

The Identification of Word Classes in Thai

Stanley Starosta

University of Hawai'i

INTRODUCTION

This article is intended as a contribution to solving the problem of finding a generally acceptable and grammatically justifiable set of syntactic word classes for Thai. It begins with a consideration of the basic question of how syntactic word classes should be established, and then applies the results to several open questions in Thai syntax, contrasting some previous approaches with constrained dependency analyses of the same phenomena in terms of the kinds of conflicting criteria which have been applied in establishing word classes. The two particular areas I will focus on are the questions of (1) which word class to assign "classifiers" to, and (2) whether or not Thai has a distinct class of prepositions.

The grammatical framework within which the following discussion will be framed is an approach to dependency grammar referred to as LEXICASE. Dependency grammar is an ancient approach to grammatical analysis (cf. Covington, 1986) which analyzes sentences in terms of pairwise relations between words, and lexicase is a version of dependency grammar which localizes information about these pairwise links in the lexical entries themselves.

For the purposes of this paper, lexicase dependency grammar can be viewed as having three salient characteristics: it is GENERATIVE (formal and explicit), it is strictly CONSTRAINED, and it is UNIVERSAL. One of the universal constraints it imposes on languages is a strictly limited inventory of word classes: no language (including Thai) may contain a word class which is not identical with, or a subclass of, one of the following eight classes: *V* (verb), *N* (noun), *Adj* (adjective), *Det* (determiner), *Adv* (adverb), *P* (preposition or postposition), *Cnjc* (conjunction), or *Sprt* (sentence particle), though not every language necessarily utilizes all eight classes (cf. Dixon, 1977). This requirement is not some kind of arbitrary edict, but rather an empirical hypothesis: if it is correct, it will make possible greater language-specific and cross-linguistic generalizations, and it can be proven wrong by proposing an alternate set (or no fixed set at all) and showing how this alternative approach makes it possible to capture more and better grammatical generalizations. This is the crucial factor that makes grammars written in different frameworks comparable: all grammars try to capture generalizations, and the grammar that does a better job of capturing generalizations is a better grammar.¹

¹As an example, Savetamalya (1989) analyzes words like Thai *níi* 'this' as a determiner, while Amara Prasithrathsint (p.c.) considers them to be adjectives. Savetamalya showed that the distribution of *níi* is quite different from the distribution of other adnominal modifiers, so that it would at least have to be considered a special subclass of adjectives. Considerations of cross-linguistic generality then tell us that a class of adnominal modifiers which always occur at the periphery of an NP and which express deictic meanings are better analyzed as determiners than as adjectives.

As a brief illustration of how this criterion may be applied, consider the advantages of proposing a fixed inventory of lexical categories for the analysis of auxiliary verbs in English or Thai (cf. Savetamalya, 1987; Indrambarya, 1994): the above eight-term inventory makes no provision for a distinct grammatical category *AUX*, *INFL*, or *I*. It thus excludes the Chomskyan *I(NFL)* analysis, an analysis which reliably results in the loss of numerous syntactic and morphological generalizations, the creation of such otherwise unmotivated transformational rules as “V-movement” and “I-movement,”² and an increase in the abstractness of the associated grammatical representations. That is, the constraint forces the linguist to adopt an analysis in which “auxiliaries” are a subclass of the class of verbs, an analysis which turns out to be demonstrably superior to the Chomskyan alternative in capturing morphological and syntactic generalizations (Starosta, 1991).

WORD CLASSES

In writing a syntactic description of a language, one thing of which we may be sure is that the description will have to refer to word classes. This follows from some simple but fundamental considerations.

The first consideration is that every language must have words. Let us start off by defining a *FREE FORM* as a stretch of speech bounded at both ends by silence. If we follow Leonard Bloomfield (1926, pp. 153–164; 1933 p. 178) in defining a word as a minimal free form, then we can find a set of free forms which are not composed of parts which themselves occur as free forms. Such forms are minimal free forms, that is, *WORDS*.

We know that these words must be identified in a grammar as grouping into classes from the fact that not every sequence of words is equally acceptable to native speakers of a language. Speakers are fairly consistent in identifying certain sequences as belonging to their language and others as not belonging to their language, and from the point of view of generative grammar, the content of a syntactic description is the internalized knowledge that allows speakers to do this in a fairly consistent and intersubjectively verifiable way. Logically, there are two ways in which that knowledge could be internally represented, an extensional way and an intensional way. The speaker might be assumed to have an internalized list of possible strings of words, which would be an extensional representation, or he might have access to a set of “well-formedness conditions” that a string must meet in order to qualify as belonging to his language. This would be an intensional mechanism. Because the set of strings is potentially infinite, we can rule out the first alternative, and focus our attention on the second.

²Claims for the necessity of V/I movement made within the Chomskyan framework, as presented for example in Pollock (1989), are circular because they assume as given that there must be a category *I* to begin with. When this category is eliminated, the generalizations captured by his analysis can be reformulated much more naturally in terms of the dependency grammar notion of *FINITE VERB DOMAIN* (cf. Savetamalya, 1987). Note that I am not denying the existence of the set of facts that V/I movement is intended to account for, but only the necessity of using powerful transformational rules to account for them.

We have just in effect defined syntax as the study of which words can co-occur in phrases, that is, non-minimal free forms. Once again, we can consider two alternative ways in which syntactic information might be encoded: we might hypothesize that the necessary well-formedness conditions on strings were encoded as a list of specific pairs of words which are allowed to occur with each other. Once again, however, such a list would be a sizable subset of the Cartesian product of the entire vocabulary of a language, a list which seems much too large to be plausibly memorized. Moreover, even if it were memorized, where would it come from? It could hardly have been memorized from experience, since speakers can consistently judge collocations as good or bad even for pairs that are very unlikely to have ever been produced before.

Again, we are faced with the necessity of positing a more abstract ability than simple listing to account for syntactic competence. I can find no alternative to assuming that each word is identified as belonging to one of a fairly small set of classes, and that the well-formedness conditions which are the content of a syntactic description are stated in terms of these classes rather than directly in terms of individual lexical entries. All we need to do then is to group words with identical properties into sets and write our grammatical rules to refer to these sets, and the problem is solved. This paper will consider some strategies for accomplishing this, and draw implications for the resolution of some controversial questions in the analysis of Thai "classifiers" and prepositions.

CONSTRAINTS VERSUS FLEXIBILITY

Throughout this paper, I will be assuming that it is desirable to find a rigorous and consistent set of syntactic criteria for uniquely determining the syntactic "part of speech" of each Thai word, and that it is desirable to fit the parts of speech for Thai into a limited universal set. However, neither of these goals is universally accepted. In the first section of this paper, I would like to consider a proposal made by Eric Schiller which rejects both of these assumptions. I will begin by quoting the first page of his paper, "Parts of speech in Southeast Asian languages: An autolexical view" (Schiller, 1992, p. 777):

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.

I will consider three general points raised by this passage, 1) the question of flexibility versus rigid categories, 2) the question of universals versus language-specific analyses, and 3) the question of polysemy versus homophony.

Flexibility versus Rigidity

Schiller's paper can be seen as a plea for flexibility in the assignment of syntactic categories. Flexibility, however, is the sworn enemy of science. The content of a theory is its constraints, and a "theory" which allows everything, as autolexical grammar does, tells us nothing. To be testable, a theory must be constrained and rigid. If new data cannot be accommodated by such a theory, then the theory has been falsified and must be corrected. If new data are found to fit into the cast iron pigeonholes provided without any alteration being necessary, then the theory has been confirmed, though of course no scientific theory can ever be proven absolutely correct.

Universals versus Language-Specific Analyses

There are at least two questions we might ask in connection with the roles of universals in language analysis: *should* languages be analyzed in terms of universal categories, and *can* languages be analyzed in terms of universal categories? For the structuralists, of course, the answer to the first question was negative; it was felt that to try to fit observations into preexisting categories was circular and compromised the objectivity of the analysis. However, it was a genuine contribution of Chomskyan linguistics to point out that if linguistics was to be a science, it had to be a search for generalizations, for simple explanations that covered a broad range of data, and that a true scientific theory should cover all the phenomena in the domain with a minimum amount of explanatory apparatus.

