Lexicography in Papuan Tip Cluster Languages:
Where will it lead?

David Michael Snyder
Summer Institute of Linguistics

1.   Introduction

Lexical studies have been the backbone of much comparative work on Austronesian languages. While many early comparative studies were based on little more than word lists, over the years a variety of more detailed lexicons have become available and later studies have made good use of them. As research into Austronesian languages continues, we need to ask ourselves both how we can best invest research resources to develop quality lexical studies, and how comparative studies can make use of this growing body of information to contribute to our understanding of language.

In order to begin to answer these questions for the Papuan Tip Cluster languages, this paper will first establish what lexical information is available and then examine how current comparative studies tend to make use of lexical information. Given that background, the paper will propose ways in which lexical research might support further comparative studies.[1]

2.   Current Status of PTC Lexicography

There are approximately fourteen language groups and 56 languages in the Papuan Tip Cluster.[2] Of these languages, twenty seven are either under study or have had some lexicographic work (beyond limited word lists) produced in the language. The nature and extent of each work varies considerably.[3]

---

[1] The intent of the author is to promote discussion, not predict (or worse yet, attempt to dictate) the direction of lexical work and comparative studies.

[2] See Ross (1988) for an excellent discussion of the languages of the Papuan Tip Cluster. Ross (p. 420) records that he "adopted [the term 'Papuan Tip'] from Lithgow (1976) who in turn derived it from Dyen (1965)."

[3] For a more detailed study of the early history of this region, see Dutton (1976), Lithgow (1976) and Taylor (1976).

Dictionaries are produced for specific audiences. The choice of audience will determine what types of information will be recorded in the dictionary. Some of the dictionaries that have been produced (such as for Muyuw) have been designed and published with local audiences in mind. Indeed David Lithgow (personal communication) has indicated that among the Milne Bay language communities, there is significant local interest in and prestige associated with dictionaries. This has resulted in several published dictionaries catering to this type of audience. Other published dictionaries (such as for Sud-Est) are linguistic field notes which have been published in the interest of disseminating information and stimulating academic research. Field notes reflect the interests and nature of the individual field research project.

Although the lexicographic work of PTC languages is not homogeneous, there is a significant amount of information available through either published documents or other relatively accessible sources. This presents interesting possibilities for comparative studies. The types of dictionaries available can be seen in the chart below. Further details can be found in the bibliography.

CHART OF DICTIONARY STATUS[4]

Language Group		
Language Name	Known sources of lexicographic work	Lexical information[5]
West-Central Group		
Doura		
Gabadi		
Kuni		
Mekeo	Desnoes, G. (n.d.) Jones, Alan A. (n.d.) Chung, Je-Soon and Jung-Ok Chung: field work since 1989	n/a n/a

[4] The language divisions are based largely on Ross (1988), with additions from Cooper and Lithgow (personal communication).

[5] The 'lexical information' column uses the following key words:

Word List - dictionaries that only give foreign equivalents for words

Simple - dictionaries which attempt definitions, but only rarely contain examples, cross references, grammatical information or the like

Complex - dictionaries which attempt definitions as well as contain examples, cross references, grammatical information, lexical sub-classes or the like
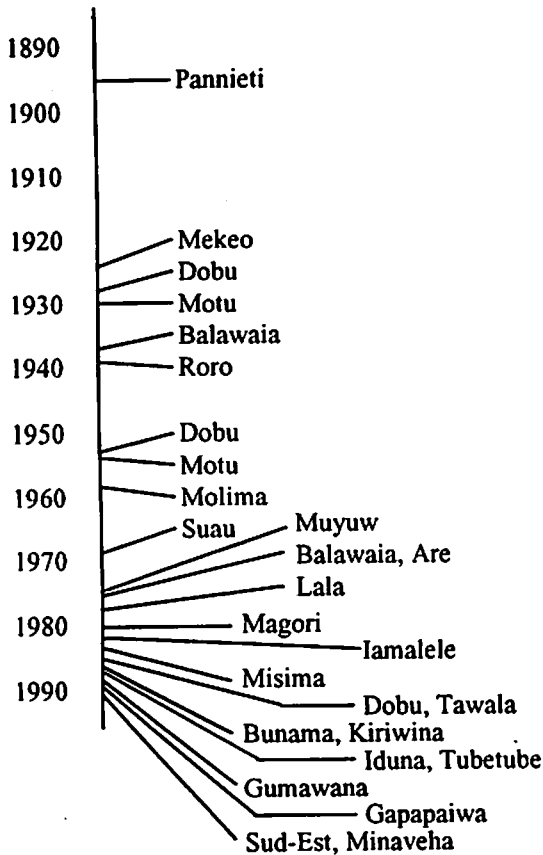
n/a - the author had only bibliographic reference to the document and could not classify it.

