Syntactic Polysemy and Underspecification in the Lexicon
Author(s): Eric Schiller
Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society (1989), pp. 278-280

Syntactic Polysemy and Underspecification in the Lexicon[0]

Eric Schiller
University of Chicago

The term Syntactic Polysemy describes a phenomenon found in a wide variety of the world's languages, where a single lexical item is found in many different syntactic positions. Unlike "zero-derivation", there is great difficulty in determining the underlying root and derived forms, and the semantic relationship between two syntactic uses of the same lexeme is not predictable. In Schiller (1984) it was suggested that the lexicon of a language which has this phenomenon is un-, or at least under-specified with regard to syntactic category membership. This proposal did not fit comfortably with most lexicalist syntactic frameworks, as syntactic properties of lexical items were "projected" from the lexicon into the syntax itself. Without a lexical foundation, the syntactic framework had no information to use in building tree structures.

Within the framework of Autolexical Syntax, however, that concept of the lexicon is not only workable, but brings significant benefits analytically. The purpose of the present paper is to explore those benefits and describe in more formal terms the approach suggested in previous work. Data from Khmer, the major language of Cambodia, will be examined.

The Khmer word baan appears in a variety of syntactic and semantic contexts. It is used as a main verb, an auxiliary verb, a noun, and in other positions as well.

As a main verb, baan means 'to get; obtain', for example:

(1) khñom cɑŋ baan pɔɔŋ mɔ̃an
I want get egg chicken
I want to get some chicken eggs.

Another common use of baan is as an aspectual marker of completion.

(2) miin baan teñ ʔəywan nuh
aunt get buy things those
Aunt bought those things.

Jacob (1968) cites three uses of baan :

i) as a main verb meaning 'to get, to obtain'.

ii) as an aspectual marker indicating a completed action. McCawley (personal communication) points out the similarity between baan and English "manage". The semantics of baan seem to combine MANAGE and GET. baan has been described as marking a completed past action when taking a VP complement, rather similar to the use of English 'got'.

iii) in final position 'to be able', as in

(3) khñom thvəə kaa mɯn baan
I do work not able
I cannot work.

This use of baan in final position is interesting. Verbs are not usually found in that position. It might be suggested that this configuration arose just to avoid ambiguity with the aspectual usage, e.g.

(4) khñom mɯn baan thvəə kaa
I didn't do work.

There are two problems with this analysis. The first is that this sort of ambiguity is not necessarily avoided in Khmer, which relies heavily on context to distinguish between various interpretations of an utterance. In addition, adverbial clauses modifying a proposition are often found at the end of Khmer sentences. So it seems reasonable to adopt the position that when baan is found at the end of the sentence, in an adverbial position, it is behaving syntactically as an adverbial, and assigning a property of, say, possibility, to the entire sentence. In other words, if it walks like a duck, quacks like a duck, and looks like a duck, it probably is a duck.

Our next example demonstrates something resembling prepositional usage, although the constituent headed by baan can also be analyzed as a $\overline{V}$:

(5) a. kɔ̄at riən khmae baan pii ʔaatət
prn. study Khmer get two week
He studied Khmer for two weeks.
b. baan pii ʔaatət kɔ̄at riən khmae
For two weeks, he studied Khmer.

(5b) shows the possibility of fronting of the temporal phrase, while the contrast in (6) is due to the difference in constituent structure, where the phrase is internal to the objective nominal phrase:

(6) a. kɔ̄at məəl siəwphəu baan pii kbaal
prn. read book get two head[CL]
He read two books.
b. *baan pii kbaal kɔ̄at məəl siəwphəu

In (6a) the books must have been read to the finish, thus bringing into play some of the aspectual nuance found in the auxiliary verb position. This construction might therefore be alternatively analyzed as a resultative construction with baan as the marker (cf. discussion of ʔaoy toward the end of this paper). The grammatical description in Ehrman and Sos (1972) is that this usage is restricted to appearance of baan 'before a quantity of completed work'.

Let us now consider the various uses of the word baan from an Autolexical viewpoint. In the framework of Autolexical Syntax (Sadock 1983, 1985, 1986), the grammar of natural language is represented in a set of autonomous modules (Syntax, Semantics, Morphology, and one or more modules, for example, for pragmatic and discourse representations), each of which contains a representation of an utterance. The Interface is where the representations of the individual modules are compared. The "interesting" phenomena of grammar are those which have representations in two or more components which are not isomorphic. An important, indeed critical, point is that a lexeme need not have a representation in all components. For the purposes of the present paper, only the syntactic and semantic components need concern us, although mention will be made of some purely pragmatic factors.

