Flying 'In' and 'Out' in Khmer and Thai

Chhany Sak-Humphry

Kitima Indrambarya*

Stanley Starosta

University of Hawai'i at Manoa

1. INTRODUCTION

By definition, a word is a minimal free form. However, to identify forms as words in isolating languages which do not have inflectional morphology is not an easy task. The Saussurean sign regards sound and meaning as the basic components of a word. Semantic difference is thus a basic criterion for differentiating words of the same shape and pronunciation (Panupong, 1978, p. 217). In the case of homophones, forms like English *run* (n) and *run* (v), however, which are similar in both pronunciation and meaning, distribution must play a significant role in distinguishing lexical entries from each other. Thus, distribution is treated as one of the components of words in Lexicase theory (Starosta, 1988, in press). Panupong (1978, p. 221) summarizes the advantage of assuming such homophones as follows:

> a) To accept this analysis, it is necessary that we regard each word as having only one function. The merit lies in our being immediately able to decide to what class a word in each sentence belongs,...
>
> b) Not having to set up classes for polyfunctional words is much more economical.
>
> c) There would be no problem in labelling the words with more than one function.

This paper attempts to establish the word classes of the following forms carrying the meaning 'enter; in' and 'leave; out,' when they are preceded by main verbs glossed as 'to fly' in two Southeast Asian languages, namely Khmer and Thai. They are *caul* 'enter' and *ceñ* `leave' in Khmer and *khâw* 'enter' and *ʔɔ̀ɔk* 'leave' in Thai.[1] The following sentences containing the forms to be tested in the paper:

---

* Now at Kasetsart University, Bangkok.

[1] Pending the proper assignment of word classes to these forms, we will gloss them in the first part of the paper as 'enter' and 'leave.'

Khmer

(1) a. caap haoe *caul* knong trung
bird fly enter inside cage
'The bird flew into the cage.'

(2) a. caap haoe *ceñ* pii trung
bird fly leave from cage
'The bird flew out of the cage.'

Thai

(3) a. nók bin *khâw* pay nay kroŋ
bird fly enter go inside cage
'The bird flew into the cage.'

(4) a. nók bin *ʔɔ̀ɔk* càak kroŋ
bird fly leave from cage
'The bird flew out of the cage.'

In the lexicase grammatical framework, there are eight and only eight basic syntactic word classes: V (verb), N (noun), Adj (adjective), Det (determiner), Adv (adverb), P (preposition or postposition), Cnjc (conjunction), or Sprt (sentence particle) (Starosta, 1988, p. 51). In principle, based on their distributions and meanings, the forms *caul, ceñ*, *khâw* and *ʔɔ̀ɔk* in the preceding sentences could be thought to be (1) prepositions, (2) adverbs, or (3) verbs in Khmer and Thai, as indicated by the following alternative glosses:[2].

Khmer

(1) a. caap haoe *caul* knong trung
bird fly enter inside cage
'The bird flew into the inside of the cage.'
'The bird flew in to the inside of the cage.'
'The bird flew to enter the inside of the cage.'

(2) a. caap haoe *ceñ* pii trung
bird fly leave from cage
'The bird flew out of the cage.'
'The bird flew out from the cage.'
'The bird flew to leave from the cage.'

[2] Other parts of speech are ruled out for the following reasons: (1) adjectives and determiners may be dependents of nouns, not of verbs such as *bin* and *haoe* 'to fly,' (2) conjunctions would coordinate words of the same category, not verbs and nouns, and (3) sentence particles occur sentence-finally unless followed by another sentence particle.

Thai

(3) a. nók bin *khâw* pay nay kroŋ
bird fly enter go inside cage
'The bird flew to the inside of the cage.'
'The bird flew in to the inside of the cage.'
'The bird flew to enter the inside of the cage.'

(4) a. nók bin *ʔɔ̀ɔk* càak kroŋ
bird fly leave from cage
'The bird flew from the cage.'
'The bird flew out from the cage.'
'The bird flew to leave the cage.'

