The syntax of nouns and noun phrases in dated pre-Angkorian inscriptions*

Chhany SAK-HUMPHRY
University of Hawaii at Manoa

1. Introduction

1.1 Background

Any diachronic analysis of Khmer must be based on the extensive corpus of inscriptions.[1] According to Jacob (1960: 351; 1965: 143; *1991), conventionally recognized periods in the development of Khmer are 'Old Khmer', 'Middle Khmer', and 'Modern Khmer'. The Old Khmer period includes the pre–Angkorian Khmer of the earliest inscriptions to A.D. 802 and Angkorian Khmer, attested from 802 to 1431 (the fall of Angkor). 'Middle Khmer' covers the period of transition from Old to Modern Khmer, from approximately 1431 to 1800, and marks the culmination of various complex changes in both the phonology and grammar. 'Modern Khmer' is considered to extend from about 1800 to the present. All three periods are, of course, historical fictions, or rather historical–linguistic constructs, in the sense that there were no abrupt demarcations between one stage of the language and the next.

1.2 Previous analyses of pre–Angkorian syntax

Whilst a number of scholars have contributed signficantly to—and indeed laid the foundations of—the historical study of Old Khmer, there are so far no syntactic analyses of Old Khmer based on a generative framework.[2] Jenner

---

[1] The Founan period is a historical period. The inscriptions from this period are in Sanskrit; the vernacular language of Founan is unknown.

[2] Cœdes 1955; Jacob 1960, 1963, 1970; Dupont 1953; Jacques 1976; Jenner 1980a, 1980b, 1981; Jenner and Pou 1980–81; Pou 1970a, 1970b; Long Seam 1973.

published a series of articles on Old Khmer syntax, and we possess now a diachronic survey of Khmer function words by Jacob.[3]

Texts of most of the pre–Angkorian inscriptions can be found in Cœdes' eight–volume *Inscriptions du Cambodge* (1937–66) or scattered throughout the *Bulletin de l'Ecole Française d'Extrême–Orient [BEFEO]*

1.3 *Primary sources*

The pre–Angkorian corpus can be divided into dated and undated inscriptions.[4] The data used in this investigation of the syntax of nouns and noun phrases of the pre–Angkorian dated inscriptions are not drawn from the original texts, but rather from the transliterated versions found in Jenner's (1983–84) unpublished work. Sixty–six texts have been taken from two volumes of Jenner's work and are listed in Appendix A.

Working with data from an ancient language imposes serious limitations on the analysis. The compilation and interpretation of the texts, the primary data for the analysis, is itself a major analytical task. As Jenner put it: "The linguistic archæologist must first consider the orthographic shape of each item; he must then translate that shape into phonological terms, analyze its morphology, determine its syntactic function, identify it with later or cognate forms, and eventually set up a hypothesis regarding its meaning" (Jenner *CPAK* 1981–82: iv). All this work must be done without a co–operative native–speaking consultant whose intuition can be accessed to confirm or discount a particular hypothesis.

1.4 *Goals and objective of this study*

The purpose of this study is twofold: (1) to apply lexicase dependency theory to the analysis of certain grammatical aspects of nouns and noun phrases of dated pre–Angkorian inscriptions to determine whether the grammatical properties of these inscriptions can be insightfully described and illuminated within this formal and explicit theory, and (2) to identify any areas in which the data prove to be incompatible with the claims made by the theory, thereby possibly necessitating a modification of the theory itself.[5] It is to be expected that in the absence of a native–speaking consultant, a narrowly constrained universal linguistic theory will provide assistance in choosing among alternative hypotheses about the correct analysis of particular constructions. At the same time, data from a natural text from another non–Indo–European language will be helpful in developing a theory that can seriously claim to be universal.

---

[3] The latter was not available to me at the time of writing; Jenner's contribution to the London Shorto Festschrift concerns Angkorian Khmer. Pou contributed a survey of some basic OKhm. constructions to this journal (1980).

[4] Jenner says that "the dated inscriptions are those which contain their own internal dating" (Jenner 1982).

[5] See Tesnière 1959 for a general reference on dependency grammar and Starosta 1988 for an introduction to the lexicase implementation of this theory.

