

PARTS OF SPEECH IN SOUTHEAST ASIAN LANGUAGES: AN AUTOLEXICAL VIEW

Eric Schiller
Lecturer in Linguistics
University of Chicago
1010 East 59th Street, Chicago IL 60637 USA

O. Introduction

The syntax of Southeast Asian languages often seems quite difficult when observed from a perspective based on the study of European languages. This complexity is often compounded when one applies a theoretical perspective which forces lexical items into fixed syntactic categories determined by what are claimed to be universal considerations. This paper uses the notion of syntactic polysemy (Schiller 1989) or syntactic flexibility (Ratliff to appear) to discuss the nature of word classes in Khmer and a few other Southeast Asian languages. Specifically, I will concentrate on several words which appear in a wide variety of syntactic contexts, not merely nouns and verbs, but also modals, adverbs, prepositions, and classifiers.

1. "Parts of Speech"

By using the Autolexical technique of separating syntactic considerations from semantic considerations (Sadock 1991), and having a distinct inventory of word classes (or categories) at each level, the often confusing problem of determining "parts-of-speech" is made much clearer. Categories which have traditionally been at least somewhat controversial, such as "relator-nouns", "classifiers", and "coverbs", are easier to deal with when syntactic, semantic, and morphological considerations are dealt with separately. These notions have a tendency to be defined in purely language-specific terms, usually by positional factors since morphology is not much help in mainland Southeast Asian languages. For pedagogical purposes it is often useful to determine lexical categories simply on the basis of co-occurrence restrictions. However, this approach runs into real problems in the languages which permit widespread deletion, as is the case with most of the isolating languages of Southeast Asia.

Consider the Khmer word presented in (1).

1) ក្នុង <knɔŋ> /knɔŋ/ 'inside'

Headley (1977) glosses it as a "predicative" with the meanings 'in, inside, within, during' and gives the following examples (inter alia):

a. ក្នុងឆ្នាំនេះ
/knɔŋ ɕhnam nih/
in year this
'during this year'

b. ក្នុងស្រុក
/knɔŋ srok/
in country
'in the country'

c. ចូលក្នុង
/coul knɔŋ/
enter inside
'Go inside!'

d. ដាក់សៀវភៅក្នុងប្រអប់
/dak siəwphəu knɔŋ prɑʔap/
put book inside box
'to put books in a box'

Just looking at these few examples, we can observe the diversity of the uses of /knɔŋ/. The semantic range is not very broad, with all senses having to do with a notion of being located inside of something, but the syntax is less clear. For the most part, /knɔŋ/ seems to be a preposition but in (1.c) either a nominal or adverbial analysis seems more appropriate. One might want to claim that (1.c) is parallel to the English translation, where one can claim that 'inside' is a preposition with a deleted object. Yet just analyzing /knɔŋ/ as a preposition (or, in Jacob's (1968) terms, a pre-nominal particle) runs into trouble, because it is most commonly found following the lexeme <nəu> in the following structure (1.e):

e. នៅក្នុងវាំង
/nəu knɔŋ vɔət/
be-in inside temple
'in the temple'

By examining this phrase alone we cannot discover the syntactic structure. One can easily imagine the trouble that can arise if we start defining our syntactic structures in terms of the presence or absence of /knɔŋ/. Even if we expand our example into a fuller sentence (1.f), it does not clarify matters:

f. គ្រូនៅក្នុងវត្ត

/kruu nəu knoŋ vɔət/
 teacher be-in in temple
 'The teacher is in the temple'

It appears that /nəu/ is the main verb, and that /knoŋ/ heads a prepositional phrase. Still, the addition of another word renders this judgement less clear (1.g).

g. គ្រូលេងនៅក្នុងវត្ត

/kruu deek nəu knoŋ vɔət/
 teacher sleep be-in in temple
 'The teacher sleeps in the temple'

Here it is clear that /deek/ is the main verb, and /nəu/ therefore must be either a preposition or a serial verb. The serial verb analysis is dubious in view of the fact that the constituent headed by /nəu/ can be fronted (1.h), which is not typical of serial verb constructions in Khmer.

h. នៅក្នុងវត្តគ្រូលេង

/nəu knoŋ vɔət kruu deek/
 in inside temple teacher sleep
 'In the temple, the teacher sleeps'

Judith Jacob (1968) treated /nəu-knoŋ/ as a compound preposition in such cases. (Note that since written Khmer does not separate words, we can not use the written language as a diagnostic.) I present these facts as an introduction to the problem at hand, namely the identification of syntactic categories or parts of speech. I will not go into any further analysis here, but note that /nəu/ was the subject of (Schiller 1984).