From these considerations, it follows that a science of language should strive to cover all the phenomena of human language with a single coherent set of principles. That is, a science of linguistics should be a search for universals. That raises the second question: *can* languages be analyzed in terms of universal categories? If not, we have to give up an attempt to construct a science of language as the term "science" is generally understood. With the stakes so high, it is obvious that this course should not be taken without first making a good-faith effort to accomplish this goal. I think Schiller has given up too soon.

It is of course difficult, outside of mathematics, to prove that something is impossible. A claim such as Schiller's that it is impossible to find a universal set of syntactic categories and/or a universal set of criteria to identify such a set will only stand as long as no one has actually produced a universal set of categories and/or promulgated such a set of criteria. However, (1) the lexibase dependency grammar framework *has* proposed a universal set of eight categories (Starosta, 1988, pp. 51–52), and tested them in analyses of parts of more than 70 different languages,

mostly non-Indo-European ones, and (2) this paper is an attempt to illustrate a universally applicable set of principles for identifying word classes.

The gist of my proposal is very simple, and hardly new: syntactic criteria (that is, distribution and co-occurrence) should be used to determine membership in syntactic word classes. Schiller refers several times to the difficulty of determining word classes in Southeast Asian languages, and the controversialness of such categories as “coverbs,” “classifiers,” and “relator nouns.” This perceived difficulty and controversialness is a direct result of his confusion about what criteria to use in determining syntactic word classes. When syntactic criteria are adopted and applied consistently, the results are gratifyingly clear and consistent: coverbs are prepositions (Clark, 1978; Starosta, 1985) and relator nouns and classifiers are nouns (Starosta, 1985; Savetamalya, 1989). The difficulty and controversialness are in the eye of the beholder.

Schiller regards the attempt to establish a rigid set of universal word classes as “a perspective based on the study of European languages.” However, any prospective universal theory has to start somewhere, and European languages are as good or as bad a place to start as any, as long as you don’t end up staying there. The real danger is that with a powerful, vague, and flexible “theory” such as the various Chomskyan approaches or autolexical grammar, it is all too easy to take, say, a Chomskyan analysis of English syntax as the “underlying” structure for all languages, and/or to impose functionally conceived parts of speech like “classifiers” on a language too polite to object. Working within an explicit and rigid monostratal theory, however, makes this impossible. The language itself gets the final say as to whether it is willing to conform to the categories provided by the universal theory. If it turns out that applying these categories in the analysis of Language L_i makes it possible to capture language-internal as well as universal generalizations, then the universal theory is vindicated, but if not, it is the universal theory that has to back down, and it is our shared knowledge of the nature of human language as a whole that reaps the rewards.

Polysemy versus Homophony

Eric Schiller’s whole article is an attempt to justify an analysis that allows words to belong to more than one category at a time, or perhaps “archicategories” would be a better term to use. The only reason that Schiller could maintain such a position is that his proposal is only programmatic, and has not been formalized within a generative framework. In this section, I will try to demonstrate the loss of generalizations that would necessarily result by setting up a different word class for every distinct range of environments in which a single root can occur.

In order to classify words, we will first need to refine our definition of a word. The definition of a word as a minimal free form is a definition in terms of shape, either phonological or orthographic. However, if we assume that shape alone is the sole criterion for identifying a word, we will be led into paradoxes and excessive complexity. As a first example, consider the following sentence:

- 1) John looked toward the bank.

This sentence is ambiguous, referring either to looking toward the raised edge of a natural body of water, or toward a place where money is kept and financial transactions take place. By a process of substitution, we can narrow down the locus of the ambiguity to the form *bank*. A grammar is normally expected to accommodate such differences in meaning, and the normal solution to this particular phenomenon is to postulate homophones: there are two words, *bank*₁ and *bank*₂, with a common form but different meanings.

Since we seem to be justified in stating that, as a rule, words with different meanings are different words, it follows that semantic difference can be used as the basis for differentiating words. (Panupong, 1978, p. 217)

There are, however, cases of words that are identical both in writing and pronunciation but different in meaning. On the semantic difference basis, one can also regard these cases as different words, which may be called homophonic-homographic words. These words fall into three types: 1) those with different functions..., 2) those with the same function..., and 3) those whose functions are different but whose meanings are related. (Panupong, 1978, p. 218)

This brings us to the classical Saussurean sign: a sign is an association of a form with a meaning, neither being predictable from the other, and *bank*₁ and *bank*₂ are HOMONYMS, different signs because although they are identical in form, they are different in meaning. Similar considerations apply to SYNONYMS, a single meaning matching two different forms. It is uncontroversial to treat such patterns as constituting separate lexical entries.

Conventional lexicography and conventional linguistic practice make a distinction between homonymy and POLYSEMY, in which a single form is treated as a single lexical entry with multiple meanings. Yet how do we draw the line objectively between these two situations? For example, is *cuff* 'a blow with the fist or an open hand' the same word as *cuff* 'a fold of cloth at the sleeve'? If you give someone a cuff, couldn't it be like giving that person the back of your hand? The dictionary says these are homonyms, apparently because they have different etymologies, but a native speaker on the street can hardly be expected to know this, and might decide they are different "senses" of a single word based on their spelling and perceived overlap in meaning.

From an objective synchronic linguistic point of view, there is no justification for distinguishing homonymy and polysemy: the different "senses" of a single form are not consistently predictable from each other³ or from their shared form, so there is no way to avoid giving them separate status. The lexicographic practice of combining them under a single "lemma" is just a compromise between synchronic linguistic reality, historical etymology, and the folk perception that a form is the same thing as a word.

The same reasoning that was used to define homophones and synonyms can also be applied to the establishment of syntactic word classes. The reasoning behind this approach was outlined by Panupong (1978, p. 218):

³Thus to cite the classical example, what principle allows us to predict the 'unmarried man' meaning of the form *bachelor* from the 'unmated young male fur seal' meaning or vice versa, or to determine that the sense of *bachelor* referring to an animal must refer to a fur seal rather than to a walrus or a sea elephant or a wildebeest?

If we were to consider [homophonous forms with different distributions] as one and the same word, i.e. a polyfunctional word, it would mean that we regard their meanings as same, but hold that a word can have more than one function. In this case, one faces two problems:

- a) How can one label the word? For mnemonic convenience one may use all the functions as the label... Such labels might serve our purpose, but they are too clumsy to be altogether suitable.
- b) One cannot decide immediately to what class a word in a particular sentence belongs...
- c) There is another drawback: to have polyfunctional words is to create more classes of words.

From a theoretical point of view, this latter point of view is especially telling. Suppose for example that we take the five sequences of letters *jog*, *walk*, *shout*, *creep*, and *monster* as five words, and try to write a grammar that accounts for the following data:

- 2) a) John will *walk* across the freeway.
b) John takes a scary *walk* on the freeway every morning.
- 3) a) John will *jog* across the freeway.
b) John takes a scary *jog* on the freeway every morning.
- 4) a) John will *shout* across the freeway.
b) *John takes a scary *shout* on the freeway every morning.
- 5) a) *John will *monster* across the freeway.
b) John takes a scary *monster* on the freeway every morning.
- 6) a) John will *creep* across the freeway.
b) John takes a scary *creep* on the freeway every morning.