The paper is divided into four sections. This first section presents an introduction. The second section discusses the criteria used to identify prepositions, verbs, and adverbs in this paper. The third section discusses how each test applies to the words in question. The fourth section presents the conclusion.

2. TESTING CRITERIA

The tests to be used for sentences in both Khmer and Thai are: (1) Stranding,[3] (2) Joint topicalization of both the word in question and the following noun phrase, referred to here as PP topicalization, (3) Choice of negation expressions, and (4) Root predicate with the choice of negation expressions used for verbs. These tests are built upon the tests used in Indrambarya (1995) and are extended here for testing other languages.[4]

2.1. Stranding

Prepositions and relator nouns cannot be stranded[5] while verbs and adverbs can (Indrambarya, 1995). From the point of view of lexicase theory, prepositions cannot be stranded because a prepositional phrase is an exocentric construction, that is, a construction whose head takes one or more structurally obligatory dependents, while a relator noun can't be stranded because there is no mechanism for passing

[3]This test is referred to as the "stranding test" rather than as the "'topicalization of the following NP test" as in Indrambarya (1995), because words following the form in question could be a prepositional phrase as well as a noun phrase.

[4]Examples of uncontroversial words for verb, adverb, prepositions, and nouns in Khmer are discussed in Sak-Humphry (1996).

[5]A noun may be differentiated from a preposition in that only the former allows a determiner as its dependent (Savetamalya 1989, Indrambarya 1994, 1995).

indices across nouns to link the relator noun with its topicalized dependent in a higher clause.

2.2. PP Topicalization

A preposition and its following NP and a relator noun with its dependent are grammatical units and may be topicalized together. However, a verb and its dependent and an adverb with a following noun cannot be topicalized together (Indrambarya, 1995). This latter claim does not follow from any universal grammatical principles, but it is a consistent generalization that can be made by observing the syntactic behavior of clear cases of prepositions, relator nouns, verbs, and adverbs.

2.3. Choice of negation word

Indrambarya (1995) has observed that different word classes are negated by different negation words when serving as predicates More specifically, in Thai the negation word *mâychây* negates NP and PP predicates, while *mây* and *mâydây* negate only verbs and (non-predicate) adverbs. Thus, if a form in question may be negated in the position in which it occurs, it is possible to separate verbs and adverbs from prepositions and nouns. The choice of a negation word only serves as a one-way test, though. That is, if a form can occur with one of the negation words, we have a clue as to its part of speech. However, if the form cannot be negated at all, the test cannot tell us what part of speech the form is . This situation arises because (1) a telic verb does not allow an embedded verb to be negated at all (, 1993a, p. 58), while an adverb does not necessarily occur with any negation word (cf. Indrambarya, 1995).

In Khmer, verbs can be negated with the negation expressions *mɨn*, *pom*, or *ʔat... (tee)* 'not,' while prepositions, verbs and nouns may be negated with the common negation expression *mɨn mɛɛn... tee*. Adverbs, on the other hand, cannot be negated at all. Negation in Khmer is summarized in Table 1.

Table 1. *Negation Pattern in Khmer*

	V	P	N
mɨn... tee 'not'	+	–	–
pom... tee 'not'	+	–	–
ʔat... tee 'not'	+	–	–
mɨn mɛɛn... tee 'not true'	+	+	+

2.4. Root Predicate with the Choice of Negation Expressions

This test consists of three parts. STEP ONE is to test whether a form may occur as a root predicate, that is, function as predicate in a simple sentence. By the lexicase Patient-to-Actor Control Rule for infinitival complements (P2a), if the form in question occurs after a verb and is a non-finite complement, it will have a higher Patient as its implied actor and therefore as its implied subject. Then, if the form in question may occur as a root predicate bearing this same higher Patient as its subject, we know that it may also be functioning as a predicate in the post-verbal position. Once we have found out that the form may occur as a predicate, the choice of negation word test is reapplied as STEP TWO to clarify and confirm the true syntactic category of the form when it occurs in that position. This is because verbs, nouns and prepositions may all function as predicates. If the form in the root predicate clause may occur with the negation word for verbs, it would then be identified as a verb. As an example, to see whether the forms *klây* 'near' and *lóm* 'fall' which occur after verbs in (5a) and (6a) are verbs, a root predicate test is applied as step one. As shown in (5b) and (6b), *klây* and *lóm* may occur as predicates. The choice of negation expressions in step two, shown by examples (5c) and (6c), further illustrates that the forms *klây* 'near' and *lóm* 'fall' in (5b) and (6b) are verbs.

