Autolexical solutions to the problem of "parts of speech" in Southeast Asian languages

Eric Schiller University of Chicago

0. Introduction

The syntax of Southeast Asian languages can appear quite difficult when observed from an ethnocentric perspective, and this complexity is often compounded when one applies a theoretical perspective which forces lexical items into fixed syntactic categories determined by putative universal considerations. This paper uses the notion of syntactic polysemy or syntactic flexibility (Ratliff 1991) 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. Clarifying what we mean by "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 easier. Let's start with categories which have traditionally been at least somewhat controversial: "relator-nouns", "classifiers", and "coverbs". These terms 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. As we shall see, however, this approach leads to analytical problems in the languages which allow widespread deletion, as is the case with most of the isolating languages of Southeast Asia.

Consider the Khmer word presented in (1).

1) ក្នុង<knoŋ> /knoŋ/ 'inside'

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

- a. ក្នុងឆ្នាំនេះ /knon chnam nih/ in year this 'during this year'
- b. గ్లైస్ట్రోగ్ /knon srok/ 'in the country'
- c. ចូលក្នុង /coul knon/ 'Go inside!'
- d. ដាក់សៀវភៅក្នុងប្រអប់
 / dak siawphau knon pro?op/
 'to put books in a box'

Even these few examples serve to show the diversity of the uses of /knon/. 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, it 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 analyzing knon as a preposition (or, as Jacob 1968 does, a pre-nominal particle) requires justification, because it is most commonly found

following the lexeme $n \ni u$ in the following structure (1.e):

e. នៅក្នុងវាត់

/ nau knon voat/ be-in in temple 'in the temple'

By examining this phrase alone we cannot discover the syntactic structure. And imagine the trouble we will get into if we start defining our syntactic structures in terms of the presence or absence of knon! A fuller sentence (1.f) does not clarify matters:

f. គ្រូនៅក្នុងវ៉ាត់

/kruu nau knon voat/ teacher be-in in temple 'The teacher is in the temple'

Although it appears that $n \ni u$ is the main verb, and that $k n \circ n$ heads a prepositional phrase, the addition of another word muddies the waters (1.g):

g. គ្រូដេកនៅក្នុងវ៉ាត់

/kruu deek nau knon voat/ teacher sleep be-in in temple 'The teacher sleeps in the temple'

And as further support we can front the constituent headed by / neu/(1.h):

h. នៅក្នុងវ៉ាត់គ្រូដេក

/ nau knon voat kruu deec/ in inside temple teacher sleep 'In the temple, the teacher sleeps'

For this reason, (Jacob 1968) treated $n = u - k n o \eta$ as a compound preposition in such cases. (Note, by the way, that written Khmer does not separate words.) 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 n = u was the subject of Schiller 1984, and that the analysis of $k n o \eta$ as a preposition is justified in Schiller 1991c.

Let us now turn to a case involving a classifier. Although classifiers are not as robust a category in Khmer as they are in, say, Thai, there are still many cases where they are obligatory. For counting ordinary people, the word used is near, which follows the numeral as shown in (2):

នាក់ <năk> /neّa?/ classifier for common people¹

a. មនុស្សពីរនាក់

/manuh pii nea?/

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):

ь. អ្នក់មកលេងជាមួយខ្ញុំថ្ងៃនេះ?

/ nea? mook leen cia-muay khnom thnai nih/ person come play with me day this 'Are you coming to play with me today?'

c. អ្នក់ណារាស់នៅផ្ទះនោះ?

/ nea? naa roah nau phteah nuh/ person wh live be-in house that 'Who lives in that house?'

d. អ្នក់គ្រូរាស់នៅផ្ទះនោះ

/ nea? kruu roah nau phteah nuh/ person teacher live be-in house that 'The teacher lives in that house.'

Are we then to suggest that in each case we have a noun, and that classifiers are just nouns? 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 noun phrases or even N's. When used as a pronoun, the word does not take modifiers but must act as a complete noun phrase, 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 near and kruu.³

We now examine a standard coverb case, where we find a lexical item used as a verb, benfactive marker and complementizer.

(3) អោយ <aoy> /?aoy/ also written as គ្នា. All meanings involve some notion of giving or transfer:

a. V: \overline{V} /___ $\overline{\overline{N}}$ $\overline{\overline{N}}$ (as a ditransitive verb) គាត់ អោយលុយខ្ញុំ

/kŏat ʔaoy luy kʰñom/

prn. give money me 'He gives me money.'

b. $V: \overline{V}/\underline{\hspace{1cm}} \overline{\overline{N}}$ (as a verb marking the benefactive⁴)

គាត់ធ្វើ អោយខ្ញុំ

/koat thvəə ʔaoy kʰñom/ pm. make give me

'He did it for me'

c. C: $\overline{S}/\underline{\hspace{1cm}}$ S (as a complementizer)