We have already seen that a particular form cannot be assumed to be a single lexical entry if it is associated with two distinct meanings. These data illustrate the further point that we also cannot necessarily equate a single Saussurean sign, a form-meaning package, with a lexical entry, for syntactic reasons: it would complicate the grammar and lose obvious generalizations. Thus supposing we tried to set up a Saussurean sign-based lexicon as part of a grammatical account of the examples above, providing for each entry its pronunciation (e.g., *walk*) and its meaning (e.g., 'ambulate'), plus its syntactic environment (e.g., the frame [Modal___PP], [*take* Det Adj___PP], or both). The result would include something like the following five items:

- | | | |
|----|----------------|---|
| 7) | <i>walk</i> | ['ambulate'; [Modal___PP], [<i>take</i> Det Adj___PP]] |
| | <i>jog</i> | ['trot'; [Modal___PP], [<i>take</i> Det Adj___PP]] |
| | <i>shout</i> | ['raise the voice'; [Modal___PP]] |
| | <i>monster</i> | ['freak'; [<i>take</i> Det Adj___PP]] |
| | <i>creep</i> | ['crawl'; [Modal___PP]; 'boring person'; [<i>take</i> Det Adj___PP]] |

The practice in folk linguistics, traditional and Chomskyan grammar, and popular lexicography is to assume that homophonous words in the same syntactic class which are felt to be similar in meaning to each other are in some sense "the same word." (In traditional and modern German linguistic practice, this is referred to as *die Einheit des Wortes* 'the unity of the word'.) If we were to assume, in accordance with this practice, that we are dealing with five and only five different lexical items in the list above, and if we tried to write a grammar that treats words with the same range of possible syntactic environments as belonging to the same syntactic class, and words

with different ranges of syntactic environments as belonging to different syntactic classes (see for example Warotamasikkhadit, 1972; Thepkanjana, 1992), then we would have to posit three different syntactic classes, C_1 , C_2 , and C_3 , for these five entries:

- 8) *walk* [+ C_1 ; 'ambulate'; Modal__PP, *take* Det Adj__PP]
jog [+ C_1 ; 'trot'; Modal__PP, *take* Det Adj__PP]
creep $\left[\begin{array}{l} +C_1 \\ \text{'crawl'; Modal__PP} \\ \text{'boring person'; take Det Adj__PP} \end{array} \right]$
shout [+ C_2 ; 'raise the voice'; Modal__PP]
monster [+ C_3 ; 'freak'; *take* Det Adj__PP]

Thus *walk*, *jog*, and *creep* would all be placed in the same syntactic class, C_1 , because they all occur in the same range of environments.

A grammar which was stated in terms of such classes, though, would be awkward and un insightful. This is illustrated by the following schematic rules, modeled on the early transformational approach to lexical classification exemplified (though inconsistently) by Warotamasikkhadit (1972):

- 9) $C \rightarrow \{C_1, C_2\} / \text{Modal_PP}$
 $C \rightarrow \{C_1, C_3\} / \text{take Det Adj_PP}$

These rules are un insightful because they are stated in terms of disjunctions, and offer no explanation as to why the class C_1 should appear in two such different environments. Moreover, the meaning difference that goes with *creep* in the two different environments is, as usual, ignored.

Suppose however that we were to decide that like an unpredictable difference in meaning, an unpredictable difference in distribution also was necessary and sufficient grounds for setting up a distinct lexical item, and that two forms that occurred in different frames ("grammatically significant environments" or "gse"; cf. Indrambarya, 1994, section 3.3.3) were homophonous but distinct lexical entries. Under that assumption, a much neater grammatical picture would emerge. Thus given the following set of lexical items, established in accordance with this principle:

- 10) *walk*₁ [+V; 'ambulate'; [Modal__PP]]
*jog*₁ [+V; 'trot'; [Modal__PP]]
shout [+V; 'raise the voice'; [Modal__PP]]
*creep*₁ [+V; 'crawl'; [Modal__PP]]

*walk*₂ [+N; 'ambulation'; [*take* Det Adj__PP]]
*jog*₂ [+N; 'rot'; [*take* Det Adj__PP]]
monster [+N; 'freak'; [*take* Det Adj__PP]]
*creep*₂ [+N; 'boring person'; [*take* Det Adj__PP]]

the schematic rules needed to account for their distributions are quite simple and straightforward:

- 11) $C \rightarrow V / \text{Modal} __\text{PP}$
 $C \rightarrow N / \text{take Det Adj} __\text{PP}$

This approach at the same time also accounts for the different meanings of *creep* associated with the different distributions: *creep*₁ is a verb, with a prototypical verbal meaning ('action or a state of being' in traditional grammar) paired with a verbal distribution, while *creep*₂ is a noun, with a prototypical nominal meaning ('name of a person, place, or thing') paired with a nominal distribution.⁴

The advantage of this homophonous entry approach is summarized by Panupong (1978, p. 218):

However, there are also some merits in having such a category:

- 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 labeling the words with more than one function.

Similar considerations apply, *mutatis mutandis*, within a single class of words. For example, if we are dividing verbs into transitive and intransitive verbs, what do we do with a form like *eat*, which can occur with or without a direct object, versus *dine*, which never takes a direct object and *devour*, which always does? Do we have to set up a third class of "ambivalent verbs" for such cases, as Udom Warotamasikkhadit (1972, pp. 15–16) did for Thai in his dissertation, or "trans-intransitive verbs," as proposed by Kingkarn Thepkanjana (1992, pp. 310, 314–321)? Vichin Panupong (1978, p. 123) answers in the negative:

It has to be pointed out that, for the purpose of the present study it has not been found necessary to set up further classes of verbs to account for words which may occupy both place 2 and place 3 in either the first or the second pair of testing sentence frames described above, on the one hand, and those which may occupy such places in either the second or the third pair, on the other. Instead, they will be regarded as homophones of one kind or another according to which two of the above pairs of testing sentence frames they can operate in. We have thus intransitive/transitive homophones in the one case and transitive/double transitive in the other....

This is also the principle I will adopt and try to motivate from basic principles in this paper. The cost of such a revision, as with the case of synonyms and semantically determined homophones, is an increase in the number of distinct lexical

⁴For those speakers who can use *creep* as a noun meaning 'act of creeping,' a third homophonous *creep*₃ would be required in the lexicon:

*creep*₃ [+N; 'act of creeping' take Det Adj ___PP]

This would not alter the number of syntactic classes required in the grammar nor the generality of the rules thereby achieved.

entries, and the gain is a reduction in the number of possible syntactic word classes and the capture of important grammatical generalizations.⁵ When a grammar attempts to cover a significant range of data, the gains can be dramatic, since the rules will consist of a list of real generalizations rather than the jumbles of subscripted symbols found in Warotamasikkhadit (1972) and its modern notational variants. As a schematic example, consider three of the six syntactic frames used by Panupong in distinguishing between intransitive, transitive, and double transitive verbs (Panupong, 1970, pp. 120–122; 1a is slightly modified):

- | | | | |
|-----|-----|---------------|-------------------------|
| 12) | 1a. | n 2 l'é:w | Intransitive verbs |
| | 2a. | n 2 n l'é:w | Transitive verbs |
| | 3a. | n 2 n n l'é:w | Double transitive verbs |

If we consider each frame to define a single word class, then we would have three distinct word classes, with some roots occurring in more than one class. If on the other hand we were to define a verb class in terms of the aggregate set of frames in which a given root can occur, then the number of classes increases exponentially. More precisely, the maximum number of classes C definable in this way for a given number of frames f is expressed by the formula $C = 2^f - 1$. Thus assuming only two distinct frames, there would be three possible word classes, but with three possible frames, there could be seven distinct verb classes:

13)		1a	2a	14)		1a	2a	3a
	V_1		+		V_1			+
	V_2	+			V_2		+	
	V_3	+	+		V_3	+		
					V_4		+	+
					V_5	+	+	
					V_6	+		+
					V_7	+	+	+

On the contrary, if we let each distinct frame define a separate lexical class, as illustrated in item 10), then the number of classes equals the number of frames, and the number of classes rises arithmetically instead of exponentially with the identification of new frames.

At this point we have reached a third approximation of the concept "word": a word is not a form, nor a form plus an associated meaning, but a form plus meaning plus distribution. This may be referred to as a "triune sign," a tripartite cluster of sound, meaning, and distribution. This concept is different from the conventional view in that forms with different distributions are conventionally considered different

⁵There is independent evidence supporting the homophonous entry analysis which involves language acquisition and inter-speaker lexical differences (cf. Starosta, 1988). However, to go into this would take us beyond the scope of this paper.

“versions” of the same word (e.g., transitive and intransitive uses of *drink* are considered the same word), while in the triune sign conception, these forms must be regarded as two different and independent words. This distinction will be crucial in the following discussion, where it will be used not only to motivate assigning forms to distinct syntactic subclasses (“word classes,” “parts of speech”), as in the case of verbs versus prepositions, but also in motivating homophonous entries within the verb and noun classes.