(5) a.

bâan	chán	yùu	klây	rooŋrian
house PAT	I	stay	near	school

'My house is near school.'

(6) a.

dɛɛŋ	plàk	rûuppân	lóm
Daeng PAT	push	statue	fall

'Daeng pushed the statue fall.' [=Daeng knocked the statue over.]

Step One:

(5) b.

bâan	chán	klây	rooŋrian
house PAT	I	near	school

'My house is near school.'

(6) b.

rûuppân	lóm
statue PAT	fall

'The statue fell down.'

Step Two:

(5) c

bâan	chán	mây/*mâychây	klây	rooŋrian
house PAT	I	NEG	near	school

'My house is not near school.'

(6) c. rûuppân mây/ mâydây/*mâychây lóm
statue NEG fall
PAT
'The statue did not fall down.'

Now STEP THREE is to determine whether the forms *klây* and *lóm* in (5b) and (6b) are instances of the same word found in examples (5a)–(6a) or whether they are lexically distinct homophones. If they are the same lexical entries, they should have the same meanings and the same selectional restrictions. The first part of this question is often difficult to answer unless the meanings of the forms in the two environments are so different that all speakers and all linguists recognize the difference. The second part is a bit easier to use, however. It will be remembered that if the forms in question are verbs, their implied subjects will be identical with the Patient (PAT) of the preceding root verb, and any noun phrase occurring as the Patient of the higher verb should also be able to occur as the subject of these forms when they are used as root clause verbs themselves, otherwise adverbs. Since the forms *klây* 'near' and *lóm* 'fall' may occur as a root predicate bearing its original Patient as subject, the forms *klây* 'near' and *lóm* 'fall' in (5a) and (6a) are found to be verbs.

3. APPLICATION OF THE TESTS

This section applies the four tests discussed in section 2 to the sentences (1a)–(4a) containing the forms *caul*, *ceñ*, *khâw*, and *ʔɔ̀ɔk* to find out the syntactic categories of these forms. Table 2 summarizes the results of the application of each of the four tests to the forms *caul*, *ceñ*, *khâw*, and *ʔɔ̀ɔk*.

Table 2. *Summary of the Results of Each Test*

Forms	Strand	Top w/PP	Choice of NEG	NEG in Root Predicate
caul 'enter'	+	-	mɨn ... tee	+/-
ceñ 'leave'	+	-	mɨn ... tee	+/-
khâw 'enter'	-	-	*N/A*	+/-
ʔɔ̀ɔk 'leave'	+	-	N/A	+/-

3.1. Stranding

When *pii trung* in (1b), *knong trung* in (2b), and *càak kroŋ* (4b) are topicalized at the beginning of the sentences, leaving *ceñ*, *caul*, and *ʔɔ̀ɔk* stranded at the ends of the sentences, the examples remain acceptable, which shows that they are not prepositions or nouns. Instead, they could be verbs or adverbs.

(1) b. knong trung noh na caap haoe *caul*
inside cage that TOP bird fly enter
'Into that cage, the bird flew.'

(2) b. pii trung noh na caap haoe *ceñ*
from cage that TOP bird fly leave
'Out of that cage, the bird flew.'

(3) b. *pay nay kroŋ nán nàʔ nók bin *khâw*
go inside cage that TOP bird fly enter
'Into that cage, the bird flew.'

(4) b. càak kroŋ nán nàʔ nók bin *ʔɔ̀ɔk*
from cage that TOP bird fly leave
'Out of that cage, the bird flew.'