Motu	Lister-Turner, R. and J. B. Clark (ca 1954) 91 pp, 6370 entries	Complex
Nara (Lala)	Clunn, S. P. and J. A. Kolia (1977) 83 pp, 950 entries	Complex
Waima (Roro)	Coluccia, P. (1939) 569 pp Coluccia, P. (1941) 428 pp Kim, Nam-Su and Duck-Shin Kim: field work since 1988	Complex Complex
Mid-Central Group		
Keapara (Aroma)		
Sinagoro (Balawaia)	Koloa, M. and J. A. Collier (1937) 123 pp, 1350 entries Kolia, J. A. (1975) 19 pp, 1330 entries Tauberschmidt, Gerhard and Hiltrud Tauberschmidt: field work since 1985	Simple Simple
East-Central Group		
Bina		
Magori	Dutton, T. E. (1980)	n/a
Ouma		
Yoba		
Dobu-Duau Group		
Bosilewa		
Bunama	Lithgow, David (1985) 22 pp, 1500 entries	Word list
Dobu	Dixon, John (1928) 205 pp, 5215 entries Grant, R. V. (1953) 126 pp, 2500 entries, 2880 index entries Lithgow, Daphne (1984)	Complex Simple Simple
Duau		
Galeya		
Mwatebu		
Sewa Bay		
Gumasi Isolate		
Gumawana (Gumasi)	Olson, Clif and Roxanne Olson (1988) 2600 entries	Complex

Bwaidoga Group		
Bwaidoga	Gibson, Stan (1990) 1100 entries	Simple
Diodio		
Fagululu		
Iamalele	Semi, Nafitali et. al. (1988) 3000 entries, 3700 index entries	Complex
Iduna (Vivigani)	Lucht, Ramona (1987a) 257 pp, 6500 entries Lucht, Ramona (1987b) 349 pp, 8750 entries	Simple (Index)
Kalokalo		
Salakahadi (Molima)	Chowning, Ann (1958) 180 pp, 4750 entries Engkvist, Leif and Helena Engkvist: field work since 1990	Simple
Anuki Isolate		
Anuki		
Are Group		
Are (Mukawa)	Giblin (n.d.)	n/a
Ariama (Miniafia)	Wakefield, David: field work since 1973	
Boanaki		
Doga		
Gabobora		
Gapapaiwa	Giblin (n.d.) McGuckin, Ed and Catherine McGuckin (to appear) 1150 entries	Complex
Ubir		
Wataluma		
Taupota Group		
Garuwahi		
Kukuya (Minavega/ Minaveha)	Lovell, Larry (to appear)	Complex
Taupota		

Tawala (Kehelala)	Ezard, Bryan (1984) 299 pp, 2100 entries, 2600 index entries	Simple
Wedau		
Suau Group		
Auhelawa (Kurada)	Lithgow, Daphne: field work since 1987	
Buhutu	Cooper, Russ: field work since 1985	
Suau	Cooper, Russ (1969) 800 pp	Word list
Tubetube	Gunderson, Steve and Jerry Gunderson (1987) 141 pp Canavan, Alan and Faye Canavan: field work since 1987	Complex
Wagawaga		
Misima Isolate		
Misima	Fellows (1894) Callister et al. (1983) Bartlett (n.d.)	n/a n/a n/a
Kakabai Group		
Dawawa	Knauber, Martin and Beate Knauber: field work since 1988	
Kakabai		
Kiriwina Group		
Budibud		
Kiriwina	Twomey (n.d.a) Senft, Gunter (1986) 4500 entries, 3000 index entries	n/a n/a
Muyuw	Lithgow, David and Daphne Lithgow (1974) 240 pp, 1900 entries, 2800 index entries	Simple
Nimowa Sud-Est Group		
Nimowa	Twomey (n.d.c)	n/a
Sud-Est	Twomey (n.d.b) Anderson, Mike (1990) 133 pp, 1650 entries, 1900 index entries	n/a Complex