It is necessary to emphasize this point here because although all full syntactic theories incorporate information on subcategorization in their lexical entries, in practice the distribution criterion is often overridden to conform to the folk taxonomy of homophonous forms referred to under the label “unity of the word” above. The most prevalent example of this is the assignment of homophonous transitive and intransitive verbs, such as English *eat*₁ ‘ingest X’ and *eat*₂ ‘perform the act of ingesting,’ to a single common lexical entry with an optional direct object specification. If we can’t resolve this conflict between linguistic principles and folk practices, syntacticians will never agree on their classifications, and thus never produce comparable sets of rules.

CLASSIFIERS AS NOUNS

Distribution

Linguists are also products of their culture and are often affected by the education they receive as part of that culture. What this often means is that they are excessively influenced by orthography, homophony, traditional notional grammar, and/or translation, and these criteria are often inconsistent with modern linguistic considerations. This is pointed out in connection with the classification of nouns in Thai by Udom Warotamasikkhadit:

I believe that any word which can be modified by one of the determiners, *nīi* ‘this’, *nán* ‘that,’ or *nóon* ‘that yonder’ is called a noun. Thus a classifier is a noun under the above definition as also in Uppakitsinlapasan (1953:21–28). I consider *khǎw* ‘thing,’ *hǎw* ‘place,’ *thīi* ‘place,’ *nay* ‘inside,’ *nǎw* ‘outside’...are nouns according to the above definition. (Warotamasikkhadit, 1992, p. 72)

I consider *sǎa khǎw khun* ‘your shirt’... a noun phrase with a string of three nouns, where *khǎw* ‘thing’ is a noun according to the above definition. We ignore the translation of *khǎw* ‘thing’ as ‘of’ in English. Prasithrathsint (1985, pp. 94–95) also considers *khǎw* ‘thing (of)’ as a noun. (Warotamasikkhadit, 1992, p. 73)

Warotamasikkhadit has given us a clear example of the use of a linguistic criterion,⁶ distribution, to correct earlier erroneous analyses based, in this case, on translation; since *nay* in certain contexts can be translated as ‘in,’ and since English *in* can be clearly determined to be a preposition, it seems that some grammarians have used that fact to extend the same analysis to Thai *nay*. On the basis of Thai-

⁶Though he usually doesn’t practice what he preaches; thus he tells us that *hǎw* ‘place’ in his (30) is a noun in the noun string *lôok hǎw santìphâap* ‘world of peace,’ but makes no attempt to linguistically justify the statement that this phrase is in fact a string of nouns.

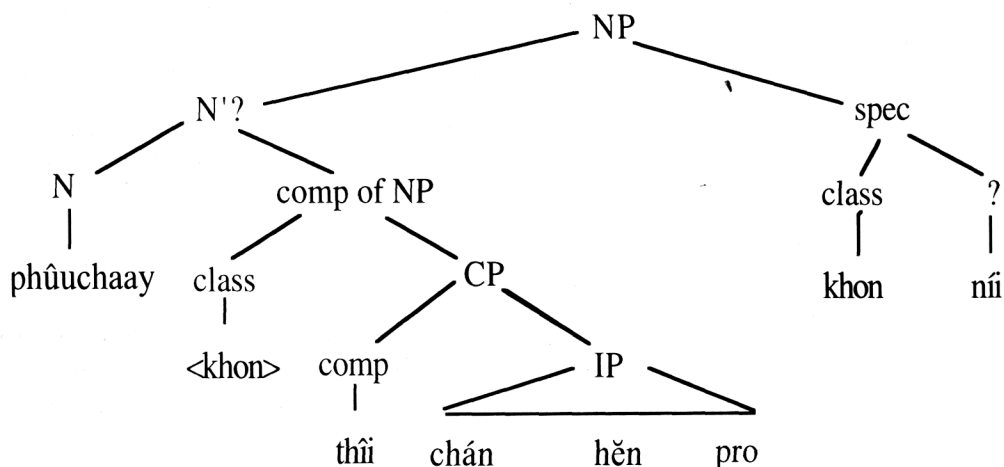
internal distributional evidence, however, *naɯ* has the distribution of a noun, and in a maximally simple and general grammar of Thai, that is how it must be categorized.

Abstract Representations and Transformations

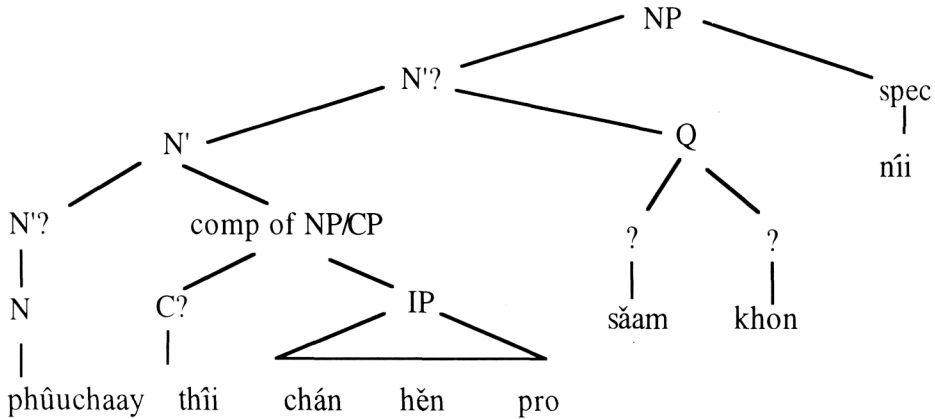
Unfortunately, the step forward taken by linguistics in raising syntactic criteria to preeminence in answering syntactic questions has been partly nullified by a subsequent development: the advent of abstract Chomskyan deep structure, transformations, empty categories, and their various equivalents in other theories influenced by Chomsky's ideas. As long as we are confined to one level of representation, the facts of the language impose a salutary discipline on our analyses: whatever we do, we must make generalizations that apply to the actual words of the language, in the actual orders in which they appear. However, as soon as that direct connection between language data and analyses is broken, a given sentence may be assigned any number of different analyses, and there will be no objective way to choose among them.

This point is illustrated for example in F. K. Lehman's Chomskyan analysis of Thai classifiers (Lehman, 1990). Lehman seems to be influenced by the functions of "classifiers" and by the term "classifier" itself, which was coined in accordance with that perceived function: a "classifier" is viewed as a word that is used to classify nouns according to shape, etc. Thus it is common in a grammatical analyses of Thai to find a class of words called "classifiers," which co-occur with nouns but which are not themselves nouns. This is illustrated by the GB-style tree diagrams 15), 16), and 17) below, which I extracted partly by guesswork from sketchy linear bracketings in Lehman's paper (Lehman, 1990, p. 114):

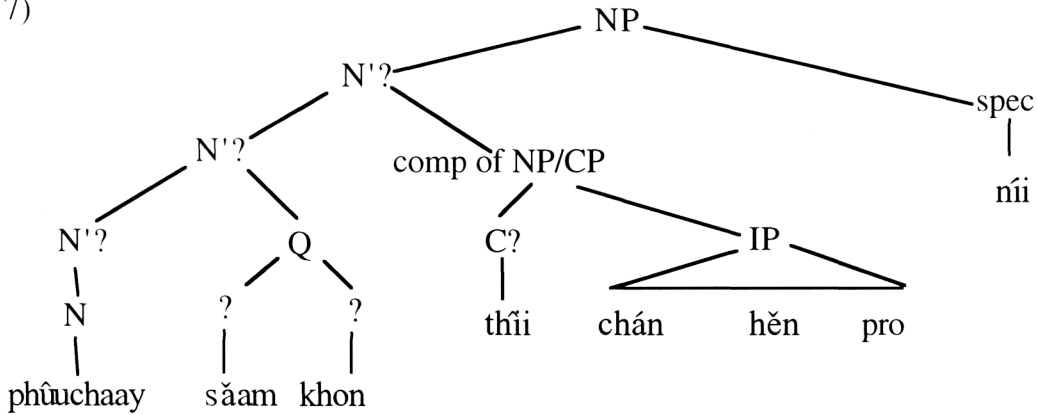
15)



16)



17)



