

MORPHOLOGICAL TYPOLOGY FROM SOUTHEAST ASIAN VIEWPOINT

Makoto Minegishi

Institute for the Study of Languages and Cultures of
Asia and Africa (ILCAA)
Tokyo University of Foreign Studies

1. Introduction

In the modern linguistic typology the notion of inflectional, agglutinative, polysynthetic and isolating languages, which dates back as early as von Humboldt (1836), is regarded as obsolete. This classification, as far as isolating type is concerned, seems still effective, especially in mainland Southeast Asia and in China where most of the languages are classified as 'isolating' and almost no grammatical clue is found in the domain of a word.

In this paper we will propose a new typological classification using the notion of variable and value. The formalized typology will provide a new perspective for possible types of grammatical description and that for geographical distribution of language types.

2. Formal Approach to Linguistic Typology

2.1 Re-definition of the Traditional Typology

Minegishi (2000a) attempts to re-define traditional morphological typology in terms of the mathematical formalization, i.e., **variable** and **value**. By introducing the notions of lexical or semantic properties (shown as L) and grammatical or syntactic properties (G) as variables within the domain of a word (W) indicated with brackets, the traditional language types are represented as below.

(1) **Inflectional language** $W = \{L, G\}$

Agglutinating language $W = \{L\}$, or $W = \{G\}$

Isolating language $W = \{L\}$

This 'algebraic' formula has been already suggested by Sapir (1921:25-26) as $(A+b)$ where A denotes a radical element and b a grammatical element, although little effort was made for elaborating the notion.

It should be noted that unlike Sapir's b , G denotes only grammatical or syntactic properties: we will not concern here with affixation in the word formation level. Thus, a word in Cambodian, like one in Thai, is represented as $W = \{L\}$, because its affixation works only for word formation. Also, although the above domain denotes a word, we will further define a domain as any syntactic unit: a word, a phrase or a sentence, according to what properties are in the scope of discussion.

2.2 New terminology for the formalized typology

Let us define the symbols used for formalizing classifications as follows:

- (2) *Parameter m* : number of the grammatical properties G in a syntactic domain,
Parameter n : number of the values for the grammatical properties,
 j, k : constant number,
 $\{ \}$: syntactic domain, i.e., any word, phrase or sentence.

We introduce here a new terminology for classification based on m , the number of variables G and n , that of values for G in a syntactic domain as follows.

- (3) **Definite category language (DCL)** both $m = j$ and $n = k$ (constant)
Indefinite category language (ICL) either m is not equal to j or n is not equal to k
Non category language (NCL) no G exists in the syntactic domain.

According to the definition above, the definite category language (**DCL**) is defined as a language whose number of G and that of values for G are both constant in a given domain. The indefinite category language (**ICL**) is a language whose

number of G is not fixed in a syntactic domain: either m or n is not constant.¹

The non category language (NCL) is that no G is existent in a syntactic domain. By assuming a word (W) as a syntactic domain, the classifications apparently correspond to the traditional inflectional, agglutinative and isolating languages respectively. We will see below how the outcome of our re-definition is different from that of the traditional typology.

3. Definiteness of a Category

Let us consider the Latin case as an example of the definite category language (DCL). A Latin noun has a definite number j of syntactic properties shown as $W = \{L, G_1, G_2, G_3\}$, in this case the number $j=3$, and each grammatical category as a variable can take a fixed number k of values. That is, any Latin noun has three variables, G_1 (i.e., *gender*), G_2 (i.e., *number*) and G_3 (i.e., *case*), and each variable can have a definite number of values.

The variable, shown here as the function, *gender*() can have one of the value '*masculine*', '*neuter*' or '*feminine*': thus $k_1=3$, the function *number*() '*singular*' or '*plural*': thus $k_2=2$, the function *case*() '*nominative*', '*vocative*', '*accusative*', '*genitive*', '*dative*', '*ablative*', or '*locative*': thus $k_3=7$, respectively. Thus, according to the formalization, a Latin word '*dominus*' can be represented as follows:

(4) *dominus* 'host'

$$\begin{aligned} W &= \{L = (\textit{domin-}), \\ G_1 &= \textit{gender}(\textit{masculine}), \\ G_2 &= \textit{number}(\textit{singular}), \\ G_3 &= \textit{case}(\textit{nominative})\} \end{aligned}$$

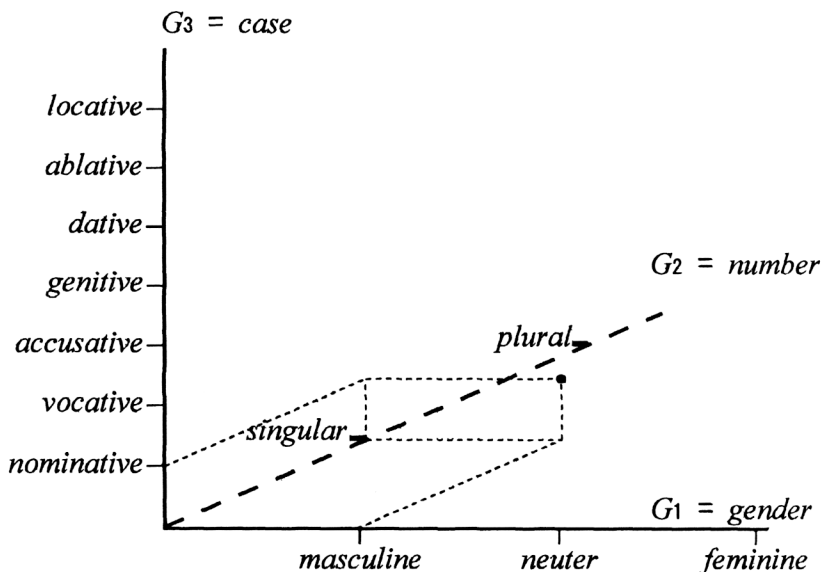
Since these grammatical properties are defined as variables, if we assume each variable as an axis, in the above case the axes G_1, G_2, G_3 , the inflectional form '*dominus*' of a word '*domin-*' can be represented as a coordinate (masculine, singular, nominative) in a three-dimensional space as in (5). It can be generalized that a word in the DCL language can be

represented as a coordinate in the j -dimension space where j is the constant number of G .

Also in the DCL, the coordinate of a word is fixed since each variable (shown as a dimension) takes only one value of possible ones at one time.²

(5)

3-Dimensional Representation of Latin 'dominus'



This is equivalent to saying that in case of DCL, since the number j of variables and k that of its values are constant, any word in the language can be represented in an inflectional table. This is the formal re-definition of the **traditional paradigm**. In other words, any language any word form of which can be represented in a traditional paradigmatic table is defined as a definite category language.

It is interesting to note that not only Latin, Sanskrit and other typical fusional languages, but also those superficially resembling isolating languages like English can be represented similarly, by expanding the domain, a word into a phrase, and by introducing a new variable *definiteness*(), which has two values, i.e., 'definite' and 'indefinite'.