On the other hand, the test shows that sentence (3b), containing the stranding *khâw*, is unacceptable. There could be two possible explanations for the ungrammaticality of this example: (1) *khâw* is either a preposition or a noun, since it cannot be stranded at the end of the sentence; or (2) *pay* too is an adverb which does not form a constituent with the following NP, and hence cannot be topicalized with it. This possibility can be tested by stranding *pay*. Since *pay* can be left at the end of the sentence, as in (3b'), unlike *nay* and *càak*, *pay* is shown to be a verb or an adverb, while *càak* and *nay* are either prepositions or relator nouns, as in (3b") and (4b') respectively.

(3) b'. nay kroŋ nán nàʔ nók bin khâw *pay*
inside cage that TOP bird fly enter go
'Inside that cage, the bird flew into.'

(3) b". * kroŋ nán nàʔ nók bin khâw pay *nay*
cage that TOP bird fly enter go inside
'That cage, the bird flew into.'

(4) b'. * kroŋ nán nàʔ nók bin ʔɔ̀ɔk *càak*
cage that TOP bird fly leave from
'That cage, the bird flew out of.'

Further testing with the insertion of determiner suggests that *nay* is a relator noun and *càak* is a preposition (Indrambarya, 1995). *Pay* on the other hand is found to be an adverb when the root predicate test is applied (see below).

3.2. PP Topicalization

This test shows that we cannot topicalize the expressions *caul knong trung*, *ceñ pii trung*, *khâw pay nay kroŋ*, or *ʔɔ̀ɔk càak kroŋ*. This fact supports a claim that *caul*, *ceñ*, *khâw*, and *ʔɔ̀ɔk* could be adverbs or nonfinite verbs, but not prepositions or nouns.

(1) c. **caul knong trung noh na caap haoe*
enter inside cage that TOP bird fly
'Into that cage, the bird flew.'

(2) c. **ceñ pii trung noh na caap haoe*
leave from cage that TOP bird fly
'Out of that cage, the bird flew.'

(3) c. **khâw pay nay kroŋ nán nàʔ nók bin*
enter go inside cage that TOP bird fly
'Into that cage, the bird flew.'

(4) c. **ʔɔ̀ɔk càak kroŋ nán nàʔ nók bin*
leave from cage that TOP bird fly
'Out of that cage, the bird flew.'

3.3. Choice of Negation Expression

In Khmer, the negation pattern *min... tee* occurs only in construction with verbs, while *min mɛɛn... tee* occurs in construction with verbs, nouns, or prepositions. Since *caul* and *ceñ* may occur with *min... tee* as well as *min mɛɛn... tee*, this test suggests that they are verbs.

(1) d. caap haoe min/min mɛɛn caul knong trung tee
bird fly NEG enter inside cage

(2) d. caap haoe min/min mEEn ceñ pii trung tee
bird fly NEG leave from cage

However, in the Thai examples below, this negation test cannot help in identifying the syntactic categories of *khâw* and *ʔɔ̀ɔk*, since neither *khâw* nor *ʔɔ̀ɔk* may be negated in the positions in which they occur. That is, the sentences (3d) and (4d) are ill-formed with the cooccurrence of any negation marker (i.e., *mây*, *mâydây* and *mâychây*). Other tests are needed to clarify this puzzle.

(3) d. *nók bin *mây/mâyday/mâychây* *khâw* pay
bird fly NEG enter go

nay kroŋ
inside cage

(4) d. *nók bin *mây/mâydây/mâychây* *ʔɔ̀ɔk* càak kroŋ
bird fly NEG leave from cage

3.4. Root Predicate with the Choice of Negation Expression

The following sentences illustrate that there exist forms *caul*, *ceñ*, *khâw*, and *ʔɔ̀ɔk* which function as root predicates.

(1) e. caap *caul* knong trung
bird enter inside cage
'The bird entered the cage.'

(2) e. caap *ceñ* pii trung
bird leave from cage
'The bird left the cage.'

(3) e. nók *khâw* pay nay kroŋ
bird enter go inside cage
'The bird entered the cage.'

(4) e. nók *ʔɔ̀ɔk* càak kroŋ
bird leave from cage
'The bird left the cage.'