Constrained Dependency Analyses

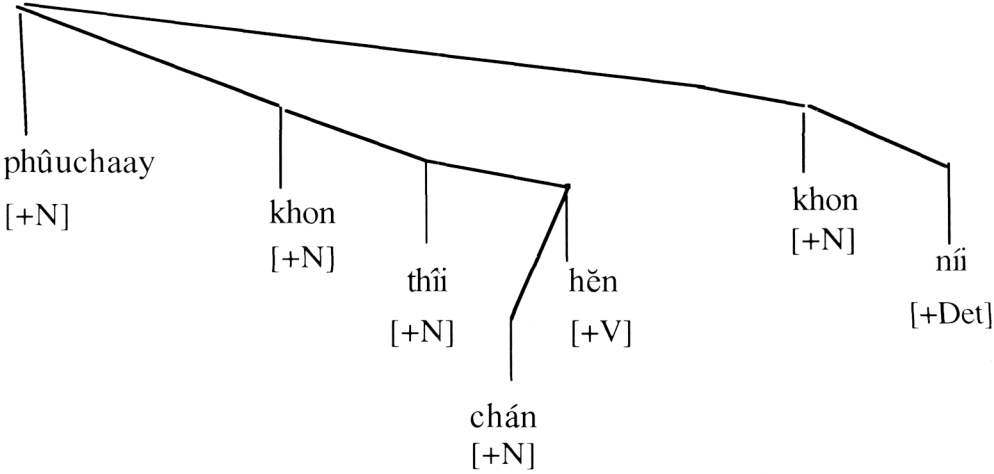
These representations may be compared with dependency analyses of the same NPs (cf. Savetamalya, 1989, as shown in 15'), 16'), and 17') below.

Classifiers as Nouns, or as Something Else?

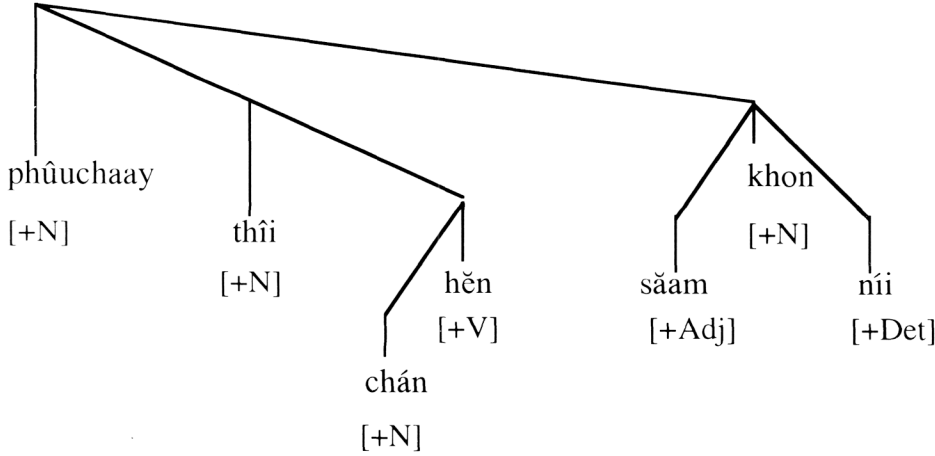
Lehman's Chomskyan representations are complex in terms of the number of different kinds of categories they require, especially when compared to the corresponding dependency analyses given in 15'), 16'), and 17'), which are based on Savetamalya's work on Thai noun phrases (Savetamalya, 1989). Even worse is the kind of formal (though never formalized) machinery it would take to make this kind

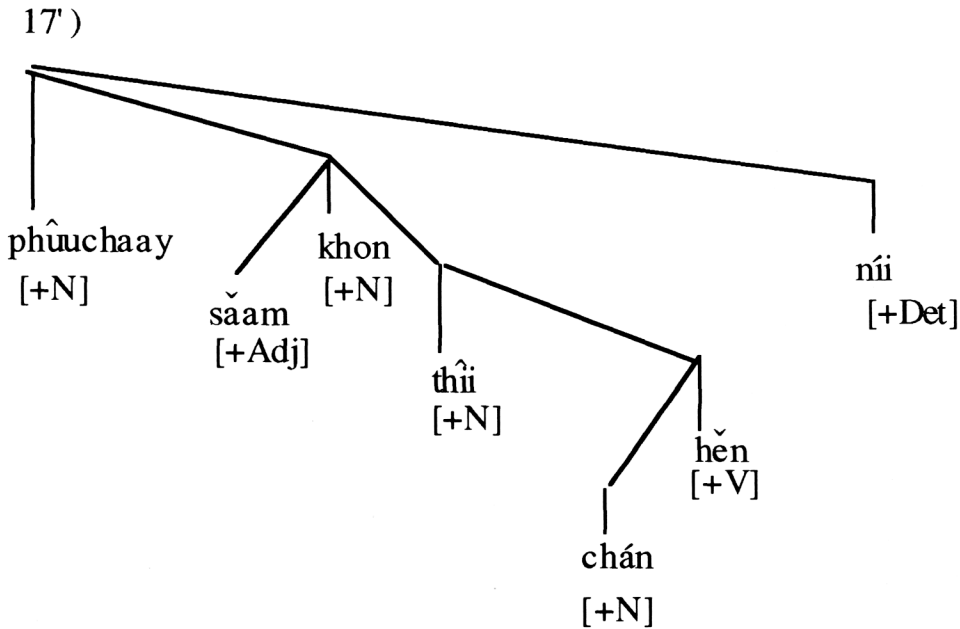
of analysis work: “feature percolation” up and down the tree (Savetamalya, 1989, p. 101), two bar levels, unfilled nodes, “minor categories” like *Quant* (p. 109), “adjunction” (cloned nodes), empty pro-elements combining with head-features to form an overt “classifier” (p. 110), and N-level and N’-level appositive adjuncts (p. 114).

15')



16')





The main descriptive problem with this analysis, in addition to the proliferation of under-motivated and excessively powerful theoretical machinery, is that there are phrases which function as NPs, and which thus by the laws of dependency grammar and X-bar theory must have an N^0 head, that is, a noun, yet they are composed of only a “classifier” and a determiner, number, or adjective, e.g. (example sentence taken from Savetamalya, 1989, pp. 154-155, 256; my bracketing):

- 18)
- | | | | | | | |
|-----|------|-------|------|-----------|------|------|
| [N' | [CLS | lêm] | [Det | níi]] | phœŋ | mâak |
| | clsf | | this | expensive | very | |
- ‘This (book) is very expensive.’

So, where is the head noun? Warotamasikkhadit applies modern syntactic distributional criteria to this problem, and draws the only possible conclusion: the “classifier” itself must be a noun. This analysis may be represented by the bracketing in 19):

- 19)
- | | | | | | |
|--------|-------|------|---------|-----------|------|
| [NP [N | lêm] | [Det | níi]] | phææŋ | mâak |
| | clsf | | this | expensive | very |
- ‘This volume is very expensive.’

Extrapolating from this result, we must in fact conclude that the most economical grammar of Thai treats *all* classifiers as nouns (Savetamalya, 1989, section 4.4). Confronting the same data, however, Lehman draws a quite different conclusion: yes, a Noun Phrase must be headed by a noun, but in the Chomskyan framework, a noun need not have a pronunciation (Savetamalya, 1989, p. 114). Thus 20) could and would have to be analyzed as having an invisible and inaudible head noun *e* modified by a visible and audible classifier,⁷ perhaps something like 20).

20)

[_{NP} [_N e] [_{CLSF} lêm] [_{Det} ní]] phææŋ mâak
 Nclsf this expensive very
 ‘This book is very expensive’; lit. ‘This volume of *e* is very expensive.’

So, Lehman is content; he can simultaneously keep his classifier as a distinct word class, follow the letter of the X-bar law, and could in principle still map his analysis (or anyone else's) onto the actual strings of the language, thereby meeting the requirements of a generative grammar in the original sense.⁸ The broader implication of such an analysis, though, is that a linguistic analysis is liberated from the earthly bonds imposed by the data of the language, and is free to assume any form the investigator fancies (usually one that turns out to look suspiciously like a Chomskyan analysis of English, though in the case of classifiers that happens not to be the case).

Thus we have come full circle: having been freed by the adoption of language-specific distributional criteria from the overweening influence of foreign grammars and translations into foreign languages, we find we have been drawn back into a scientific limbo by the increased power and abstractness and loss of explicitness brought in with the Chomskyan paradigm. The moral to be drawn from this is I think clear: Lehman's adoption of Chomskyan empty words results in an equally empty and unfalsifiable analysis, and Warotamasikkhadit was right in adducing language-specific criteria to conclude that classifiers and words like *bon* 'top' and *khǝɔŋ* 'of' are nouns. Cross-linguistic considerations of determiner distribution support this language-specific analysis.