A particularly interesting, and typically Southeast Asian type of word is the classifier. Although classifiers are not as robust a category in Khmer as they are in languages such as Thai, there are still many cases where they are obligatory. For counting ordinary people, the word used is /nĕəʔ/, which follows the numeral as shown in (2.a):

- 2) អ្នក <ʔnək> /nĕəʔ/ 'person'
 នាក់ <nək> /nĕəʔ/ classifier for common people¹

a. មនុស្ស ពីរនាក់

/mənuh pĭi nĕəʔ/

person 2 CL

'two persons'

But the word is also used as a pronoun (usually second person²) as shown in (2.b) and is also the head noun in compounds (2.c,d):

- b. អ្នកមកលេងជាមួយខ្ញុំ ថ្ងៃនេះ?
 /nĕəʔ mɔɔk leən cĕə-muəy kʰnɔm tʰəi nĭhʔ/
 person come play with me day this
 'Are you coming to play with me today?'

- c. អ្នកណាស់នៅផ្ទះនោះ?
 /nĕəʔ naa rɔəh nəu ptĕəh nuhʔ/
 person wh live be-in house that
 'Who lives in that house?'

- d. អ្នកគ្រូរស់នៅផ្ទះនោះ
 /nĕəʔ kruu rɔəh nəu ptĕəh nuh./
 person teacher live be-in house that
 'The teacher lives in that house.'

One might suggest that in each case the classifier can be analyzed as a noun, and that classifiers are merely a subcategory of nouns. There is a major flaw in this treatment. In Khmer, classifiers generally do have the form of nouns, but they do not show the syntactic behaviour of nouns, in that they do not combine with adjectives and cannot be full NP's or even N. When used as a pronoun, the word does not take modifiers but must act as a complete N, as is usually the case with pronouns. Thus (2.d) cannot be interpreted as meaning 'You, teacher, live in that house' or 'Your teacher lives in that house.', though the last reading might be obtained by switching the order of /nĕəʔ/

¹ This spelling is used for the classifier only. Headley (1977) points out the similarity to Malay *anak* 'child', which better fits the <ʔnak> spelling.

² As pointed out in Schiller (1988), the pronominal system of Khmer is much messier than the literature indicates.

and /kruu/.³

One of the most widely discussed "part-of-speech" questions is that of 'coverbs'. The term 'coverb' has been applied to lexical items which are used as a verb, preposition and complementizer.

(3) អោយ <aoy> វាoy/ (also written as ឲ្យ) all meanings involve some notion of 'give'

- a. V: វ/___ វ (as a ditransitive verb)

ពាត់អោយលុយខ្ញុំ

/kɔət ʔaoy luy kʰom/

prn. give money me

'He gives me money.'

- b. P: វ/___ វ (as a preposition marking the benefactive)

ពាត់ធ្វើអោយខ្ញុំ

/kɔət tʰvəə ʔaoy kʰom/

prn. make give me

'He did it for me'

- c. C: វ/___ S (as a complementizer)

ពាត់ធ្វើអោយស្រួល

/kɔət tʰvəə ʔaoy sruəl/

prn. make give easy

'He made it easy'

- d. V: វ/___ S (as a causative verb which takes sentential complements)

វគ្គអោយខ្ញុំរៀនវិទ្យា

/kee ʔaoy kʰom rien khmae/

prn. give me study Khmer

'They made me study Khmer' or (!) 'They let me study Khmer'

- e. V: វ/___ S (as a causative verb which takes sentential complements)

អោយវគ្គស្រួលខ្ញុំទៅហើយ

/ʔaoy tæ sruəl kʰom tæu haey/

give only easy me go already

'Provided that it is easy, I'll go'

There are many fascinating semantic interactions in the data listed above, but it is not difficult to draw connections between the various meanings. The

³ The hedge is due to the fact that /nəəʔ/ is highly restricted as a pronoun in many dialects,

relationship between the causative /t^hvəə ʔaoy/ where /ʔaoy/ seems to be a complementizer, and the simple causative /ʔaoy/ in (3.d), with its ambiguous meaning might be viewed as simply deletion of /t^hvəə/ but this would only be possible with /ʔaoy/ interpreted as a verb, as Khmer syntax seems to demand a verb in every well formed declarative sentence.