That these forms may occur as root predicates, as shown in (1e)–(4e), implies the following possibilities: 1) the forms in question in (1a)–(4a), as well as the ones in (1e)–(4e), are verbs, or (2) the forms in question in (1a)–(4a), as well as the ones in (1e)–(4e), are non-verb predicates (noun or preposition predicates), or (3) the forms in question in (1a)–(4a) are not verbs but have homophonous verbal counterparts, the verbs in (1e)–(4e).

By reapplying the choice of negation test, we may be able to identify whether the forms in a root predicate clause (1e)–(4e) are themselves verbs. Consider (1f)–(4f):

(1) f. caap mɨn/mɨn mEEn caul knong trung tee
bird NEG enter inside cage
'The bird did not enter the inside of the cage.'

(2) f. caap mɨn/mɨn mEEn ceñ pii trung tee
bird NEG leave from cage
'The bird did not leave from the cage.'

(3) f. nók *mây/mâydây/*mâychây* khâw pay nay kroŋ
bird NEG enter go inside cage
'The bird did not enter the inside of the cage.'

(4) f. nók *mây/mâydây/*mâychây* ʔɔ̀ɔk càak kroŋ
bird NEG leave from cage
'The bird did not leave the cage.'

As shown above, the forms *ceñ*, *caul*, *khâw*, and *ʔɔ̀ɔk* may occur with the negation words for verbs, namely *mɨn...tee* for Khmer and *mây* and *mâydây* for Thai in a root clause. Hence, these forms in (1e)–(4e) are found to be verbs.

Now it needs to be determined whether these four words are instances of the same word found in our original examples (1a)–(1d) or whether they are lexically

distinct homophones. If they are the same lexical entries, they should have the same meanings and the same selectional restrictions. As mentioned earlier, it is difficult to determine whether forms in two different environments have the same meaning unless the meanings of the forms in the two environments are distinctly different. The selectional restriction is easier to apply. If the forms in question are verbs, their implied subjects will be identical with the Patient of the preceding root verb, and any noun phrase occurring as the Patient of the higher verb should also be able to occur as the subject of our forms when they are used as root clause verbs themselves.

However, if we can find some class of nouns which may occur as the matrix Patient but which may not occur as root clause subjects, the forms which may occur in the root clauses are selectionally and thus lexically distinct from homophonous post-verb forms. By testing the forms in question with transitive verbs with direct objects that are inanimate and not normally perceived as able to move by themselves, if the post-verbal forms may not occur as root verbs of their own clause with the same set of nouns as subjects, then they are selectionally and thus lexically distinct from the homophonous post-verb forms, and the latter cannot be verbs. The only remaining possibility then is that they are adverbs. This is in fact what is shown by the following tests in (7)–(10).

(7) a. koet ruñ <u>tok</u> caul knong banthom
he push table enter inside room
'He pushed a table into the room.'

(8) a. koet ruñ <u>tok</u> ceñ pii banthom
he push table leave from room
'He pushed a table out of the room.'

(9) a. dɛɛŋ lâak <u>tôʔ</u> khâw pay nay hɔ̂ŋ
Daeng drag table enter go inside room
'Daeng dragged a table into the room.'

(10) a. dɛɛŋ lâak <u>tôʔ</u> ʔɔ̀ɔk càak hɔ̂ŋ
Daeng drag table out from room
'Daeng dragged a table out of the room.'

For each example, the underlined word is the Patient of the higher clause. By the lexicase P2a infinitival complement rule, if the following word is an infinitival complement of the higher verb, this word is its implied subject. Consequently, it should be possible to construct another sentence in which the following word is the main verb and the underlined word is its subject. The following examples test this prediction.

(7) b. *tok caul knong banthom
table enter inside room
'A table entered the room.'

(8) b. *tok ceñ pii banthom
table leave from room
'A table left the room.'

(9) b. *tó? khâw pay nay hɔ̂ŋ
table enter go inside room
'A table entered the room.'

(10) b. *tó? ʔɔ̀ɔk càak hɔ̂ŋ
table leave from room
'A table left the room.'