PREPOSITIONS

The second application of the principles outlined above will be the question of whether Thai has prepositions, and this time I will cite Udom Warotamasikkhadit on the other side of the issue. According to him in "There are no prepositions in Thai" (Warotamasikkhadit, 1992), Thai has no preposition class, but lexicase dependency analyses by Pranee Kullavanijaya (Kullavanijaya, 1974), Amara Prasithrathsint (Prasithrathsint, 1985), Saranya Savetamalya (Savetamalya, 1989), Kitima

⁷For convenience, we may refer to this as the “empty head analysis.”

⁸In practice this is never done. With a theory of this power and unfalsifiability, there would be no point in doing so: given any analysis, it will be possible to map it onto any data, so why bother actually doing it?

Indrambarya (Indrambarya, 1990, 1993a) and Supriya Wilawan (Wilawan, 1993) maintain that it does. Since there is no significant difference among these six investigators regarding the grammaticality of the crucial examples, the differences in the resulting conclusions must be traceable to differences in the language-specific or theoretical criteria applied in allocating words to categories. The six lexicase studies follow the principles laid out in the first part of this paper: (i) if two forms differ in meaning *or* distribution, then they are distinct lexical items, (ii) if two lexical items occur in the same grammatically significant environment, then they are members of the same syntactic (sub)class, and (iii) the resulting analyses must be generative, constrained, and consistent with a universal theory.

The result of this research was substantial agreement about the fundamental question of the correct inventory of word classes for Thai. Lexicase dependency grammar defines a preposition as a word that forms an EXOCENTRIC construction with a phrase, where an exocentric construction is defined as one in which the dependent phrase is an obligatory part of the construction. All six lexicase analyses found that there are two distinct classes of words in Thai that differ in distribution, and that in terms of the universal criteria imposed by the theory, these two classes must be identified as verbs and prepositions respectively.

By contrast, Warotamasikkhadit's claims to the contrary do not satisfy any of the above requirements. It is instructive to compare the approaches and results of the three lexicase analyses with those employed in his paper (1992), in order to see how differing criteria can result in radically different conclusions. For convenience, Warotamasikkhadit's proposals and criteria can be grouped into the following categories. (Page numbers refer to his paper; italics added unless otherwise noted.)

“Unity of the Word”

Same Form, Different Meaning \Rightarrow Same Word Class

Non-generative approaches to grammar, including notional grammar and Chomskyan “generative” grammar, often confuse syntactic categories with grammatical function. This tradition goes back a long way, as reflected in a quotation from Warotamasikkhadit's preposition paper:

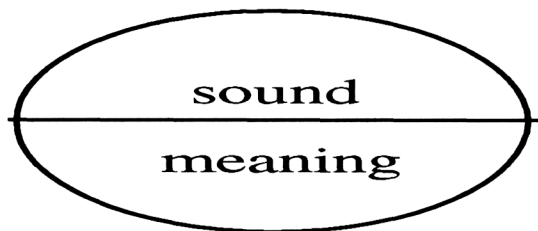
...the class of prepositions hardly has a real existence [in Sanskrit], but is represented by certain adverbial words which are to a greater or less extent used prepositionally. (Whitney, 1889, p. 403, cited in Warotamasikkhadit, 1992, p. 70)

But what does it mean for a word to be “adverbial” but to be “used prepositionally”? W. D. Whitney was working in a tradition in which word classes were defined “notionally,” that is, semantically, which typically meant in terms of their translation into English or some other European language. Thus “adverbial” here probably means something like “answers the question *how, when, where, or how much,*” whereas “used as” seems to refer to syntactic distribution. In modern linguistics, starting with the structuralists and followed with some lapses by transformational grammarians, syntactic classes are determined by distribution, and something which “is used as” a preposition *is* a preposition by definition in such constructions. Within

a modern and explicit framework, the statement “adverbial words which are...used prepositionally” is logically incoherent.

Same Form, Same Meaning, \Rightarrow Same Word

“Same meaning, same form \Rightarrow same word” is the assumption incorporated in the Saussurean sign:



I showed earlier that it results in the loss of important syntactic generalizations. It is also subject to abuse, in that semantics is very subjective, and anyone can claim that any two vaguely similar words “mean the same” if they are pronounced the same. This intuitive semantic approach is exemplified repeatedly in Warotamasikkhadit’s paper. A good illustration of this point is his analysis of the examples given as 21) below. (His original example numbers are given at the right.)

21)

- | | |
|--|-----|
| a. khăw taam hăa lûuk | (1) |
| he follow search child | |
| ‘He was searching for his child.’ | |
| b. khăw dœn taam hăa lûuk | (2) |
| he walk follow search child | |
| ‘He was walking in search of his child.’ | |
| c. khăw dœn taam lûuk | (3) |
| he walk follow child | |
| ‘He walked, following his child.’ | |
| d. khăw dœn taam thănŏn | (4) |
| he walk follow street | |
| ‘He walks along the street.’ | |
| e. dèk dèk nâŋ taam raaw sàphaan | (5) |
| child sit follow rail bridge | |
| ‘Children sit on the bridge rail.’ | |

According to Warotamasikkhadit (1992, p. 70),

Action verbs like *taam* ‘to follow’,...are all verbs in their occurrence, but their translations into English are often prepositions....I consider them verbs because *they* can be negated *when*

occurring as main verbs signifying propositional semantic interpretation or when occurring as the last member of a verb string signifying the same semantic interpretation as the main verb.

In the examples in 21), *taam* is consistently assigned 'follow' as a word gloss, but note that it has quite different English translations: 'following' in (c), 'along' in (d), 'on' in (e), and nothing at all in (a) and (b). If *taam* has the same meaning in all five sentences, why does it have this many different translations? Of course translation cannot be used to justify a syntactic analysis, but it does perhaps bear some weight in evaluating semantic claims. In such a situation, how can a linguist justify saying that the meaning is the same, unless he has already decided that they are the same on the basis of the form alone, and without reference to any semantic evidence?

Warotamasikkkhadit's analysis here mixes distributional with semantic criteria. A verb is first defined as a word that can be negated, which in itself is a respectable distributional criterion, but then any word that has the same shape and arguably the same meaning is assumed to be a verb by association,⁹ regardless of which environment it occurs in. Aside from the fact that the first part of the statement is descriptively incorrect, since adverbs and predicate nouns and prepositions can also be negated in Thai,¹⁰ this procedure necessarily leads to paradoxes and the loss of generalizations in the same way that it does in the example at the beginning of the paper in which the English word *walk* is assigned to a single word class in all its appearances. Thus words that have the distribution of verbs can indeed be negated by *mây*, but homophonous words that occur in non-predicate preposition environments cannot be negated at all (Indrambarya, 1993a). If we assume that the latter forms are prepositions rather than verbs, this fact is easy to explain, but in Warotamasikkkhadit's analysis it is an anomaly.

Same Form ⇒ Same Word Class

Sometimes the structural parallels are not even spelled out at all, but merely asserted. For example:

- 22)
- | | | | |
|-----------------------------------|------|--------|-------|
| khăw | dəən | taam | lûuk |
| he | walk | follow | child |
| 'He walked, following his child.' | | | |

- 23)
- | | | | |
|------------------------------|------|--------|--------|
| khăw | dəən | taam | thănǒn |
| he | walk | follow | street |
| 'He walks along the street.' | | | |

⁹Note the assumed identity reflected in the use of the pronoun *they*.

¹⁰As Kitima Indrambarya has pointed out (Indrambarya, 1993a), Thai words that can be shown to be verbs and adverbs by independent criteria are negated by *mây*, while noun-headed and preposition-headed predicates are negated by *mây chây*.

Our argument is that (3) and (4) are structurally the same and *the meaning of taam* 'to follow' *remains the same* in both sentences. (Warotamasikkhadit, 1992, p. 71)

However, in actual fact there is no argument here at all, only an unsupported assertion. Observing that (22) and (23) are superficially similar in sequence in no way constitutes a proof that they are identical in structure, and the claim that the meaning is the same is contradicted by the English glosses. Because judgments of synonymy are inherently quite subjective, the unsupported assertion of a native speaker cannot be taken at face value in such matters, especially when the native speaker is part of a culture with a long literary tradition and when he has already shown an overwhelming inclination to equate any two elements that happen to be written and pronounced the same way.

