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:

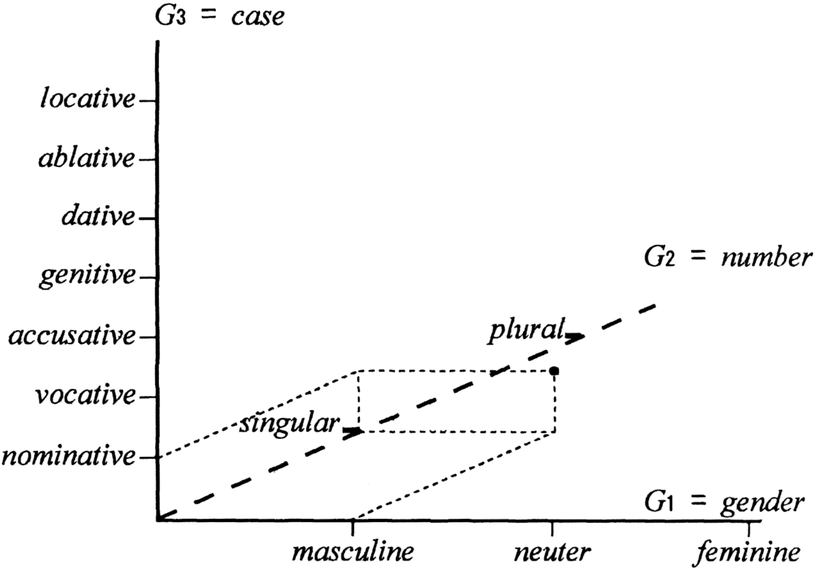
(4) dominus 'host'
W={L=(domin-),
G_1 *= gender(masculine),*
G_2*= number(singular),*
G_3 *=case(nominative)}*

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.

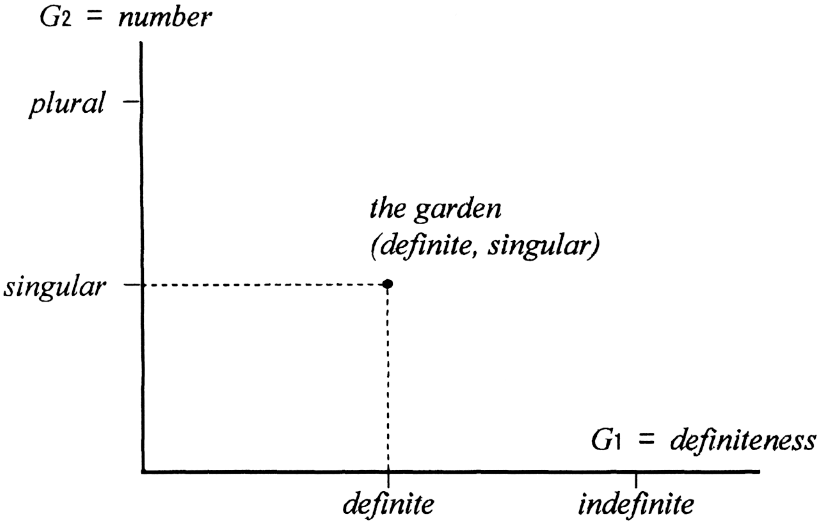
It is interesting to note that not only Latin, Sanskrit and other typical fusional languages, but also those superficially resembling isolating languages like English can be represented similarly, by expanding the domain, a word into a phrase, and by introducing a new variable *definiteness( ),* which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

(6) the garden
NP={L, G_1, G_2,}
={L=(garden),
G_1=definiteness(definite), G_2=number(singular)}

In case of English any noun phrase, not a word, can be described in a two-dimensional plane, whose axes are *definiteness( )* and *number( )*. The values for the *definiteness( )* is either *'definite'* or *'indefinite'*, those for the *number( )*, *'singular'* or *'plural'*.

Thus, English noun phrase 'the garden' can be located in a two-dimensional surface shown in (7).