គាត់ធ្វើ អោយស្រួល

/koat thvaa ?aoy srual/ prn. make give easy 'He made it easy'

d. $V: \overline{V}/__S$ (as a causative verb which takes

sentential complements) គែ អោយខ្ញុំរៀនខ្មែរ

/kee ?aoy kʰñom rien kʰmaɛ/

prn. give I study Khmer 'They made me study Khmer' or(!) 'They let me study Khmer'

e. V: V/___S (as a causative verb which takes sentential complements)

អោយតែស្រួលខ្ញុំទៅហើយ

/?aoy tas sruəl k^hñom təu haey/ give only easy I go already

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

There are many fascinating semantic interactions in the data listed above, and it is not difficult to draw connections between the various meanings. The relationship between the causative thvaa raoy where raoy seems to be a complementizer, and the simple causative raoy in (3.d), with its ambiguous meaning might be viewed as simply deletion of thvaa, but this would only be possible with raoy 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 trow, 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 with regard to the parallel situation in Thai (Lekawatana 1975 inter alia.)

(4) a. V: S/N ___ (as an adjective) ច្ឆាំយនោះមិនត្រូវទេ

comlasy nuh mun traw tee answer that not correct Prt. 5
'That answer is not correct.'

b. $V: \overline{V}/$ ___ $\overline{\overline{N}}$ (as a transitive verb) ពូថៅជ្រួលួសមកត្រូវជើង

puuthau croluah mook traw caan ax $slip^6$ come hit leg 'The ax slipped and hit his leg.'

c. V: V/___V (as an auxiliary verb) គាត់ត្រូវពូកែ

> koat traw puukas he should be-skillful 'He ought to be skillful.'

- d. V: V/__V (as an auxiliary verb)
 ខ្ញុំត្រូវទៅផ្សារថ្ងៃនេះ
 k nom traw tau phaaa thhai nih
 I must go market day this
 'I must go to the market today.'
- e. $V: \overline{V}/_{\overline{V}}$ (as a verb which takes sentential complements)
 គាត់ត្រូវ ឆ្កែខាំ

 koat traw chkas kham he experience dog bite 'He got bit by a dog.'
- f. V: V/__N (as a transitive verb, syntactically)
 គាត់ត្រូវគ្រប់
 koat traw kroap
 prn. experience/hit bullet
 'He was struck by a bullet.'

g. N: V/V ___ (as a noun) មនុស្ស នោះដឹងខុសត្រូវៈ

man that know wrong right 'That man knows right from wrong.'

h. N: P/P ___(as a noun) មនុស្ស នោះដិងខុសពីត្រូវៈ

> menuh nuh den khoh pii trew man that know wrong from right 'That man knows right from wrong.'

i. A: \overline{V}/V ___ (as an adverb) គាត់ឆ្លើយសំនួត្រូវ

> koat chlaay somnua traw he answer 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 depends on the lexicographer's analysis of the semantics.

The Autolexical approach involves complete autonomy between modules 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. It is often the case that syntactic noun phrases correspond to bound variables 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 we have many cases where mismatches occur, for example the adverbial function of 'today' is filled by a

morphological noun, and the verb 'seem' corresponds to something closer to a logical operator than a standard predicate. There is insufficient room in this brief paper for further detail, but see Sadock 1991 for an overview of the theory.

Under this approach, since we see the word $tr \ni w$ in every syntactic environment except that of a preposition, we can then provide the following lexical entry.

(5) ត្រ្ជីវ<troov> /traw/

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 semantics 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 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 of this word 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 have the default specifications [-pronominal] and [-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 khñom, 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 near, discussed earlier.

So if we adopt the position that lexical items are neither prepositional nor pronominal as a default, we expect then that trew 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), trew 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 informants, when asked how many different $tr \ni w$'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 $tr \ni w$, we come up with the entry in (6), taking [-prepositional] as a default:

(6) ត្រី្ជ like troov > / traw/

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.

When we find two different uses of the word $n \neq u$ sideby-side, then the same device is used (9).

(9) កូននៅ២ផ្ទះ /koun nau nau p^hteah/ 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.

Now consider a fact which at first seems to counterexemplify the proposed analysis. We do not find $k \neq t$ '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 $k \supset mn \neq t$ '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 only recapitulates notions which are well established in linguistic theory:

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

In the rest 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, 1976,1979).

Semai, a Mon-Khmer language 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.

(11) / dyɔ̃:lyɔ̃:l/ 'the appearance of an object floating down a river and getting stuck'

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

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).

/rladî:wdîw hi-nε: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 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 contain 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) /pnpriding/ e.g., 'appearance of a completely

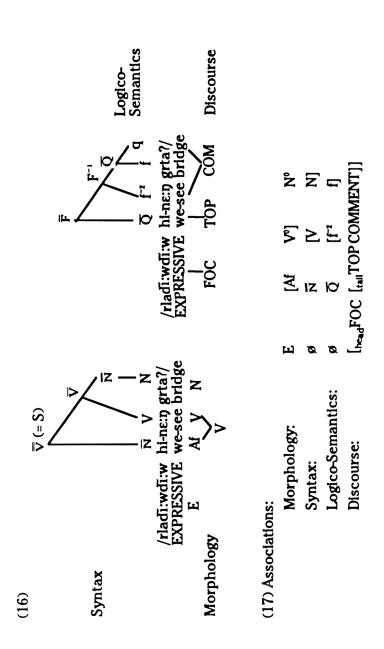
bald head, big and smooth like a papaya'

Morphology: E Logico-Semantics: nil

Syntax: 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.