The Chomskyan pro-argument

Chomskyan analyses in general do not qualify as generative because they do not provide formal and explicit rules. What they normally do provide is informal arguments, but even the arguments are sometimes replaced by what might be called, by analogy with "pro-form," a "pro-argument": the word "clearly" or one of its synonyms. Warotamasikkhadit avails himself of the same strategy at several points in his paper:

It is evident that the words which are called prepositions in the traditional Thai grammars do not constitute a single form class like English as defined by Curme...or by Fries.... (Warotamasikkhadit, 1992, p. 70).

24)

khăw	càak	bâan	maa	(6)
he	leave	home	come	
'He left home (to come here).'				

It is evident that *càak* 'to leave, to depart', in (6) is a verb where *bâan* 'home' is the object. (Warotamasikkhadit, 1992, p. 71)

"It is evident" is not an argument, but simply an assertion. Bare and unsupported assertions of course have no status in a science, including the science of language, so statements such as the above carry no weight in resolving the question at hand.¹¹

¹¹The use of "it is evident" here calls to mind Chomsky's "clearly," which is employed in similar situations to similar effect.

Transformational Analyses

Prepositions versus conjunctions

The procedure followed by Warotamasikkkhadit in the examples discussed in his Section 4, “Conjunctions mistakenly called prepositions,” is simple: (1) find an example of a form occurring in a particular desired environment, e.g., *kàp* in the environment of a clause, like 25) below, (2) use that environment to determine the word class of the form in question, e.g., “conjunction”; and (3) for examples in which that form does not occur in the desired environment, e.g., 25b), create an ad hoc “underlying representation” in which it does. When a linguist follows the Chomskyan practice of assuming powerful transformational machinery without bothering to formalize it, the result is a vacuous analysis. Nice examples of this are Warotamasikkkhadit’s analyses of *phîa* ‘for,’ *tææ* ‘but,’ and *kàp* ‘with.’ Only the *kàp* examples will be considered here:

25)

phôo	kin	khâaw	kàp	mææ	(kin khâaw)	(51)
father	eat	rice	and	mother	(eat rice)	
‘Father eats with mother.’						

It clearly shows that *kàp* ‘with, and’ in (51) is a conjunction. (Warotamasikkkhadit, 1992, p. 74)

“Clearly” here is the classical Chomskyan pro-argument, a word that takes the place of an actual logical argument. In fact, however, 25) by itself shows nothing. A careful analysis shows Warotamasikkkhadit’s example to be really a conflation of two structurally quite different examples, 25a) and 25b):

25a)

phôo	kin	khâaw	kàp	mææ	kin	khâaw
father	eat	rice	and	mother	eat	rice
‘Father eats and mother eats.’						

25b)

phôo	kin	khâaw	kàp	mææ
father	eat	rice	with	mother.
‘Father eats with mother.’				

In 25a), *kàp* is a true conjunction, and in 25b) it is a preposition. Conjunctions and prepositions differ in their constituent or dependency structure. A coordinating conjunction occurs in the middle of a construction, and forms a polycentric construction with one or more preceding and one or more following phrases of the same kind, while a preposition occurs at the periphery of a construction, and forms an exocentric construction with a preceding or following phrase. This difference in constituency can often be tested by “movement” phenomena: thus if Warotamasikkkhadit’s analysis were linguistically valid, both of the following

sentences should be impossible, because if *kàp* is really a conjunction, *kàp mǎæ* (*kin khâaw*) is not a constituent and should not be susceptible to ‘preposing’:

26)

**kàp mǎæ kin khâaw ná? phôo kin khâaw*
 and mother eat rice father eat rice
 ‘And mother eats, father eats.’

27)

kàp mǎæ phôo kin dūay tɛɛ kàp chǎn
 and mother father eat together but with me
phôo mây khay kin dūay ləəy
 father not used to eat together at all.
 ‘With mother, father eats together, but he never eats together with me.’

Yet while 26) confirms the prediction, sentence 27) is unproblematic and thus disconfirms it. On the other hand, if Warotamasikkhadit had proposed the opposite analysis, that *kàp* is a preposition in both examples, then 27) would be supporting evidence and 26) would be counterevidence. The only conclusion to draw from this seemingly paradoxical situation is that *kàp* in 25a) and 26) is a conjunction, and *kàp* in 25b) and 27) is a preposition.

Verbs versus adverbs

The analysis of *klay* utilizing an unmotivated and completely gratuitous (and completely unformalized) rule of *càak* deletion on p. 72 further illustrates the same point:

28)

bāan khǎw yùu klay càak bāan chǎn (25)
 home he be far leave home I
 ‘His house is far from my house.’

29)

bāan khǎw yùu klay bāan chǎn (26)
 home he be far home I
 ‘His house is far from my house.’

klay ‘far’ in (25) is a member of a verb string *yùu klay càak* ‘to be far away from’ showing that it is a verb and *klay* ‘far’ in (26) is structurally and semantically the same as in (25) except *càak* to leave, to depart’ is deleted in (26). (Warotamasikkhadit, 1992, p. 72)

The claim that the forms *klay* in 28) and 29) are “semantically the same,” though subjective, seems unproblematic here. However, no evidence is given to support the

claim that the two examples are *structurally* the same. In a Chomskyan deep structure, though, all things are possible.

Verbs versus suppletive prepositions

It is not the case that a non-generative transformational analysis will *always* regard words with the same spelling and pronunciation as the same word. Warotamasikkhadit's analysis of *kāp* in the following example is instructive in this respect:

- 30) phỏm tham man kàp mii (52)
I make it with hand
'I made it with my own hand.'

- 31) khǎw chǎy chón kin khâaw (53)
 he use spoon eat rice
 'He uses a spoon in eating.'

- 32) khăw kin khâaw dûay chónn (54)
he eat rice with spoon
'He eats rice with a spoon.'

kàp 'with' in (52) is a *variant* of *dúay* 'with,' an instrumental marker, which is *derived from* *cháy* 'to use,' to be discussed below... (54) is *derived from* (53) where a *movement transformation* is applied and *cháy* 'to use' *becomes* *dúay* 'with' to signify an instrumental case.... *kàp* 'with' can be a *variant* of *dúay* 'with' as in (52). (Warotamasikkhadit, 1992, p. 74)

I consider kææ 'to' derived from hây 'to give' to signify a dative case. (Warotamasikkhadit, 1992, p. 75)

For these examples, the “same-form = same word” principle has been overruled by a competing principle which is just as prevalent in popular syntactic analysis, and just as unmotivated by the syntactic facts of the language: analysis by paraphrase or by English translation. What does it mean in a formal grammar to say that *kàp* ‘with’ in (30) is “a variant of *dûay* ‘with,’” or that this *kàp* is “derived from” *cháy* ‘to use’ in (31)? Just what does a “variant” look like in a formal grammar, and what kind of a rule is it that “derives” one word from another? Even Chomskyan grammar does not go this far in pulling forms and rules out of a silk hat, and we have to go back to the generative semantic excesses of the late 60s to find analyses with comparably flagrant disregard for principles of rigorous linguistic analysis.

Syntactic Analysis Using Syntactic Criteria

As good examples of using syntactic criteria to determine syntactic classes of words, I recommend Kitima Indrambarya's "The Status of the Word Hay in Thai" (1990) and "The Grammatical Function of the Word hay in Thai" (1993b). These papers address a similar question to that addressed by Warotamasikkhadit, but use different methods and arrive at quite different results. She accepts two things from the outset: (1) forms that differ in pronunciation, meaning, *or* distribution are different lexical entries, and (2) the best criteria for doing a syntactic classification of the words of a language are syntactic criteria, that is, distribution. She carefully and logically examines the examples of Thai constructions containing *hây*, and finds that from a syntactic point of view, there are six different words pronounced as *hây* in Thai. Five of them are illustrated in the following example (Indrambarya, 1990, p. 33):

33) (KI's (1))

nóy bòok wâa mêe hây tǒy hây kuncêe deen
 Noy say that mother cause Toy give key Dang
 hây sôm rôt míchanán ca? tii hây
 cause repair car otherwise will hit
 'Noy said that mother ordered Toy, under the threat of being hit, to give the key to Dang to fix the car for (mother).'