From the above chart, one can notice that there has been little work in the East-Central group, the Anuki isolate, and the Taupota group. The Sewa Bay and Are groups have some

```
1890
            Pannieti
1900

1910

1920      Mekeo
          Dobu
1930      Motu
          Balawaia
1940      Roro

1950      Dobu
          Motu
1960      Molima
          Suau      Muyuw
1970                Balawaia, Are
                    Lala
1980            Magori
                      Iamalele
1990          Misima
                    Dobu, Tawala
                Bunama, Kiriwina
                      Iduna, Tubetube
              Gumawana
                    Gapapaiwa
              Sud-Est, Minaveha
```

work, but less than half of the languages have dictionaries available for them. More than half of the languages in the other language groups in the PTC region have a mix of simple and complex dictionaries now available, or else have work in progress.

Given the apparent abundance of lexical resources, one might be led to believe that there is little that has been unstudied. In fact, the opposite is the case. Most comparative linguists have not had the bulk of this information available to them. Viewing the data chronologically reveals the reason for this.

In the 25 years following the 1894 publication of the Pannieti dictionary, no other dictionary was completed.[6] The next twenty five years saw the completion of five dictionaries, and the third twenty five years saw the completion of five new dictionaries. The last twenty five years have seen over fourteen new dictionaries.[7] Thus only recent comparativists have had access to most of these dictionaries.

3.   Comparative Linguistics and Lexicography

Comparative linguistics often begins with the collection of word lists in order to determine genetic relationships. As more becomes known of the linguistic situation, linguists begin to delve into more detail in order to better understand the relationships between the languages.

[6] Blust (personal communication) indicates Bromilow's Dobu vocabulary of several thousand entries was published in 1904, so this statement (as well as others) probably needs revising. The general thrust of the statements, that most dictionary work is recent, is still valid.

[7] This excludes the seven undated dictionaries, most of which were probably compiled in the past 25 years.

At the beginning of the 1980s John Lynch wrote an article on the future of linguistic study in Melanesia. Comparing linguistic research in Melanesia with that of Oceania in general, Lynch notes that the Melanesian linguist is confronted not only with proportionally more languages, but with more difficult access to language data as well.

Looking ahead to the decade of the eighties, Lynch states:

"The first necessity for comparative linguistic research is descriptive linguistic research; we need more, and better, descriptions of the grammars, the phonologies, and the vocabularies of as many Melanesian languages as possible."

He goes on to say:

"A second important step, it seems to me, is the need for low-level reconstruction of a number of families... It is now time ... to begin comparative work from the bottom up: i.e., begin reconstructing the proto-languages of the various families, and then of the stocks which contain these families and so on until, by this application of the comparative method, the overall picture can be built up more scientifically than it has been so far. A natural spin-off of this undertaking will undoubtedly be that we will be able to make cultural-historical statements in much greater detail..."

Lynch's assessment of the situation was very appropriate. The 1980s did see steady growth in descriptive linguistic work. A massive comparative study based in part on this work was produced by Malcolm Ross in 1986 and published by Pacific Linguistics in 1988. Yet even as his comparative study was going to press, new descriptive data was shedding more light on the linguistic situation he described. He notes at the end of his chapter on the Papuan Tip Cluster (1988:212) that as the volume was going to press, new data from Cliff Olson working in the Gumawana (Gumasi) language was suggesting a reanalysis of the genetic relationship of Gumawana.[8] As Lynch foresaw, new and more detailed descriptive data is resulting in better comparative analysis.

Lynch looked beyond a mere description of the early linguistic situation. He foresaw that linguistic data would shed light on the peoples who spoke those languages. Rehg and Bendor (1990) use comparative lexicographic evidence in order to support a hypothesis concerning Micronesian prehistory. Their study looks not only at lexical forms, but at semantic domains. They discuss the introduction of loan-words in the languages and note that "failure to take into account lexical transfer...will result in erroneous reconstructions and may lead to spurious subgroupings hypotheses." They use evidence of lexical transfer to reconstruct the social interaction of peoples.

Achievements in lexicography in the recent past have resulted in new comparative work. Within the Austronesian realm, Robert Blust's four part series on Austronesian

---

[8] See also Ross (1992) for further information on the reanalysis of the genetic relationship of Gumawana.

etymologies has appeared. Blust has attempted to demonstrate not only the phonological form of the proto lexeme, but has attempted a brief definition as well.

