR%3E2.0.CO%3B2-C

Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos
Morris Swadesh
Proceedings of the American Philosophical Society, Vol. 96, No. 4, Studies of Historical Documents in the Library of the American Philosophical Society. (Aug. 22, 1952), pp. 452-463.
Stable URL:
http://links.jstor.org/sici?sici=0003-049X%2819520822%2996%3A4%3C452%3ALDOPEC%3E2.0.CO%3B2-P

Towards Greater Accuracy in Lexicostatistic Dating
Morris Swadesh
International Journal of American Linguistics, Vol. 21, No. 2. (Apr., 1955), pp. 121-137.
Stable URL:
http://links.jstor.org/sici?sici=0020-7071%28195504%2921%3A2%3C121%3ATGAILD%3E2.0.CO%3B2-2