There is little to say about the morphology. Essentially, Khmer words are all of the same morphological category, and cannot be inflected. Compounding is possible, however, and quite prolific. There are some items which appear to have affixes (prefixes and infixes), but these are simply vestiges of morphology from the era of Old and Middle Khmer, which had productive affixation. New derivations are extremely rare, and it is safe to say that these are formed by analogy. We will therefore assign a default specification in the morphological component such that all lexical items are stems, subject to compounding. There is, however, productive reduplication of complete morphemes (partial reduplication, formerly productive, is completely fossilized now) usually resulting in a semantic nuance of intensification, pluralization or frequentive aspect.

The syntactic representation is a monostratal, non-transformational Context-Free Phrase Structure Grammar with an IDLP framework and slash categories, as in GPSG. Semantics, taken here to be the dimension of predicate-argument structure and quantification (possibly of thematic roles and case as well) will be represented by a tree structure which has dominance relationships but which lacks any intrinsic linear ordering.

The syntax of Khmer seems especially complicated, given the lack of inflection which in other languages can help to solve questions of analysis. Since many words are syntactically polysemous, promiscuously occupying any slots into which their meaning (as opposed to combinatoric semantics) reasonably allows, I will adopt the position of Radical Syntactic Polysemy:

(7) *The default syntactic condition is unspecified with regard to syntactic category.*

This is to say that words are syntactically free to attach to any terminal node in the syntax. There are numerous exceptions of course, but they seem to be principled. Pronouns are a fixed category, although the situation with reflexives is unclear (Schiller to appear). Borrowed words, especially those of Sanskrit or Pali origin, seem to have fixed lexical categories. (Although many languages, including English, seem to apply some special rules to borrowed words, this is by no means universal. Tibetan is syntactically polysemous without regard to the origin of words ("There is no real distinction between different classes of words, and the same word can be used as a noun, adjective, or verb - all depends on its position in the sentence" (Roerich and Phuntshok 1971)), and even Sanskrit borrowings can be used as any part of speech (Agha p.c.). But in Khmer the pronouns and fixed class borrowings are a helpful diagnostic for syntactic analysis.

In addition, items which are historically the result of the prolific morphological processes which were productive until a couple of hundred years ago belong to a single lexical category. Affixes had nominalizing, causativizing, transitivizing and other grammatical functions.

The freedom enjoyed by lexical items is not matched by the set of node admissibility conditions which define grammatical structures in Khmer. The apparent complexity of syntactic expression is, I suggest, a result of attempting to maintain an analysis in which category membership for lexical items is fixed. When one permits the lexical items to occupy a variety of syntactic slots, then the number of

syntactic constructions can be greatly constrained. In fact, the syntax of Khmer, often regarded as horribly complex, may not be so difficult after all, if we permit non-isomorphism of syntactic and semantic trees, and allow items to have no representation in certain components (so that words with purely pragmatic function do not receive a representation in the syntax, for example.) These assumptions are standard Autolexicalist positions.

In most Southeast Asian languages, context is a powerful tool which licenses omission of arguments in syntactic structure and which interacts intimately with both syntax and semantics. In this paper such interactions will not be explored, but two analytical positions must be stated so that the syntactic analysis will not seem deficient.

First of all, responsive particles are treated as lacking both syntax and semantics, and are used for purely pragmatic reasons, to affirm or deny propositions. Specifically, words which correspond to English 'yes' (/baat/ and /caa/, used by men and women respectively) and which are always utterance initial, and /tee/, a final particle used whenever an overtly or covertly negated proposition is part of the utterance or when a question is being asked, will not be assigned syntactic or semantic representations. Some discussion of tee will be presented below.

Also assigned to the pragmatic component(s) are such matters as choice of pronoun (involving a very elaborate status mechanism), honorifics, and illocutionary mechanisms. See Eilfort (PhD thesis in progress) for more on the illocutionary aspects of Autolexical theory.