The comparative etymological work of Blust is very interesting, but it leaves many unanswered question in my mind. His treatment of *kelas* illustrates this well.

> *kelas 'to peel, skin off'. BON *kelás* 'pull off, as cloths', BAL *kelas* 'lay bare, peel, skin', NGG *gola* 'scrape, plane', *golah-i* 'to scrape, plane; chafe, bruise, rub skin off', MOTA *gor* 'rasp, scrape'.
> NOTE: Also MLG *helaka* 'skinned, peeled'.

What sorts of things can be peeled? Is it only animal skin? Can a food such as banana be peeled as the English gloss might imply? Is an instrument used in the process? Is the action always intentional, or might it be accidental? Questions of this sort are important to expose the full meaning of the word. Some of Blust's sources could have no doubt shed light on these questions, yet many sources could not. Answers to such questions could lead to interesting discoveries of semantic innovations in language groupings.

Current comparative studies normally result in two products. One product is the reconstruction of the earlier stages of languages. The other product is insight into the degrees and directions of language change. Ross (1988) illustrates this in his discussion of phonological innovations that speech communities developed. He uses this information in association with other facts and assumptions to propose a history of early people movements. However, efforts in comparative lexicography seem to be limited only to etymological reconstruction, as exemplified in Ross (1988). While this is important, one cannot help but wonder what insights into the linguistic communities of Oceania could be obtained by comparing similarities and differences in the semantic features of lexemes.

Laycock (1986) notes that comparative work can be based on virtually any feature of language. He notes that most comparative work is based on retentions and innovations of some aspect of language. Lexical studies are usually based on shared forms of lexemes, not on a careful study of the semantic properties of the lexemes. Laycock suggests a research project in which 'semantic conflations' are compared. He has noted that some languages use one lexeme to refer to two apparantly separate items (such as one lexeme referring to head hair and tree leaves). Laycock began to investigate inter-language relationships, but unfortunately it appears that little came of this study before his death.

What could be learned from the study of the semantic features of lexemes? Might new roots (à la Blust 1988) have been introduced and old ones lost? Do collocational patterns, metaphors, and basic concepts shift from one language group to another? Perhaps such studies could point to underlying differences in the cognitive framework of the people speaking these languages and possibly supply new evidence about genetic relationships. If comparative lexicography could answer questions like these it could shed light not only on language, but on the very heart of cognition as well.

4.   Future Directions

The lexicons that are available to linguists at this time would allow for new comparative studies. It appears that studies similar to the work of Rehg and Bender (1990) could be undertaken. Undoubtedly significant work toward a Proto-PTC dictionary along the lines of Blust's Austronesian etymology work could be undertaken based on current lexicons. But if comparative work is to be based on semantic features, the existing lexicons and current methods used to compile lexicons are inadequate.

Comparative lexicography, if it involved details such as comparing the differences between the semantic features of lexemes or comparing collocational patterns of lexemes, would be difficult. A basic question which needs to be carefully considered is whether such study is feasible given the current resources. But to answer that question, one needs to consider what the cost might be to undertake such study.

Cooperative research could only be effective if a common model were used by the various parties involved in data collection. I would suggest a model based on the work of Anna Wierzbika.[9] Yet significant work would need to be undertaken to develop a practical model that lexicographers could consistently use to describe the lexemes to be studied.[10]

As lexicographers know all too well, lexical analysis and dictionary compilation in any language is an extremely time consuming task. Exhaustive study of major portions of multiple lexicons is probably out of the question. However it may be possible to study limited semantic domains. It is the opinion of this author that the easiest domains to study are those domains which are verbal.[11] Examples of such domains are: speech act verbs, verbs of motion, and verbs of perception. The most difficult domains to study are concrete objects such as foods, animals, plants, and the like.