Below is a list of the various *hây* forms and an indication of Indrambarya's distributional reasons for distinguishing them (Indrambarya, 1990):

52)

- hây*₁: ditransitive verb; [___ NP₁ NP₂]
- hây*₂: ditransitive verb; [___ NP₁ PP₂]
- hây*₃: ditransitive verb; [___ NP₁ NP₂ S], Patient-to-Patient control
- hây*₄: causative verb; [___ NP S], Patient-to-actor control
- hây*₅: impersonal causative verb; [___ NP S], Patient-to-actor control, no referential subject
- hây*₆: benefactive/malefactive adverb; same syntactic distribution as the underived adverb *wáy* 'lying'

The following sentences taken from Indrambarya's paper exemplify these different but homophonous lexical entries:

34)

- a) lék *hây*₁ còtmăay deen (32)
 Lek give letter Dang
 'Lek gave Dang a letter.'
- b) lék *hây*₂ còtmăay kêe deen (33)
 Lek give letter to Dang
 'Lek gave a letter to Dang.'

- c) chán hây₃ nánŋh dèkdèk ʔaan (34)
 I give book children read
 'I gave a book to the children to read.'
- d) nít hây₄ lék càt dòokmáay (56a)
 Nit cause Lek arrange flower
 'Nit had Lek arrange the flowers.'
- e) hây₅ khruu maa ciŋciŋ chán kóɔ mây klua (88)
 though teacher come true I ASP not fear
 'Even if the teacher really comes, I am not afraid.'
- f) nídaa thŋh krapǎw hây₆ (wiinaa) (93)
 Nida carry bag for Weena
 'Nida carried a bag for (Weena).'

All of these distinct items can be justified on the basis of occurrence in distinct grammatically significant environments, syntactic properties parallel to the properties of uncontroversial members of other classes, and consistent complement frame and meaning for each separate item. The cost is homophonous lexical items, and the gain is grammatical generalizations.

CONCLUSION

So does Thai have word classes? Yes, because every language has word classes; otherwise it would be impossible for its speakers to learn languages and to distinguish between sentences of their language and random strings of words. Are the syntactically defined word classes in Thai different from the word classes found in other languages of the world? Do we find different kinds of categories, sets of items with fuzzy and indeterminate boundaries, or different specific classes of words, classes unknown to English, Swahili, and Tagalog? The answer is "yes" from the point of view of non-generative and unconstrained grammatical frameworks such as Chomskyan grammar or autolexical grammar, since in such frameworks, anything goes. They make no concrete testable claims, and thus can never be disproven. From the point of view of lexicase dependency grammar, however, the answer is "no." When Thai is observed from the point of view of a sufficiently constrained and generative framework of analysis such as lexicase dependency grammar, it turns out that clearly defined syntactic word classes do emerge, and that they fit comfortably into the same categories found in other unrelated human languages. These classes permit the capture of language-specific and cross-linguistic grammatical generalizations and confirm the cross-linguistic validity of the theory within which they are defined.

ACKNOWLEDGMENT

I would like to thank Kitima Indrambarya and Supriya Wilawan, Amara Prasithrathsint, and the other participants in the London panel, before which an oral version of the paper was given, and William O'Grady for their helpful comments and criticisms at various points in the evolution of this paper.

REFERENCES

- Bloomfield, L. (1926). A set of postulates for the science of language. *Language*, 2, 153–64; reprinted in M. Joos (Ed.), *Readings in linguistics* (pp. 26–31). New York: American Council of Learned Societies.
- Bloomfield, L. (1933). *Language*. New York: Henry Holt and Company.
- Clark, M. (1978). *Coverbs and case in Vietnamese*. Pacific Linguistics, Series B, No. 48. Canberra: Australian National University Press.
- Covington, M. A. (1986). Grammatical theory in the Middle Ages. In T. Bynon & F. R. Palmer (Eds.), *Studies in the history of western linguistics* (pp. 23–42). Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1977). Where have all the adjectives gone? *Studies in Language*, 1, 19–80.
- Indrambarya, K. (1990). The status of the word Hay in Thai. University of Hawai'i *Working Papers in Linguistics*, 22, 33–72.
- Indrambarya, K. (1993a). *Are there prepositions in Thai?* Paper presented at the Third Annual Meeting, Southeast Asian Linguistics Society, Honolulu, 7–10 May.
- Indrambarya, K. (1993b). The grammatical function of the word hay in Thai. In S. Luksaneeyanawin (Ed.), *Proceedings of the Third International Symposium on Language and Linguistics: Pan-Asiatic Linguistics, Vol. 3* (pp. 1163–1177). Bangkok: Chulalongkorn University.
- Indrambarya, K. (1994). *Subcategorization of verbs in Thai*. Unpublished doctoral dissertation. University of Hawai'i, Honolulu.
- Kullavanijaya, P. (1974). *Transitive verbs in Thai*. Unpublished doctoral dissertation. University of Hawai'i, Honolulu.
- Lehman, F. K. (1990). Outline of a formal syntax of numerical expressions, with especial reference to the phenomenon of numeral classifiers. *Linguistics of the Tibeto-Burman Area*, 13, 89–120.
- Panupong, V. (1970). *Inter-sentence relations in modern conversational Thai*. Bangkok: The Siam Society.
- 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). Innsbrucker Beiträge zur Sprachwissenschaft. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Pollock, J-Y. (1989). Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry*, 20, 365–424.
- Prasithrathsint, A. (1985). *Change in the passive constructions in written Thai during the Bangkok period*. Unpublished doctoral dissertation. University of Hawai'i, Honolulu.
- Savetamalya, S. (1987). A reanalysis of auxiliaries in Thai. University of Hawai'i *Working Papers in Linguistics*, 19, 1–44.
- Savetamalya, S. (1989). *Thai nouns and noun phrases: A lexicase analysis*. Unpublished doctoral dissertation. University of Hawai'i, Honolulu.
- Schiller, E. (1992). Parts of speech in Southeast Asian languages: An autolexical view. In S. Luksaneeyanawin (Ed.), *Pan-Asiatic linguistics: Proceedings of the Third International Symposium on Language and Linguistics, Vol. 2* (pp. 777–790). Bangkok: Chulalongkorn University.

- Starosta, S. (1985). Mandarin case marking: A localistic lexicase analysis. *Journal of Chinese Linguistics*, 13, 215–266.
- Starosta, S. (1988). *The case for lexicase: An outline of lexicase grammatical theory*. London: Pinter Publishers.
- Starosta, S. (1991). The great AUX cataclysm: Diachronic justification for a synchronic analysis. In B. Lakshmi Bai & B. Ramakrishna Reddy (Eds.), *Studies in Dravidian and general linguistics: A festschrift for Bh. Krishnamurti*. Department of Linguistics, Osmania University, Hyderabad. [Unabridged preliminary version in University of Hawaii *Working Papers in Linguistics* 17, 95–114].
- Thepkanjana, K. (1992). Transitivity continuum in Thai. In S. Luksaneeyanawin (Ed.), *Pan-Asiatic linguistics: Proceedings of the Third International Symposium on Language and Linguistics*, Vol. 1 (pp. 308–323). Bangkok: Chulalongkorn University.
- Warotamasikkkhadit, U. (1972). *Thai syntax: An outline*. The Hague: Mouton.
- Warotamasikkkhadit, U. (1992). There are no prepositions in Thai. In C. Bamroongraks, W. Khannittanan, L. Permcharoen, S. Anivan, N. Bishop, G., D. Srinarawat, S. Chenvidyakarn, P. Thongsawang, M. Peterson, J. Hinds, Y. Hoonchamlong, & N. Ronakiat (Eds.), *The International symposium on language and linguistics* (pp. 70–76). Bangkok: Faculty of Arts, Thammasart University.
- Wilawan, S. (1993). *The reinterpretation of so-called 'serial verb constructions' in Thai, Khmer, Mandarin, and Yoruba*. Unpublished doctoral dissertation. University of Hawai'i, Honolulu.