(7)
2-Deimensional Representation of English Noun Phrase 'the garden'



It should be pointed out here that in DCL, a grammatical category can be defined as a set of *Gs*. That is, for example, being a Latin noun means that the word is a coordinate in the three-dimensional space, G_{number}, G_{gender}, G_{case}. Similarly, being a Latin verb means it can be represented in the G_{mood}, G_{aspect}, G_{tense}, G_{person}, G_{number} space. Furthermore, by expanding the syntactic domain into a phrase instead of a word, an English noun phrase can be placed in G_{number}, $G_{definiteness}$: on the two-dimensional surface, and an English verb phrase, in the G_{mood}, G_{aspect}, G_{tense}, G_{person}, G_{number} space.[3]

Thus ignoring the name and value of each function, both can be shown in the generalized model DCL.

4. Agglutinative Languages: whether DCL or ICL?

As we have seen that in spite of the superficial difference of fusional and isolating, both Latin and English can be classified as DCLs. We will see the case of two languages, Turkish and Japanese. Both are said to be of the typical agglutinative type. Here we will see how they differ from each other in their case marking system.

Turkish has case markers as follows:

(8) (a) *ev* (a house /houses) $NP = \{L, G_1\ case(nominative)\}$
(b) *ev-i* (a house /houses) $NP = \{L, G_1\ case(accusative)\}$
(c) *ev-den* (from a house) $NP = \{L, G_1\ case(ablative)\}$
(d) *ev-de* (at a house) $NP = \{L, G_1\ case(locative)\}$
(e) *ev-e* (to a house) $NP = \{L, G_1\ case(dative)\}$

where NP (Noun Phrase) is a syntactic domain.

It should be noted that no noun in Turkish can have more than two case markers at a time, which means that its case marking can be represented as a paradigmatic table as shown above. Thus by definition, it can be regarded as DCL. In Japanese, however, two case markers *-made* (till), *-kara* (from) can be followed by other case markers such as nominative and accusative,[4]

(9) (a) *koko* (here) $NP = \{L, G_1 = case(0)\}$
(b) *koko-ga* (here) $NP = \{L, G_1 = case(nominative)\}$
(c) *koko-o* (here) $NP = \{L, G_1 = case(accusative)\}$
(d) *koko-ni* (here) $NP = \{L, G_1 = case(dative)\}$
(e) *koko-de* (here) $NP = \{L, G_1 = case(locative)\}$
(f) *koko-made* (till here) $NP = \{L, G_1 = case(allative)\}$
(g) *koko-made-ga* (up to here) $NP = \{L, G_1 = case(allative, nominative)\}$
(h) *koko-made-o* (up to here) $NP = \{L, G_1 = case(allative, accusative)\}$
(i) *koko-made-ni* (here) $NP = \{L, G_1 = case(allative, dative)\}$

(j) *koko-made-de* (here) *NP = {L, G1 = case(allative, locative)}*
(k) *koko-kara* (from here) *NP = {L, G1 = case(ablative)}*
(l) *koko-kara-ga* (from here) *NP = {L, G1 = case(ablative, nominative)}*
(m) *koko-kara-o* (from here) *NP = {L, G1 = case(ablative, accusative)}*
(n) *koko-kara-ni* (from here) *NP = {L, G1 = case(ablative, dative)}*
(o) *koko-kara-de* (from here) *NP = {L, G1 = case(ablative, locative)}*

These examples illustrate that Turkish has a 'definite category' regarding its case marking system as its neighboring European languages do, but Japanese is different in that it allows duplicated case marking: as a result assuming the case as grammatical property G_{case} and representing it as one dimension line, a case-marked noun phrase in Japanese cannot be located properly as a coordinate on the line because one *variable has two values at one time: thus Japanese should be classified as ICL.*

Regarding the number *n* of values for each grammatical property *G* in ICL, it seems that in most cases a category *G* has either a specific value, or no value, which means the category is optional.

Assuming that the optionality of the latter case is denoted as *G*(0), the number n in ICL in general is 2: either a specific value or 0.

That is, in ICL a grammatical property is usually optional, thus it is not essential part of a word.

It follows that in the above Japanese example (9) (l), the NP should be represented as follows:

(10) *NP* (*Compound Noun Phrase*) = *{{L} {G} {G} ...}*
NP {{L = (*koko*)*}{*G_{case}*(ablative)}{*G_{case}*(nominative)} ...}*
NP {{L = (*koko*)*}{*G_{case}*(-kara)}{*G_{case}*(-ga)} ...}*

where the NP is regarded as a **compound** of mutually independent domains indicated with { }s.