I would suggest that the work begin with speech act lexemes, verbal lexemes related to eating and drinking, as well as lexemes referring to a small group of ocean life forms such as squids and octopuses. The reason for these three choices are: speech act lexemes are frequently studied and well developed models of analysis and definition exist (see Wierzbicka (1987)); lexemes of eating and drinking seem to be somewhat stable across Austronesian language boundries, thus providing a more homogenous area of study; and

[9] Note especially Wierzbicka (1985a, 1985b, 1989).
[10] I am doubtful whether a cooperative research project dependent upon consistent data collection by multiple field workers is even possible. The chances of success seem much more probable if the field work and research were undertaken by one entity.
[11] This assumes that the researcher has access to text material and that most of the comparative work is done at a distance from the places where the languages are spoken. The basis for this statement is the fact that verbal concepts are basically relational concepts and many of their semantic features can be discovered through the study of text material. Concrete nominal concepts embody large amounts of information about the world and culture. Most of this information can be known only through observation and elicitation. This type of information is rarely present in text material. For further discussion of this point, see Snyder (forthcoming).

lexemes referring to squids and octopuses allow for limited[12] study of non-verbal lexemes that would most likely be lexicalized throught the area.

Thus a comparative effort based on traditional lexical data would require the development of a model of data collection, a model for comparing the data, a standardized corpus of data, and perhaps a cooperative team effort.

An alternate, though perhaps not yet feasible, method of comparative study is study of collocational patterns in texts. Such study could potentially determine semantic sets through collocational patterns. For example, a verb such as "push" implies an agent which causes the action; or the object which is "pushed" may be limited to objects of a certain size or orientation. A study of such verbs would provide clues about what things people perceive of as being able to cause events, or about what objects people perceive of as having similar "size or orientation". The retentions and innovations among such sets of agents and objects might provide interesting insights into language change.

5.   Conclusion

This paper has discussed the status of lexical work among the languages of the Papuan Tip Cluster. It has also looked at the type of comparative work that has been done among Austronesian languages. Based on that information, I have suggested ways in which the current lexical data might provide useful as well as areas of future lexical and comparative research that might prove useful. Whether such studies will ever be undertaken depends in part upon the interests and resources of Austronesianists studying the languages of the Papuan Tip Cluster.[13]

References

Anderson, Mike. 1990. *Sud-Est Field Notes*. Dictionaries of Papua New Guinea, vol. 11. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Bartlett, H.K. n.d. A vocabulary of the language of Paneati and Misima - Papua. Noumea: South Pacific Commission microfilm.

Blust, Robert A. 1980. Austronesian etymologies. *Oceanic Linguistics* 19:1-181.

_____. 1983-4. Austronesian Etymologies - II. *Oceanic Linguistics* 22-23:29-150.

_____. 1986. Austronesian Etymologies - III. *Oceanic Linguistics* 25:1-231.

_____. 1988. *Austronesian root theory*. Amsterdam: John Benjamins.

_____. 1989. Austronesian Etymologies - IV. *Oceanic Linguistics* 28:111-180.

[12]Perhaps this area of non-verbal objects would prove too limited.
[13]The author welcomes discussion on the paper.

Callister, Sandra et al. 1983. *Baba ana talisi ana buki: Misima dictionary*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Chowning, Ann. 1958. Molima - English dictionary. ms.

Clunn, S. P. and J. A. Kolia. 1977. *A Lala grammar sketch and vocabulary*. Port Moresby: Institute of Papua New Guinea Studies.

Coluccia, P. 1939. Dictionaire Roro - Français. ms.

_____. 1941. Dictionaire Français - Roro. ms.

Cooper, Russ. 1969. Suau word list and concordance. ms.

Desnoes, G. n.d. (ca. 1925) Dictionarie de Mekeo. ms.

Dixon, John W. 1928. *English - Dobu dictionary*. Salamo: Methodist Mission Press.

Dutton, T. E. 1976. History of research in Austronesian languages: eastern part of south-eastern mainland Papua. In Wurm, 129-40.

_____. 1980. A Magori vocabulary. ms: Australian National University.

Dyen, Isidore. 1965. *A lexicostatistical classification of the Austronesian languages*. International Journal of American Linguistics Memoir 19. Baltimore: Waverly Press.

Ezard, Bryan. 1984. Tawala dictionary. ms: SIL, Ukarumpa, Papua New Guinea.

Fellows, S. B. 1894. Grammar of the Pannieti dialect, British New Guinea, together with a comprehensive vocabulary. *Annual Report on British New Guinea, 1892-1893*:78-92. Melbourne: Government Printer.

Giblin, Eric n.d. A grammar and dictionary of the Mukawa language, with Gapapaiwa equivalents. Noumea: South Pacific Commission microfilm.

Gibson, Stan. 1990. Bwaidoga Dictionary. ms: SIL, Ukarumpa, Papua New Guinea.