However, these examples in (7b)–(10b) are all semantically anomalous in a way that (7a)–(10a) are not. Consequently, the root verbs do not have the same selectional restrictions as the putative verbs in the post-verbal position. That is, if *khâw* is a verb, we have no explanation for the illformedness of (9)b; therefore, it is not a verb, and can only be an adverb.

In fact, our mystery words all turn out to be adverbs belonging to an independently motivated class of deverbal adverbs (cf. Wilawan, 1994), and their semantic properties when viewed from this perspective are quite consistent with the other members of that adverb class (cf. Indrambarya, 1994): they encode the orientation of the motion of the Patient, taking the speaker as the point of reference. Syntactically, they have no direct connection to a following NP if any (i.e., (1a)–(4a)), which is why a following NP can be topicalized, leaving them stranded unless followed by another adverb, and why they can occur with no following NP at all, depending on the class of the regent verb; e.g.,

(1) g. caap haoe *caul*
bird fly enter
'The bird flew in.'

(2) g. caap haoe *ceñ*
bird fly leave
'The bird flew out.'

(3) g. nók bin *khâw*
bird fly enter
'The bird flew in.'

(4) g. nók bin *ʔɔ̀ɔk*
bird fly leave
'The bird flew out.'

4. CONCLUSION

In conclusion, when the forms *caul* and *ceñ* in Khmer follow the main verb *haoe* 'to fly,' they cannot be interpreted as prepositions or nouns, but only as adverbs or verbs, as suggested by the stranding test and PP topicalization. The choice of negation expressions and the root predicate with the negation word *min... tee* suggests that *caul* and *ceñ* in (1) and (2) could be verbs. However, the root predicate test with inanimate objects tells us that they are rather deverbal homophonous adverbs, and suggests that the negation test be revised to allow *min... tee* to negate directional adverbs, which would result in a more consistent analysis. Further investigation will shed more light on this issue.

The forms *khâw* and *ʔɔ̀ɔk* in the Thai data are more difficult to interpret because the choice of negation-expression test is not applicable to the Thai data as it serves only as a one-way test. Moreover, the application of the stranding test is obscured for the form *khâw* due to the presence of the following adverb *pay*. Nevertheless, the inanimate object variation of the root predicate test tells us that they are lexically distinct from the homophonous root predicates. The only remaining conclusion then is that they are deverbal adverbs.

REFERENCES

Indrambarya, K. (1994). *Subcategorization of verbs in Thai: A lexicase dependency approach.* Unpublished doctoral dissertation. University of Hawai'i, Honolulu.

Indrambarya, K. (1995). Are there prepositions in Thai? In M. Alves (Ed.), *Papers from the third annual meeting of the Southeast Asian Linguistics Society* (pp. 101–117). Arizona State University Program for Southeast Asian Studies.

Panupong, V. (1978). Some basic problems of semantics concerning certain types of homophonic-graphic words in Thai. In W. U. Dressler & W. Meid (Eds), *Proceedings of the Twelfth International Congress of Linguists* (pp. 217–221). Insbrucker Beiträge zur Sprachwissenschaft. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Sak-Humphry, C. (1996). *Khmer nouns and noun phrases: A dependency grammar analysis.* Unpublished doctoral dissertation. University of Hawai'i, Honolulu.

Savetamalaya, S. (1989). *Thai nouns and noun phrases: A lexicase analysis.* Unpublished doctoral dissertation. University of Hawai'i, Honolulu.

Starosta, S. (1988). *The case for lexicase.* London: Pinter Publishers.

Starosta, S. (in press). The identification of word classes in Thai. In K. Tingsabadh & A.S. Abramson (Eds.), *Essays in Tai linguistics.* Bangkok: Chulalongkorn University Press.

Wilawan, S. (1993). *A reanalysis of so-called 'serial verb constructions' in Thai, Khmer, Mandarin Chinese, and Yoruba.* Unpublished doctoral dissertation. University of Hawai'i, Honolulu.

Wilawan, S. (1994). Categorial reanalysis: Deverbal adverbs. *Phaasaa lae phaasaasaat [Journal of Language and Linguistics], 12,* 71–88.