I have argued elsewhere that a Khmer sentence should be analyzed as a $\overline{V}$, that is a maximal projection of a head verb (Schiller to appear). The following can be taken as a partial set of the rules required for Khmer by the syntactic component:

(8) SR1: $\overline{V} \rightarrow (XP)\ \overline{V}$ (for topic structures)
SR2: $\overline{V} \rightarrow (\overline{N})\ \overline{V}$ (i.e. Khmer is a pro-drop language)
SR3: $\overline{V} \rightarrow V_{[aux]}\ \overline{V}$
SR4: $\overline{V} \rightarrow \overline{V}\ \overline{V}$ (for serial verb structures)
SR5: $\overline{V} \rightarrow V\ (\overline{N})$
SR6: $\overline{N} \rightarrow \overline{N}$ (Num) (Cl) (Dem) (i.e. Khmer is a classifier language)
SR7: $\overline{N} \rightarrow N\ (N)\ (A)$
SR8: $\overline{P} \rightarrow P\ \overline{N}$
SR9: $V^n \rightarrow V^n$ Adv
(Any constituent headed by V can be followed by an adverb)

Note that Khmer is clearly a head-first language. I have omitted discussion of quantifiers for the purposes of this paper, because Khmer shows negative-raising effects and taɛ 'only' shows rather remarkable clitic-like propertiers which deserve full discussion elsewhere (e.g. Schiller *f.c.*).

Returning to baan, it may seem that one might extend the notion of GET of the word baan to obtain a passive-like element as well, though few books or dictionaries list such a function. In fact, there is a fairly common construction which parallels the English "get" passive:

(9) a. Sok baan rɔ̃atta kaa cuəy ʔaoy təu riən
Sok get government help give go study
Sok got government help so that (he could) go to study.

The various uses of the word clearly share something in common, and deserve investigation in the metaphorical framework of Lakoff (1987). But sticking to just the syntax, on the basis of the data presented above, a first approximation of a lexical listing for baan might be:

(10) /baan/

to get	$V[\overline{V}/\overline{N}]$	F^{-2}
get-passive	$V[\overline{V}/\overline{V}]$	F^{-2}
to have done	$V_{[+AUX]}[\overline{V}/\overline{V}]$	M
can-do	$A[\overline{V}/\overline{V}]$	M

where F^{-2} indicates that in the combinatoric semantics, the item is a function which combines with two arguments to form a proposition, while M indicates that in the combinatoric semantics the item is a modifier. The listing above exemplifies the property of Syntactic Polysemy that I claim holds for certain languages. Translation clouds the issue, but it seems clear that these items are related, but not necessarily derived in a traditional sense. There are three reasonable strategies:

1. List every syntactic usage in the lexicon.

2. Have lexical redundancy rules which create lexical items of different categories from a root of a single category.

3. Allow for some sort of underspecification which allows the lexical item to remain syntactically promiscuous.

I think that a combination of the first and the third paths is clearly preferable. The second option is clearly empirically inadequate, since there is a great deal of variation in the degree of syntactic promiscuity. Were the semantic consequences of the zero-derivation process clear, this might be a reasonable try. Even so, one is hard pressed to explain the distribution of various syntactically polysemous items via zero-derivation, as one would have to assume that all speakers of the language chose to apply this optional derivational path to almost exactly the same range of phenomena. Zero-derivation can explain the possibility of syntactically polysemous items, but cannot account for restrictions seen in, for example, the case of baan.

But if we take the third route, we can underspecify the lexical item in a number of different ways, allowing for certain syntactic features (here we adopt the GPSG mechanism of decomposing syntactic categories into feature specifications) to be free in polarity. Naturally many of the uses will become lexicalized over time, and not only in terms of the syntactic category, but also in the form of idioms. We can adopt the position suggested earlier that the lexical syntax in Khmer is underspecified. We can go even further, and suggest that the semantics has a default relationship to the syntax.

A *default*, in the sense intended here, refers to a set of relationships which are presumed to hold betweeen syntax and semantics. For example, a transitive verb is expected to have two logical arguments in the semantic component. An intransitive verb should have a single argument. An adjective or adverb is supposed to modify a semantic constituent. A nominal element in the syntax should be a quantified entity in the semantics.