Structure (10) above illustrates that defining **word class** in ICL should be in terms of **syntagmatic relations**, i.e., the word's possibility of co-occurrence with other linguistic forms. In the above example, the word 'koko' has a nominal property because it can be followed by a case marker: '-kara' has a property as a case marker because it can follow a noun, etc. Thus 'word class' in ICL is a relative notion whereas 'grammatical category' in DCL is an absolute one in that it can be defined in **paradigmatic relations**, i.e., definable regardless of the existence of other words or phrases.

5. Non Category Languages

Non category languages can be represented as follows.

(11) $W = \{L\}$

This simple formula (11) looks adequate for representing the traditional 'isolating' type of languages as well, in that no grammatical clue G exists as any linguistic form in the domain W. Furthermore, it might look too simple in that most linguistic theories expect the existence of G as any linguistic form in a syntactic domain W, P (Phrase) or S (Sentence). We may call this classic **morphosyntactic** viewpoint: syntax is inseparable from morphology as in the case of Latin grammar. Consequently most topics of the linguistic theories, taking the morphosyntactic view for granted, concern forms and their representation; such as meaning of morphemes, case marking, etc. As a result, modern linguistics would be at a loss for terminology to describe the above formula.

Let us consider now the relations of words in a sentence of NCL by expanding the above (11).

(12) $S = \{L\}\{L\}\{L\}...$

(12) shows that a sentence in NCL consists of mutually independent Ls. At first glance this representation reminds us of a compound. For example, a compound noun can be denoted similarly as follows.

(13) $Compound\ Noun = \{\{L\}\{L\}\{L\}...\}$

Since we do not have sufficient space, we will examine the implication of these formulae. They predict the following difficulties in syntactic analysis.

5.1 Definition of Word Class

As is the case of ICL, defining **word class** in NCL is possible in terms of the word's distribution in a syntactic domain. This is because a word in NCL as well as in ICL, but unlike one in DCL, has no support of *j*-dimension space. It means that although NCL superficially resembles English which has lost most of its case marking, the definition of word classes in the former should be done on the basis different from the latter. That is, defining a word class in NCL relies basically on the basis of distribution, i.e., syntagmatic relations, whereas defining a grammatical category in DCL is possible by the support of the *j*-dimension, or the expanded notion of the paradigm.

Similarly, for example, we cannot assume without close examination that if two verbs co-occur in a sentence, one is a **main verb** and the other is a **co-verb**. Neither can we regard that one is a modal **auxiliary verb** and the other is a verb. The above model (13) shows the possibility of analyzing a series of verbs as a compound verb phrase, as *there is no inflection in* NCL.

Or rather we have to start from the very beginning: what the property of the auxiliary verb, if any, is, and how it differs from that of the ordinary verb, etc. Also, this will lead *us to the prospect that* **Serial Verb Construction** should be considered seriously in this regard.

5.2 Word Order Treated as Hierarchical Structure

We must be very careful when we consider the existence of something that has no explicit linguistic form in NCL. One of such cases is word order analysis.

Taking the example of the word order in English, the subject and the object are defined, according to the generative grammar, in terms of their location in the hierarchical syntactic structure; the subject NP_1 is directly dominated by the node S and whose sister is VP, the object NP_2 is directly dominated by the node *VP* and whose sister is *V*.

(14) $S \dashrightarrow NP_1\ VP$
(15) $S \dashrightarrow VP\ NP_2$

In addition to this definition, the case theory and the theta-role theory are assumed necessary in order to restrict the function of syntactic position. This analysis is, from the author's point of view, equivalent to the following formulae.

(16) $NP_1 = \{L, G_{case}\ (nominative)\}$
(17) $NP_2 = \{L, G_{case}\ (accusative)\}$

This is to assume G_{case} in English is realized as zero-suffix in the morphological level and represent it as a position in hierarchical structure.

It should be emphasized that such an analysis cannot be applied without scrutiny. Consider the following cases in Thai.

(18) chán pay chiaŋmày
I go Chiang May
I go to Chiang May.