But to look at these words as a mere case of mild polysemy is to miss an important point. Consider the word /trəw/, which has the meanings 'hit, come into contact with, experience, must, should, correct, right', and appears in an even wider range of configurations. It is also sometimes claimed to be a marker of "passive", although this analysis has been properly criticized by for the parallel Thai case by (Lekawatana 1975), *inter alia*.

- (4) a. V:S/Ṽ___ (as an adjective)

ចម្លើយនោះមិនត្រូវទេ

cəmləy nuh mun trəw tee
answer that not correct Prt.⁴

'That answer is not correct.'

- b. V:Ṽ/___Ṽ (as a transitive verb)

ព្រួចច្រឡំលើមេក្រវិញដើម

puuthau cɹɔluəh mɔɔk trəw cəəp
ax slip⁵ come hit leg

'The ax slipped and hit his leg.'

- c. V:Ṽ/___V (as an auxiliary verb)

គាត់ត្រូវត្រូវការ

kɔət trəw puukae

he should be-skillful

'He ought to be skillful.'

- d. V:Ṽ/___Ṽ (as an auxiliary verb)

ខ្ញុំត្រូវទៅផ្សារណាមួយ

khñom trəw təu p^hsaa t^hɲai nih
I must go market day this

'I must go to the market today.'

⁴ See Eilfort & Schiller (1990) for a discussion of this particle.

⁵ Lit. 'miss (passing by or beyond)' There may be an understood deleted object here.

- e. V:V̄/___V̄ (as a verb which takes sentential complements)*

គាត់ត្រូវ ឆ្កែ ខាំ

kōāt trəw c^hkae kham

he experience dog bite

'He got bit by a dog.'

- f. V:V̄/___N̄ (as a transitive verb, syntactically)

គាត់ត្រូវ គ្រាប់

kōāt trəw krōap

prn. experience/hit bullet

'He was struck by a bullet.'

- g. N:V̄/V___ (as a noun)

មនុស្ស នោះ ដឹង ខុសត្រូវ

mənuh nuh dəŋ khoh trəw

man that know wrong right

'That man knows right from wrong.'

- h. N:F/P___ (as a noun)

មនុស្ស នោះ ដឹង ខុស ពី ត្រូវ

mənuh nuh dəŋ khoh pli trəw

man that know wrong from right

'That man knows right from wrong.'

- i. A:V̄/V___ (as an adverb)

គាត់ ឆ្លើយ លំនួច ត្រូវ

kōāt c^hlaey sɔmnuə trəw

he answer question correctly


'He answers the question correctly.'

The traditional approach to the handling of these facts is to create multiple listings in the lexicon. If the meanings are close enough, dictionary makers may include them in a single listing, leaving those meanings which seem to differ significantly in supplementary listings. In the present case much would depend on the lexicographer's analysis of the semantics.

The Autolexical approach involves complete autonomy between modules (called Dimensions) of a grammar, among them syntax, semantics, and morphology. Regular correspondences between categories of these different dimensions are noted in the form of default relations. Simple examples are that syntactic noun phrases usually correspond to bound variables (or referential entities) in the formal semantics, transitive predicates tend to be

verbs, and so on. Since there is nothing that corresponds to the Projection Principle (Chomsky 1981), it is not the case that the semantic category follows from the syntactic category or vice versa. In English one finds many cases where mismatches occur, for example the adverbial function of 'today' is fulfilled by a morphological noun, and the verb 'seem' corresponds to something closer to a logical operator than a standard predicate.

Under this approach, since we see the word /trəw/ in every syntactic environment except that of a preposition, we can then provide the following lexical entry.

- (5)  <troov> /trəw/
 Syntax: [-P]
 Semantics: *default*

This listing indicates that the word can be a member of any syntactic category except that of preposition (or particle), and that the formal logical semantics⁷ will be determined by the default relations I mentioned above. So that if the word is in a syntactic position of a noun, it will be treated semantically as a noun. If it is in the position of a main verb, it will have the semantics of a predicate, and so on. This is not to say that one cannot select one of the syntactic/semantic meanings as central. (Croft 1991) argues that there are strong cross-linguistic correlations between semantic properties and syntactic categories, so it is natural that the uses of a word which are closest to the predicted prototype would be the most common ones.