Since baan appears in a wide range of syntactic and semantic uses, and each of those uses is normal (that is, when found in a given syntactic slot the semantic function is that which is appropriate to that position), what we have then, is the following entry:

(11) /baan/ GET
Syntax: [X $_{V}$ _____ Y]
Semantics: *default*

The entry shows that the phonological form /baan/ has a general meaning something like 'get' (Generalizing across languages and syntactic forms is difficult without an adequate metalanguage, and this is purely a descriptive device, at present.) which can appear in any syntactic environment which forms a verb phrase. Note that no syntactic category is listed. The syntactic entry allows any terminal node meeting the requirements of the listed specification to accept the lexical item.

Thus the lexical conditions for the word will be met whenever it appears in a syntactic position which holds a normal relationship with the representation of the lexical item in the semantic component. So we can expect that it is possible to use a lexical item as an N, V, A, or P. But if the categorial cabinet is so empty, why then is there such a great statistical tendency to find the word baan used in main verb or auxiliary position? The answer lies in the non-combinatoric semantics, i.e. the real "meaning" of the word. In short, there is no prediction that all items will, in fact, be found in all positions, merely that the underspecification allows it to occupy these positions. A combination of the real "meaning" of the word and the Grounding Principle mentioned below is enough to account for the distribution of most underspecified lexical items. This also leaves open the possibility that two speakers might choose different representations when hearing a string, say, a prepositional phrase construction as opposed to verb+object, which have the same surface syntactic forms in Khmer (see discussion of nəu below). In the following examples, the notation is to be read as follows (using 12a. as an example): The lexical item baan is attached to the node label V, forming a verb phrase when combined with a noun phrase to its right. In (12) I have added a reference to the particular syntactic rule in which the word participates.

(12) a. baan kaa(r) 'reliable, able, capable, sure to bring good results'
V:$\overline{V}$/___$\overline{N}$ (as a transitive verb: SR5)
kŏat baan kaa
prn. GET act/result
He is a reliable person.

b. baan koun proh pii nĕak 'to have two sons'
V:$\overline{V}$/___$\overline{N}$ (as a transitive verb:SR5)
kŏat baan koun proh pii nĕak
prn. GET child male two persons
He has two sons.

c. baan cətt 'to become more daring"
V:$\overline{V}$/___$\overline{N}$ (as a transitive verb:SR5)
kraoy Sok lɯɯ rɯəŋ nih baan cətt nah
after Sok hear story this GET heart very
After Sok heard the story, he became bolder.

d. V:$\overline{V}$/___$\overline{V}$ (as an auxiliary verb:SR3)
kɔ̃at mɯn baan mɔɔk tee
he not GET come POL
He couldn't come.

e. A:$\overline{V}$/V___ (as an adverb: SR9)
kɔ̃at cɑŋ dəŋ thaa nɛ̃ak naa klah ʔaacbaoh chnaotbaan
he want think say person any some can throw ballot able
He wants to know who is eligible to vote.

(12a) seems idiomatic, closest in meaning to the main verb use 'obtain', i.e. 'He gets results', which involves an acceptable translation of kaa. (12b) only allows baan to be used as the matrix verb. (12c) is interesting because of the modifier 'very' which must go with the verb phrase, since there is no justification for translating cətt as 'bolder'. Here baan has its main verb syntactic use encoded in the idiom. In (12d), it seems that the meaning 'can' is as much involved as the completive aspectual meaning, which just illustrates the point that there isn't a set of separate meanings involved here, but rather a single lexical item with a wide but non-divisible range of meanings. In (12e) we find that the the ability to cast a vote is predicated from the lower clause, and that no long-distance relationship with the subject of the higher clause is possible.

(12) shows syntactic polysemy on a limited scale, with only verb-phrase forming syntactic functions permitted. trəw has the meanings 'hit, come into contact with, experience, must, should, correct, right', and appears in an even wider range of configurations. It is sometimes claimed to be a marker of Passive, although this analysis has been properly criticized by Lekawatana (1975).