The associations are all in conformity with the default relationships between dimensions discussed in 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 in frameworks which require everything to have a syntactic and semantic function, I think that the Autolexical approach produces descriptions of "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" (Schiller 1991c) and the related question of verb serialization (Schiller 1991), expressives (maybe?) 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.

Acknowledgements

I would like to thank Martha Ratliff, Jerry Sadock, Jim McCawley, Sokhoum Khaek, Gérard Diffloth, Zixin Jiang and Bill Croft for many helpful and useful discussions on the matters covered in this paper. None of them is to be held responsible for any errors of commission, omission, or sloppy thinking that might exist in this or any other version of the material presented here. Comments and criticism are welcome by snail mail to Eric Schiller, Department of Linguistics, University of Chicago, 1010

East 59th Street, Chicago IL 60637 or via e-mail to schiller@sapir.uchicago.edu.

Notes

- ¹ This spelling is used for the classifier only. Headley (1977) points out the similarity to Malay anak 'child', which better fits the <?nak > spelling.
- ² As pointed out in Schiller (1988), the pronominal system of Khmer is much messier than the literature indicates.
- ³ The hedge is due to the fact that /nea?/ is highly restricted as a pronoun in many dialects.
- ⁴ In the presentation of this paper I did not take a correct position on the category of this item, listing it in the handout as a preposition. I am now convinced that it is not, based on new evidence presented in Schiller (1991c).
- ⁵ See Eilfort & Schiller (1990) for a discussion of this particle.
- ⁶ Lit. 'miss (passing by or beyond)' There may be an understood deleted object here.
- ⁷ See (Lekawatana, 1975) for discussion of comparable Thai construction.
- ⁸ The terms 'head' and 'tail' were introduced in Jerrold Sadock's 1991 Autolexical Seminar at the University of Chicago.

References

- Benedict, Paul K. 1975. Austro-Thai: language and culture, with a glossary of roots. New Haven: HRAF Press.
- Chomsky, Noam. 1981.Lectures on Government and Binding. Dordrecht: Foris.
- Croft, William. 1991. Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information. Chicago: University of Chicago Press.
- Diffloth, Gérard. 1972. "Notes on Expressive Meaning." 1972. In The Best of CLS, ed. Eric Schiller et. al. 313-328. Chicago: Chicago Linguistic Society.
- Diffloth, Gérard. 1976. "Expressives in Semai." In Austroasiatic Studies, ed. L.C.Thompson & S.Starosta P.N.Jenner. 249-

- 164. Honolulu: University of Hawaii Press.
- Eilfort, William and Eric Schiller. 1990. "Pragmatics and Grammar: Cross-Modular Relations in Autolexical Theory." Proceedings of the 26th Annual Regional Meeting of the Chicago Linguistics Society, Vol.1.
- Headley, Robert K. 1977. Cambodian English Dictionary. Washington DC: Catholic University of America Press.
- Jacob, Judith M. 1968. Introduction to Cambodian. London: Oxford University Press.
- Lekawatana, P. 1975. "The So-Called Passive in Thai." In A Tai Festschrift for William J. Gedney, ed. W. Gething. Honolulu: University of Hawaii.
- Ratliff, Martha. 1991. "Cov, the Underspecified Noun, and Syntactic Flexibility in Hmong." JAOS 111.4: 694-703.
- Sadock, Jerrold M. 1991. Autolexical Syntax: A Theory of Parallel Grammatical Representations. Chicago: University of Chicago Press.
- Schiller, Eric. 1984. Syntactic Polysemy and the Lexicon in Khmer. M.A. Thesis, University of Chicago.
- Schiller, Eric. 1989. "Syntactic Polysemy and Underspecification in the Lexicon." BLS 15 (1989): 278-290.
- Schiller, Eric. 1990. "The Typology of Serial Verb Constructions." Proceedings of the 26th Annual Regional Meeting of the Chicago Linguistics Society, Vol.1.
- Schiller, Eric. 1991. An Autolexical Account of Subordinating Serial Verb Constructions. Ph.D. thesis, University of Chicago.
- Schiller, Eric. 1991b. "An Autolexical Account of Oceanic Verbal Prepositions." 6th International Conference on Austronesian Linguistics, Honolulu. May 21-24,1991. To appear in Chicago Working Papers in Linguistics, Volume 7 (1991).
- Chicago Working Papers in Linguistics, Volume 7 (1991). Schiller, Eric. 1991c. "Coverbs: a Dissent". Paper presented at the meeting of the Michigan Linguistics Society, Ann Arbor. October 4, 1991. To appear in Chicago Working Papers in Linguistics, Volume 8 (1992).