A few additional observations should be made. First of all, the use of this item in a wide variety of syntactic positions is quite similar to that seen in Thai (a member of the Tai-Kadai family) and Hmong (a member of the Hmong-Mien family). The forms used in those languages may even be etymologically related (Gérard Diffloth and Martha Ratliff, p.c.), though the languages are either unrelated to Khmer (Benedict 1975) or very distantly related (as I argued in Schiller (1987).)

One fact about the use in each language is that it cannot be employed as a preposition or as a pronoun. Prepositions and pronouns tend to form closed classes in most languages, and it does not seem unreasonable to take the position that lexical items will have default specifications of [- pronominal] and


⁶ See (Lekawatana, 1975) for discussion of comparable Thai construction.

⁷ The term logical semantics refers only to the truth-conditional combinatoric semantics. Naturally the "real-world" meaning of the item also needs to be

[- prepositional]. A coverb is therefore a case of a verb acquiring the feature [+ prepositional], while cases of nouns becoming pronominal involve the acquisition of the [+ pronominal] feature. A case of the latter is Khmer /kʰnɔm/, which was once a noun meaning 'slave' (a meaning preserved in the verbal use of the word as 'to serve') and which is now a first person pronoun. An additional example might be /nĕəʔ/, discussed earlier.

If we adopt the position that lexical items are [- prepositional, - pronominal] as a default, we expect then that /trəw/ will function in all other syntactic positions, and the data illustrates that this does, in fact, seem to be the case. In addition, there seems to be a pattern that in a modifier position (roughly - to the right of the constituent with which it combines), /trəw/ has a semantic core of correctness or appropriateness, but in other positions (to the left of the constituent with which it combines) it seems to have a semantic core of contact, or experience. Of course the modal use counterexemplifies this, but then auxiliaries often have semantics rather similar to adverbs.

It is interesting that native speakers, when asked how many different "/trəw/s" they have, give a variety of replies. A few insist that there is only one, often the reply is two (one for the "contact" meaning and one for the "correct/obligation" meaning, and sometimes a greater number is given.) But the answer is almost always phrased in semantic terms, almost never in syntactic terms, even by linguistically sophisticated informants and teachers. In any event, leaving aside the difficulty of finding a way to express, in English, the central meaning of /trəw/, we come up with the entry in (6), taking [-Prepositional] as a default:

- (6)  <troov>/trəw/
 Syntax: *unspecified*
 Semantics: *default*

This discussion of the perception of the number of different /trəw/'s is hardly scientific, but there is one interesting piece of evidence showing that Khmers, at least literate ones, do find some unifying theme despite the variety of uses. This evidence comes from the writing system. There is a considerable amount of morphemic reduplication in Khmer, used to express plurality among other things. In such cases a special reduplicative marker is used instead of writing out the word twice, e.g.

specified, for each of the uses.

- (8) <dūk dūk> /tuuk-tuuk/ 'boats' =

When we find two different uses of the word /nəu/ side-by-side, then the same device is used (9).

- (9) ကုန်နေတဲမိသီး
 /koun nəu nəu p^htēah/
 child still be-in house
 ‘The child is still inside.’

While I would not want to make too much out of this orthographic convention, it does indicate that the semantic range is not so great as to make the use of such a device uncomfortable.

There is a piece of evidence which at first seems to counterexemplify the proposed analysis. We do not find /kit/ 'think' used as a noun, although there is no obvious reason why this should not be so. The answer lies in the existence in the lexicon of /kɔmnɪt/ 'thought', a form created during the period when Khmer enjoyed productive derivation. This is an example of a general principle of primacy of the lexicon, whereby one does not create a form if an appropriate form already exists. This explanation applies to most languages, for example English, which lacks a form **fastly*. We can state it as a principle (10), though it should be noted that this is hardly anything new:

- (10) **Grounding Principle:** *If an appropriate lexeme exists in the lexicon, do not use the underspecified form instead.*