(13) a. V:S/$\overline{N}$___ (as an adjective)
cɑmlaəy nuh mɯn trəw tee
answer that not correct POL
That answer is not correct.

b. V:$\overline{V}$/___$\overline{N}$ (as a transitive verb)
puuthau crɔluəh mɔɔk trəw cəəŋ
ax slip come hit leg
The ax slipped and hit his leg.

c. V:$\overline{V}$/___V (as an auxiliary verb)
kɔ̃at trəw puukaɛ
he should be-skillful
He ought to be skillful.

d. V:$\overline{V}$/___$\overline{V}$ (as an auxiliary verb)
khñom trəw təu phsaa thŋai nih
I must go market day this
I must go to the market today.

e. V:$\overline{V}$/___S (as a verb which takes sentential complements)
kɔ̃at trəw chkaɛ kham
he experience dog bite
He got bit by a dog.

f. V:$\bar{V}$/___$\bar{\bar{N}}$ (as a transitive verb, syntactically)
kɔ̃at trəw krɔ̃ap
prn. experience/hit bullet
He was struck by a bullet.

g. N:$\bar{V}$/V___ (as a noun)
mənuh nuh dəŋ khoh trəw
man that know wrong right
That man knows right from wrong.

h. N:$\bar{P}$/P___ (as a noun)
mənuh nuh dəŋ khoh pii trəw
man that know wrong from right
That man knows right from wrong.

i. A:$\bar{V}$/V___ (as an adverb)
kɔ̃at chlaəy sɑmnuə trəw
he answer question correctly
He answers the question correctly.

Since we see the word in every syntactic environment except that of a preposition, we can then provide the following lexical entry.

(14) /trəw/
Syntax: [-P]
Semantics: *default*

There are a few noteworthy observations to be made. First of all, the use of this item in this 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 completely unrelated to Khmer (following Benedict's Austro-Thai) or very distantly related (as I argued at this very conference two years ago.)

One fact about the use in each language is that it cannot be used 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.

So 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 have semantics rather similar to adverbs. 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 following entry, taking [-Prepositional] as a default:

(15) /trəw/
Syntax: *unspecified*
Semantics: *default*

Now consider another fact which at first seems to counterexemplify the proposed analysis. We do not find kɨt '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 kɔmnɨt 'thought', a form created during the period of 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.

(16)*If an appropriate lexeme exists in the lexicon, do not use the underspecified form instead.*

It should be mentioned that Jerry Sadock has been working with the idea of primacy of various components, such that where conflicts exist, the more concrete component seems to force a resolution in its favor, e.g. when Morphology and Syntax collide, Morphology wins.

One must keep in mind, however, that Syntactic Polysemy is a phenomenon which appears in natural languages to different degrees, and that no claim is being made that there is complete freedom in the lexicon. It seems that in most languages, personal pronouns form a fixed nominal class. It is reasonable to assume that the process of acquisition of Khmer syntax involves observations that some words, for example the personal pronouns, only appear in a nominal setting, while other words appear with a wider variety of uses. It is relevant, perhaps, that the most promiscous lexical items are very basic to the vocabulary. In addition to baan and trəw, considered above, here are a few more syntactically polysemous common words with representative examples (by no means an exhaustive list, either of examples, or of examples of examples):

(17) nəu 'IN'

a. kʊ̃at rʊ̃ah nəu phnum peñ
prn. (a)live in Phnom Penh
He lives in Phnom Penh.

b. kʊ̃at nəu phteah
prn. in house
He is in the house.

c. kʊ̃at nəu rien khmaɛ
prn. in study Khmer
He is still studying Khmer.

d. nəu knoŋ sɑmot mien trəi craən
in inside sea have fish many
In the sea there are many fish.

(18) ʔaoy 'GIVE'

a. kʊ̃at ʔaoy luy khñom
prn. give money me
He gives me money.

b. som niʔyiey ʔaoy cbah
please speak so-that clear
Please read clearly.

c. kʊ̃at thvəə ʔaoy pibaaʔ
prn. do so-that difficult
He made it difficult.

d. kɔ̃at teñ trəi ʔaoy khñom
He buy fish for me
He bought the fish for me.

e. ʔaoy taɛ taok khñom təw
give only inexpensive I go
If it's cheap, I'll go.

(19) daoy

a. kee mɔɔk daoy laan
prn. come by car
They came by car.

b. daoy yɔɔbɔl khñom kɔ̃at trəw təu thvəə kaa
according opinion I prn. must go do work
In my opinion, he ought to go to work.

c. khñom daə daoy tɔnlee
I walk by/along river
I walk along the river.

d. siəwphəu nih sɑɔsee daoy puu Sok
book this write by uncle Sok
That book was written by my uncle.

e. kɔ̃at thvəə daoy piʔbaa
prn. do with difficulty
He did it with difficulty.

f. kɔ̃at trəw rien daoy kɔ̃at mɯn cɛh khmaɛ
prn. should study becausehe not know Khmer
He has to study because he doesn't know Khmer.