Grant, R. V. 1953. *A school dictionary in the Dobu language*. Rabaul: Methodist Mission Press.

Gunderson, Steve and Jerri Gunderson. 1987. Tubetube-English dictionary. ms: SIL, Ukarumpa, Papua New Guinea.

Jones, Alan A. n.d. Mekeo area word lists. ms: Australian National University.

Kolia, J.A. 1975. A Balawaia grammar sketch and vocabulary. In *Studies in languages of central and south-east Papua*, ed. T. E. Dutton, 107-226. Pacific Linguistics C-29. Canberra: Australian National University.

Koloa, M. and J. A. Collier. 1937. Balawaia-English dictionary.

Laycock, D. C. 1986. Papuan languages and the possibility of semantic classification. In *Papers in New Guinea Linguistics*, no. 24, 1-10. Pacific Linguistics A-70. Canberra: Australian National University.

Lister-Turner, R. and J. B. Clark. 1954. *A dictionary of the Motu language of Papua*. Second edition, edited by Percy Chatterton. Sydney: NSW Government Printer.

Lithgow, Daphne. 1984. *Dobu - English dictionary*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Lithgow, David. 1976. History of research in Austronesian languages: Milne Bay Province. In Wurm, 157-70.

_____. 1985. Bunama dictionary. ms: SIL, Ukarumpa, Papua New Guinea.

Lithgow, David and Daphne Lithgow. 1974. *Muyuw dictionary*. Dictionaries of Papua New Guinea, vol. 1. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Lovell, Larry. to appear. *Minavega field notes*. Dictionaries of Papua New Guinea, vol. 13. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Lucht, Ramona. 1987a. Iduna - English dictionary. ms: SIL, Ukarumpa, Papua New Guinea.

_____. 1987b. English - Iduna dictionary. ms: SIL, Ukarumpa, Papua New Guinea.

Lynch, John. 1981. *Readings in the comparative linguistics of Melanesia*. Studying Pacific Language Series, no. 4. Port Moresby: University of Papua New Guinea.

McGuckin, Ed and Catherine McGuckin. to appear. *Gapapaiwa field notes*. Dictionaries in Papua New Guinea, vol. 12. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Olson, Clif and Roxanne Olson. 1988. Gumawana dictionary. ms: SIL, Ukarumpa, Papua New Guinea.

Rehg, Kenneth L. and Bryon W. Bendor. 1990. Lexical transfer from Marshallese to Mokilese: A case of intra-Micronesian borrowing. *Oceanic Linguistics* 29:1-26.

Ross, M. D. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Pacific Linguistics C-98. Canberra: Australian National University.

_____. 1992. The position of Gumawana among the languages of the Papuan Tip Cluster. *Language and Linguistics in Melanesia* 23:139-65.

Semi, Nafitalai, Imanueli Lobati, Kaliofasa Mokoi, Ilamwatu Iakobo, John Beaumont, and Margaret Beaumont. 1988. *Bonada ma 'adi saivila (Iamalele–English dictionary)*. Summer Institute of Linguistics.

Senft, Gunter. 1986. *Kilivila, the language of the Trobriand Islanders*. Berlin: Mouten de Gruyter.

Snyder, David Michael. forthcoming. Towards a defining vocabulary for Tok Pisin. M.A. thesis, University of Papua New Guinea.

Taylor, A. J. 1976. History of research in Austronesian languages: Western part of South-Eastern Mainland Papua. In Wurm, 141-55.

Twomey, Kevin. n.d.a. Kiriwinian vocabulary and grammar. ms. Losuia: Catholic Mission.

_____. n.d.b. Nimowan dictionary. ms.

_____. n.d.c. Vanatinai dictionary. ms.

Wierzbicka, Anna. 1985a. *Lexicography and conceptual analysis*. Ann Arbor, Mich.: Karoma Publishers.

_____. 1985b. A semantic metalanguage for a cross-cultural comparison of speech acts and speech genres. *Language in Society* 14:491-513.

_____. 1987. *English speech act verbs*. New York: Academic Press.

_____. 1989. Semantics primitives — an expanding set. *Quaderni di Semantica* 10:101-21.

Wurm, S. A., ed. 1976. *Austronesian languages*. Vol. 2 of *New Guinea area languages and language study*. Pacific Linguistics C-39. Canberra: Australian National University.