3. Expressives

In the remainder of the paper, I would like to make an attempt to show that the Autolexical approach can even handle one of the trickiest problems of Southeast Asian languages. I am referring to the thorny question of expressives (Diffloth 1972, 1976). Semai, a Mon-Khmer languages of Malaysia, has a productive system of expressives which do not fit into the normal syntactic categories of verb, noun or particle/preposition. They are easily identified by their rich semantics and their morphology, which is quite distinct, involving reduplication of the major syllable, e.g. (11):

- (11) /dyʃ:lyʃ:l/
 'the appearance of an object floating down a river and getting stuck'

In addition, only in expressives will one find a bi-consonantal prefix which contains a copy of the first and last consonants of the root (12):

- (12) /dldyʃ:l/ 'appearance of an object which goes on floating down'
 /dhɒŋɔh/ 'appearance of nodding constantly'

There are many other morpho-phonological properties discussed in Diffloth's work. But suffice it to say that there is a clear morphological category involved.

Now the syntax is more of a problem. (Diffloth 1976) argues persuasively that expressives are not nouns, verbs, adjectives or adverbs. They do not occupy syntactic positions associated with these categories. Instead, they are found preceding sentences or isolated noun phrases. An example of a full sentence is (13).

- (13) /rladī:wdīw hi-nɛ:ŋ grta?/
 EXPRESSIVE we-see bridge
 (used to describe the look of a bridge)

N.B. /hi-/ is a pronominal agreement marker which can co-occur with a full lexical pronoun.

One might then posit a unique syntactic category for expressives, but there is another alternative, which is that expressives play no role at all in the formal syntax of the language. The expressive can appear in any position which is not inside a word boundary, despite the fact that Semai is, as far as I know, a fixed word-order language.

The semantics of expressives are very complex. (Diffloth 1976) notes that "The meanings of expressives seem to be extremely detailed and idiosyncratic, describing a situation perceived as a whole, as an independent clause would. On the other hand, the same expressive can be used to describe a variety of situations which at first glance seem to be quite different but share a common core which could be defined as a cluster of elementary sensations." In other words, there seems to be a direct mapping between the perceptual world and the choice of expressive. Or, in some languages (including Semai), the mapping may actually be between perceptions and phonological material,

since expressive meaning can be varied in subtle ways by making small changes to the phonetics, e.g. (14).

- (14) a. /prbuñbuñbuñ/ 'noise of bubbles in water'
 b. /prbucbucbuc/ 'noise of small bubbles in mud'
 c. /prbusbusbus/ 'noise of big bubbles in mud'

This mapping seems to bypass ordinary lexical semantics, and as (Diffloth 1976) noted, the the semantics of the expressive can only be analyzed by decomposing the word into its phonetics, perhaps even its distinctive features. Moreover, the expressive is non-propositional, and thus plays no role in the logico-semantics.

But there is a discourse function involved. It creates a detailed an vivid picture in the mind of the hearer, and my reading of Diffloth's articles and P.c.'s indicate that the expressive does in fact contains critical background (scene-setting) information for the discourse but this is also new information. In Semai, expressives can be considered to occupy the head⁸, as opposed to tail, position in the discourse-functional module. Access to a lot more data, especially complete discourses, is required before drawing any further conclusions, however.

We can provide the following Autolexical entry for an expressive (15), keeping in mind that the gloss is only one possible use (thus the e.g.):

- (15) /pŋprlɔːŋ/ e.g., 'appearance of a completely bald head, big and smooth like a papaya'

Morphology: E

Syntax: nil

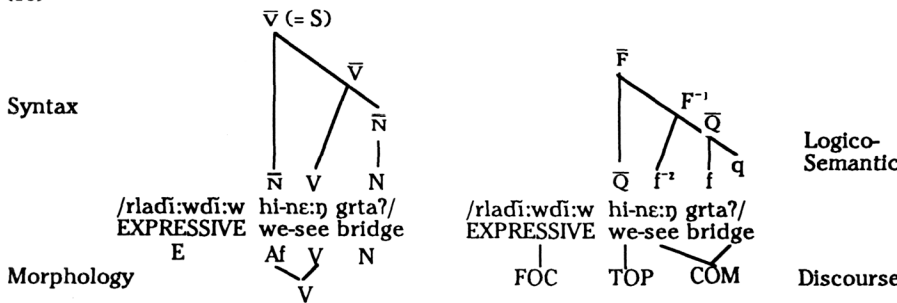
Logico-Semantics: nil

Discourse Function: [HEAD]

A possible multi-dimensional representation of the expressive sentence in (13) is presented in (16), keeping in mind that the association of the expressive with focus is a particular property of the word order of the sentence. Were the expressive to occupy a different position it would not be the focus. Thus the lexical entry of the expressive does not include information about focus, since expressives per se are not obligatorily focussed.