I have no time to discuss these examples, or many others with similar wide ranges of application, but present them for your information. Another word, təu will figure in the discussion below.

The theory of Autolexical Syntax is still fairly new, and its ongoing development may be of interest to some of you. But there are immediate benefits in using this framework for descriptive purposes. Perhaps the most important benefit is the elimination of the proliferation of syntactic categories (or, in another framework, cases) which are employed for the description of isolating languages. Let us take an example from Khmer, namely, the grammatical descriptive devices employed in Jacob (1968).

> *"A noun is a word which may occur immediately following pre-nominal particles (q.v.)."*

Her pre-nominal particles are words such as pii 'from', ʔaɛ 'at', knoŋ 'in'.

> *"They may precede noun constructs immediately and thus form adverbial constructs which may occur in several different positions in the various sentence-forms and may be pronounced with separate phrasing..."*

She provides a list of 48 items which function as pre-nominal particles. Of those, 27 are syntactically polysemous, and 18 are compounds with at least one part showing syntactic polysemy. So, for example, nəu-knoŋ : nəu is used as a main verb meaning 'to be located at' and is also used as a preposition meaning 'in'; knoŋ is restricted to pre-nominal positions. nəu knoŋ is listed, as is each of the components. But in almost all of the listed compounds,

the first component is also available for use as a verb, and the second portion varies tremendously. So the second component can be analyzed as the complement or object of the verb, and often there is evidence to support this. For example, knoŋ is suspiciously nominal in its form (vestiges of nominal nasal infix). In any event, this heterogeneous group of pre-nominal particles is very hard to define, and thus cannot adequately function as a test for noun-ness. Jacob includes among the functions of these particles that

> *"They act as ad hoc nominalizers of words of other categories (verbs, numerals, adverbs, and even of miniature sentences)."*

In other words, what follows a pre-nominal particle is interpreted syntactically as nominal.

(20) khñom təu ciə rɔhah
I go be fast
I shall go quickly.

Now the problem here is that both təu and ciə are generally treated as verbs. Under Jacob's analysis, the latter acts as a nominalizer, although it seems justified only because one can replace ciə rɔhah with a garden variety nominal like phtẽah 'house':

(21) khñom təu phtẽah
I go house
I'm going home.

But the analytical difficulties quickly multiply with the addition of a few more words:

(22) a. khñom təu knoŋ phtẽah
I go in house
I'm going into the house.

b. khñom daə təu phtẽah
I walk go house
I'm walking home.

c. khñom daə təu knoŋ phtẽah
I walk go in house
I'm walking home.

In these examples, təu seems to function sometimes as a verb, and sometimes as a preposition. Other analyses involve the "coverb" concept and this analysis might work for təu, but is less appropriate for knoŋ, which is sometimes treated as a "relator noun". This latter concept seems sensible in view of the frequent use of the word to mean "inside", or "the inside of..." Suppose we try to simplify our analysis by treating knoŋ as a noun. In (22c), we have an example of $\bar{N} \rightarrow N\ N$ which is a frequently seen construction in Khmer. But syntactically, knoŋ is probably a preposition, since it cannot be modified in any way and cannot take a determiner. We could, of course, try to classify it as a pronoun, but this seems intuitively weird from a syntactic point of view, and in any event it cannot appear in some pronominal positions such as possessor. There is still the possibility of a syntactic class of relator nouns. In any case we can maintain təu

as a verb (in the syntactic representation), with daə təu analyzed either as a compound or serial verb construction , a decision which will not be taken here.

In a similar fashion we need not create a lexical category for numeral classifiers, as these can be either nouns or, less frequently, stative verbs. But not all categories can be collapsed into the standard set of syntactic classes of noun, verb, adverb, preposition. Khmer has certain particles which are used only at the beginning of an utterance (such as 'yes', 'no') and a set of emphatic particles. In the analysis proposed here, these items would have NIL syntax, rather than the default case. They can be treated as terminal nodes of a pragmatic or discourse tree. This also applies to the politeness particle tee which has sometimes been treated as part of a discontinuous structure of the negative operators mɯn and ʔɔt in addition to its use as a "question particle". Given the following set of data one can easily understand how tee became analyzed as a negative particle.