(18) shows that the verb pay' (go) behaves like transitive verb which takes the proper noun `Chiang May' functioning as GOAL as its direct object.

(19) chán pay rótmee
I go bus
I go by bus.

(19) shows that the verb 'pay' as a *Vt* which takes 'bus' functioning as INSTRUMENT as its direct object.

These examples clearly show that, in case a noun without case marker directly follows a verb, it is not always the case that it can be regarded as a combination of 'transitive verb' and its 'object': there is no one-to-one relationship between semantic relation and syntactic position. It would rather be possible to analyze them as *V-N* compound.

6. Geographical Distribution

We will see below the geographical distribution of the three language types proposed here. It should be noted that the distribution shown below is a tentative one based on firstly the possibility of paradigmatic description of the predicative complex, and secondly on the possibility of double case marking of the noun phrase. [5]

The classification presented here provides a new perspective for the language type distribution as follows:

1. Indo-European languages, regardless of the superficial differences in their degrees of fusionality, can be classified as DCL.

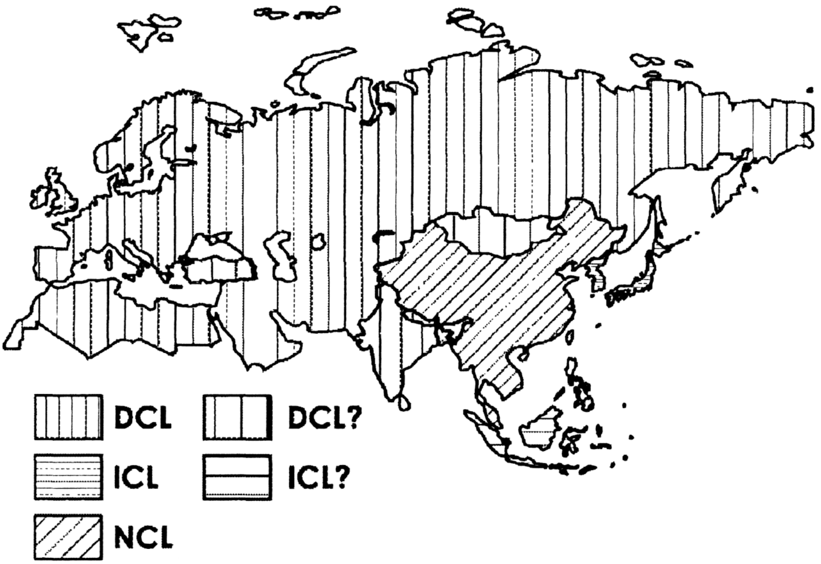
2. Since Finnish, Arabic and Hebrew words are shown in a clearly finite paradigm in previous descriptions, they are obviously regarded as DCL.

3. Some languages like Turkish and Dravidian languages, which used to be classified as the agglutinating type, would be re-defined as DCL, although closer examination would be necessary.

4. If the above languages, including not only Indo-European, but also their neighboring languages of genetically different origins, such as Afro-Asiatic, Uralic, Altaic, Dravidian, can be classified as DCL, it follows that DCLs have wide and continuous geographical distribution.

5. By contrast, NCLs or isolating ones, such as Thai, Laotian, Khmer, Vietnamese, and Chinese, are concentrated in mainland Southeast Asia and East Asia.

6. Although Japanese is ICL, we do not have the other languages classified as such for now. Probably, Korean may be ICL, as its syntax is almost similar to that of Japanese. Austronesian and Paleo-Asiatic languages might be candidates for ICL.



Map 1: Geographical distribution of language types

In order to find ICLs other than Japanese, closer re-examinations concerning Korean, African, Paleo-Asiatic, Pacific and American indigenous languages, etc. would be necessary. This necessity of close scrutiny should be emphasized because most of the linguistic descriptions, whether traditional or modern, as to these languages have been done from the viewpoint of DCLs: there must be definite *G* in a syntactic domain. Even Japanese 'inflection' has been so far described with a 'pseudo-paradigm' which is far from comprehensive. The example of Japanese double case marking clearly shows that such description is incomplete.

7. Conclusion

In this paper we have attempted to re-define the traditional typology in terms of mathematical formalization.