1.5 Methodology

The corpus of this study is composed of the texts presented in Philip N. Jenner's unpublished work, *"Textes Vieux Khmers faisant partie du Corpus des Inscriptions du pays khmer"* (*Tome 1.1 et 2, Tome 1.3, 1983–4*). I began analyzing the texts of the earliest of the dated inscriptions, which had already been segmented into their constituent sentences by Jenner's translations of each text in French. Each example cited in this thesis is labeled with a number identifying its location in these texts. Each word in the texts was checked against Jenner's pre–Angkorian lexicon and labeled in terms of the categories made available by the lexicase theory, and an explicit lexicase dependency representation was drawn for each sentence. These representations conform to the constraints imposed by the lexicase theory. The purpose of this task was not to check on Jenner's translation of the text, but rather to investigate the structure of the sentences of the pre–Angkorian dated inscriptions within the lexicase framework. However, this study indicates places where the requirements of the theory suggested alternative translations or analyses that differ from those proposed by Jenner.

1.6 The Lexicase model

This investigation of nouns and noun phrases of pre–Angkorian dated inscriptions is formulated within the lexicase dependency grammar framework developed over the last twenty years, primarily by Stanley Starosta and was carried out in accordance with the principles and constraints of this theory.

A lexicase grammar is a grammar of words. It has no deep structure, no transformations, and no phrase structure rules. It represents the structure of a sentence solely in terms of a network of dependency relations obtaining among pairs of words in the sentence. Co–reference relationships are described in terms of coindexing words. The syntactic and semantic properties of words are characterized in terms of contextual and non–contextual features of lexical entries (Starosta [forthcoming]). This theory attempts to capture cross–linguistic generalizations, and makes a claim about human language in general.

A lexicase grammar is a set of generalizations about the internal compositions, external distributions, and lexical relationships of the words in the language (Starosta 1988: 2). The relationship among lexicase rules can be represented in terms of the flow chart in Appendix B.

The analysis presented here does not include the set of formalized rules that would be required by a complete grammar. It does, however, provide fully specified dependency representations of a broad range of example sentences relevant to a description of the nouns and noun phrases of the Old Khmer dated inscriptions. For each representation, the functional label of each binary dependency relationship is specified in terms of an indexed contextual feature in the matrix of the regent term of the relation. Such labels include case relations [CRs], case forms [CFs], predication [prdc], or finiteness [fint] of the dependents.

This study is divided into the following sections:

2. Overview of dated pre–Angkorian syntax

The focus of this analysis is on noun and noun phrase structures. Because so many examples of the constructions in the data presented here are non–sentential, and since this is an accurate reflection of the language of Old Khmer inscriptions, a brief description of the overall structure of sentence patterns of the dated pre–Angkorian inscriptions is needed to set the foundation of this work. This overall description provides a sketch of the general syntactic properties of the Old Khmer language, many of which are relevant to a description of the constructions which depend on N[oun]s in N[oun]P[hrase] structures. In this study, the term 'Old Khmer' [OKhm.] is used to mean 'Old Khmer dated pre–Angkorian'.

2.1 *Basic sentence patterns*

A basic sentence structure has only one clause. A sentence is any phrase that has a word marked [+prdc] as its lexical head, that is, a verb, or a prdc–inflected preposition or a noun. The focus of this section is on the analysis of sentences with and without verbs. The outline of this presentation is: (1) a brief description of verbless sentences and their subcategorization, (2) verbal sentences, and (3) conjoined sentences.

2.1.1 *Verbless sentences*

This section examines sentences without verbs as the head of the constructions. In principle, lexicase allows two types of verbless sentence constructions: (1) the nominal predicate where the regent noun N_2 is the head of the construction, bears the feature [+prdc], and has N_1 as its nominative patient dependent; and (2) the predicate prepositional phrase where the preposition P_1 itself bears the [+prdc] feature.

The Old Khmer pre–Angkorian dated inscriptions have only one example of an independent clause with a nominal predicate NP (see example 8 of section 4.2.2.1) and no examples of prepositional predicate phrases as described above. Instead, most of the verbless sentences occur as relative clauses in NPs composed minimally of two nouns, a regent noun N_1 and a nominal dependent N_2, which is