(23) a. khñom min təu phtēah tee
I not go house POL
I'm not going home.

b. look təu phtēah tee
Mr. go house POL/QUERY
Are you going home?

c. look təu phtēah rɯɯ tee
Mr. go house or POL/QUERY
Are you going home, or not?

d. look təu phtēah rɯɯ mɯn təu phtēah tee
Mr. go house or not go house POL/QUERY
Are you going home, or not?

The last example is a very typical areal construction for questions. By treating tee as a pragmatic entity with no syntactic or semantic representation, the remainder of the sentence poses no problem. Again, simplicity within each component is maintained.

In this brief paper an Autolexical treatment of Syntactic Polysemy has been outlined. The claim of underspecification or even possibly unspecification in the lexicon is strong, but not unique to Autolexical Syntax (see, for example, the HPSG application of underspecification in Pollard and Sag (1987). The combination of Autolexical Syntax with the concept of lexical underspecification gives rise to optimism concerning the description and analysis of isolating languages. Syntactic Polysemy can also be applied to morphologically productive languages such as Nootka, where the roots seem to be syntactically underspecified. Areas of investigation which might benefit from this approach are the problem of serial verb constructions, verb concatenation (cf. Matisoff 1973) and potential explanations for (Hawkins 1983) proposed implicational word order universals.

Endnotes

0 I would like to thank James D. McCawley and Jerrold M. Sadock for their helpful comments both on my 1984 M.A. Thesis, where I first discussed the syntactically polysemous nature of the Khmer Lexicon, and in many courses and conversations over the past 6 years. Gérard Diffloth introduced me to, and guided

my study of Khmer with a syntactically skeptical outlook, which helped me to avoid many analytical traps. Sokoum Khek has been a perfect informant for many years, and Im Proum, Kim Sue Lee, Sina Tan and other teachers and students at the 1987-1988 Southeast Asian Summer Studies Institutes provided invaluable additional data. Ivan Sag encouraged me to look to HPSG for analytical devices which resemble those presented here. My fellow students at the University of Chicago have stimulated my thoughts and helped to correct errors of omission and commission. Paul Deane, Bill Eilfort, Ligang Li, John Richardson, Zixin Jiang and Hans Smessaert have had a direct influence on the work presented here. All of the above are absolved from any complicity in any errors of fact or analysis which appears in this paper.

References

Ehrman, M. E. and K. Sos. (1972). Contemporary Cambodian: grammatical sketch. Washington, DC, Foreign Service Institute.

Hawkins, J. A. (1983). Word Order Universals. Quantitative Analyses of Linguistic Structure. New York, Academic Press.

Huffman, F. E. (1970). Modern Spoken Cambodian. Yale Linguistic Series. New Haven, Yale University Press.

Jacob, J. M. (1968). Introduction to Cambodian. London, Oxford University Press.

Kheang, L. H., M. E. Ehrman and K. Sos. (1974). Contemporary Cambodian: the social institutions. Washington, DC, Foreign Service Institute.

Lakoff, G. (1987). Women, Fire and Dangerous Things. Chicago: University of Chicago Press.

Lekawatana, P. (1975). The So-Called Passive in Thai. A Tai Festschrift for William J. Gedney. Honolulu, University of Hawaii.

Matisoff, J. A. (1973). The Grammar of Lahu. University of California Publications in Linguistics. Berkeley, University of California Press.

Pollard, C. and I. Sag. (1987). Information-Based Syntax and Semantics Volume 1: Fundamentals. CSLI Lecture Notes #13. Stanford, CSLI.

Roerich, G. N. and T. L. Phuntshok. (1971). Textbook of Colloquial Tibetan. New Delhi, Manjusri.

Sadock, J. M. (1983). "The necessary overlapping of grammatical components." CLS. 19(2): 198-221.

Sadock, J. M. (1985). "Autolexical Syntax: A Proposal for the Treatment of Noun Incorporation and Similar Phenomena." NLLT. 3: 379-439.

Sadock, J. M. (1986). "An autolexical view of pronouns, anaphora, and agreement." UCWPIL. 2: 143-164.

Schiller, E. (1984). Syntactic Polysemy and the Lexicon in Khmer. M.A. Thesis, University of Chicago

Schiller, E. (to appear). "The Parametrics of INFL." ESCOL. 5

Schiller, E. (f.c.). "Heading in the Right Direction: An Autolexical Account of Numeral Classifier Constructions".