New criteria proposed are as follows:

Criterion 1 whether one or more grammatical properties (*G*) are included in any syntactic domain, e.g., a word or a phrase.

From this, we can classify the non category language (NCL) on the one hand, and the rest, i.e., definite category language (DCL) and indefinite category language (ICL) on the other hand. Although the re-definition gives a threefold classification almost similar to the traditional ones, isolating, fusional and agglutinative types, the former is different from the latter in that there is no ambiguity in classifying languages since the above re-definition is mathematically formalized. We can therefore distinguish each type clearly.

Criterion 2 if *G* is included in a domain, whether *m*, the number of *G* and *n*, that of the values for *G* is definite or not (whether both *m=j* and *n=k* are constant or not).

This criterion shows the idiosyncrasy of DCL, distinguished from the rest; ICL and NCL. By definition, a word in the DCL, whose *n=j* (*constant*), can be located as a coordinate in *j*-dimensional space, whereas ICL and NCL share a common feature that a word cannot be represented in a finite paradigm, thus characteristics of their word should be described on syntagmatic basis.

The following table shows the idiosyncratic and shared characteristics of the three language types.

(20)

Characteristics	**DCL**	**ICL**	**NCL**
Traditional Typology	fusional or agglutinative	agglutinative	isolating
Number of *G* in a domain	definite	indefinite	none
Paradigmatic representation	possible	impossible	
Category or class definition	by (extended) paradigm	by syntagmatic relation	
Syntactic Strategy	preferring hierarchy	preferring compounds	

These similarities of ICL and NCL imply that there is another possible way of description as to these languages; that is, a description based on syntagmatic relations and complex word and phrase formation as opposed to the paradigmatic description currently applied to DCL and some ICL.

Notes

[1] ICL is formerly called '**Infinite Category Language**' in Minegishi (2000a), but renamed here as 'indefinite', considering the mathematical characteristics of the category.

[2] The representation in *j*-dimensional space is only a metaphorical one, since each value on an axis is not a mathematically ordered number.

[3] Chomsky (1965:42-44) describes English auxiliary elements AUX as follows: *AUX → Tense (Modal) (Perfect) (Progressive)*, whose elements correspond the variables G_{tense} G_{mood}, G_{aspect}, respectively.

[4] The terms for each case are tentative ones.

[5] These two standards might contradict each other, as in the case of Dravidian and Indo-European languages in Indian subcontinent, which are classified here as DCL since they apparently have predicative paradigms, in spite of the possibility of double case marking in a few cases.

References

Anderson, S. R. 1985. Typological distinctions in word formation. Language Grammatical Categories and the Lexicon (Typology and Grammatical Description Vol. 3), ed. by Shopen, T., 3--56, 427. Cambrige: Cambridge U.P.

Chomsky, N. 1965. Aspects of the Theory of Syntax. 251. Cambridge: The M.I.T. Press.

Hayasi, Tooru. 1989. Toruko-go [Turkish Language]. The Sanseido Encyclopedia of Linguistics (Vol.3), ed. by Kamei, Takashi et al., 1383--1395. Tokyo: Sanseido.

Humboldt, Wilhelm von. 1836. Ueber die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts, Hrsg. von Eduard Buschmann, Berlin: Gedruckt in der Druckerei der Koeniglichen Akademie der Wissenschaften.

Kazama, Shinjirou. 1992. Setsubigata gengono doushi fukugoutaini tsuite [On the predicative compound of languages of suffixational type]. Kitano gengo [Languages of the North Pacific Rim: Types and History], ed by Miyaoka Osahito., 241--260, 454. Tokyo: Sanseido.

Minegishi, Makoto. 2000a. Ruikeiron-kara mita bunpou riron [Linguistic theories as viewed from linguistic typology]. Gengo Kenkyu 117:101--127.

Minegishi, Makoto. 2000b. Koritsugo kenkyuu-no houhousei-ni tsuite [Towards a Descriptive Study of `Isolating Languages']. Journal of Asian and African Studies 60: 237--247.

Sapir, Edward. 1921. Language. New York: Harcourt, Brace and World: 242.

Spencer, Andrew.1991. Morphological Theory. Oxford: UK and Cambridge, USA: 512.