⁸The terms 'head' and 'tail' were introduced in Sadock's 1991 Autolexical Seminar at the University of Chicago.

(16)



(17) Associations:

Morphology:	E	[Af	V ⁰	N ⁰
Syntax:	∅	N̄	[V	N]
Logico-Semantics:	∅	Q̄	[f ²	f]
Discourse:	[_{head} FOC [_{tail} TOP COMMENT]]			

The associations are all in conformity with the default relationships between dimensions discussed by Eilfort & Schiller (1990), and the bracketing mismatch between morphology and syntax is the standard one for pronominal clitics. Much more could, and should be said on this topic, but this must be left to investigators who have better access to and command of languages with expressives such as these. The point is not to defend this particular analysis, but rather to show how the model of radical autonomy of components can be useful in descriptive work. The Autolexical model imposes a tremendous burden - that of explicitness at every level of description. There are no rugs under which dirty data can be swept.

A much more elaborate analysis of the discourse facts in particular are needed here. But when compared with the analyses available if frameworks which require everything to have a syntactic and semantic function, I think that this Autolexical approach produces descriptions of so-called "exotic" language facts which are far more plausible and useful, especially from a pedagogical standpoint.

These are just a few of the applications of the Autolexical technique to problems in the analysis of Southeast Asian languages. By relaxing the connection between syntactic and semantic category, it is possible to better understand the existence of "coverbs" which have not (yet) been bleached of their meaning

and the related question of verb serialization (Schiller 1991), **expressives** (perhaps?) and the constituent structure of sentences without recourse to complex mechanisms of syntactic theory, especially those motivated by a **Projection Principle** or obligatory match between syntactic and semantic category (GPSG, **Categorial Grammar**). The **Autolexical** approach provides useful tools for the description of Southeast Asian syntax which can be applied in pedagogical as well as descriptive and theoretical work.

References

- Benedict, P. K. (1975) **Austro-Thai: language and culture, with a glossary of roots**. HRAF Press.
- Chomsky, N. (1981) **Lectures on Government and Binding**. Foris.
- Croft, W. (1991) **Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information**. University of Chicago Press.
- Diffloth, G. (1972) 'Notes on Expressive Meaning.' In E. Schiller (ed), (1988) **The Best of CLS**. Chicago Linguistic Society.
- Diffloth, G. (1976) 'Expressives in Semai.' In L.C.Thompson et al. (eds), (1976) **Austroasiatic Studies**, University of Hawaii Press, 249-164.
- Eilfort, W. and E.Schiller. (1990) 'Pragmatics and Grammar: Cross-Modular Relations in Autolexical Theory.' **Proc 26th CLS**. Chicago Vol 1.
- Headley, R. K. (1977) **Cambodian - English Dictionary**. Catholic University of America Press.
- Jacob, J. M. (1986). **Introduction to Cambodian**. Oxford University Press.
- Lekawatana, P.(1975) 'The So-Called Passive in Thai.' In W. Gething (ed) (1975) **A Tai Festschrift for William J. Gedney**. University of Hawaii.
- Sadock, J. M. (1991) **Autolexical Syntax: A Theory of Parallel Grammatical Representations**. University of Chicago Press.
- Schiller, E. (1984) **Syntactic Polysemy and the Lexicon in Khmer**. M.A. Thesis, University of Chicago..
- Schiller, E. 'Syntactic Polysemy and Underspecification in the Lexicon.' **Proc 15th BLS** 15. Berkeley 278-290.
- Schiller, E. (1990) 'The Typology of Serial Verb Constructions.' **Proc 26th CLS**. Chicago Vol 1.
- Schiller, E. (1991) **An Autolexical Account of Subordinating Serial Verb Constructions**. Ph.D. thesis, University of Chicago, 1991.
- Schiller, E. (1991) 'An Autolexical Account of Oceanic Verbal Prepositions.' 6th Int'l Conference on Austronesian Linguistics, Honolulu. 1991.