

Typological notes on copula constructions

KAZAMA, Shinjiro
Tokyo University of Foreign Studies

This paper is a cross-linguistic research on copula constructions in diverse languages, focusing on their forms and their correlation to the typological characteristics of the languages examined. I introduce two cover-terms for further discussion below: linker, which is used to construct an endocentric structure, and copula, which is used to construct an exocentric structure. Cross-linguistically, these formatives may be independent words but may also be affixes or even a certain change in word form. Four types of languages can be predicted in terms of the presence (+) or absence (-) of these two formatives.

- | | |
|----------------------------|---------------------------|
| I. Linker (-) Copula (-) | II. Linker (+) Copula (-) |
| III. Linker (-) Copula (+) | IV. Linker (+) Copula (+) |

Language is constrained by the principle of linearity, whereby linguistic elements must be arranged in the linear order. Therefore, every language has ways to mark continuity and discontinuity in terms of phrase structure. Many languages have a rigid constituent order whereby continuity and discontinuity can be distinguished. From the perspective of information structure, some languages have a system in which constituent order is rigid in terms of restrictiveness or in which articles play a role in marking continuity and discontinuity.

Keywords: copula, linguistic typology, word order, ezafe, referentiality (specificity)

1. Introduction
2. Notes on previous works
3. A cross-linguistic examination of the form of the copula sentence and the modifying structure
4. Conclusion

1. Introduction

This paper is a cross-linguistic research on copula constructions in diverse languages, focusing on their forms and their correlation to the typological characteristics of the languages examined. I will show that what characterizes copula constructions is their exocentric nature, as opposed to endocentric nature found in the modifier-modified relationship (as in noun phrases).

It is useful here to give a definition of the copula construction. A copula construction is a construction in which the predicate is a noun or nominal phrase in general. Thus, the copula construction in my definition here equals to a nominal predicate construction. A construction may be a copula construction without respect to whether it contains a copula word, even though this way of defining a copula construction has its problems, which will be noted in sections below.

2. Notes on previous works

2.1. Previous works on copula constructions

Among many works on copula constructions and nominal predicate constructions, I only focus on three which deal with typologically various types of

languages: Hegeveld (1992), Pustet (2003), and Nishiyama (2003).

Hegeveld (1992) examines 37 languages which are genetically and geographically not homogeneous, focusing on non-verbal predication (including existentials) found in these languages.

The definition of the copula is, according to Hengeveld, an element which is 'semantically empty'. Hence, the words like *become* and *remain* constitute a subclass of copula, while the words like *seem* do not. Hengeveld also examines the historical development of copula words and their grammaticalisation processes.

Pustet (2003) is another cross-linguistic survey of copula constructions. Pustet (2003: 5) follows Hengeveld (1992: 32) in defining the copula, i.e., as the auxiliary word without semantic content. He examined more than 160 languages, focusing on the relationships between the word class and lexical meaning of the predicate on the one hand and occurrence of the copula on the other. Pustet suggests a hierarchy based on which the relationship between the word class of the predicate and occurrence of a copula is explained: some languages (such as Tagalog) do not require a copula without respect to whether the predicate is a verb or a noun or a word of any other word class, while others (like Bambara) employ a copula for verbal as well as nominal predicates. With respect to the relationship between the lexical meaning of the predicate and occurrence of the copula, Pustet suggests four semantic features, i.e. dynamicity, transience, transitivity, and dependency, and examines presence or absence of the copula in these types of predicates. He also claims that the historical sources of copulas are typically verbs and pronouns.

Nishiyama (2003) focuses on the semantics of copulas. In classifying copula constructions, he pays special attention to the referentiality of a noun phrase and the saturability of the head noun. According to his typology, the Japanese copula construction *A wa B da* 'A is B' (NB: A is marked by topic *wa*) is classified into five semantic types¹: predicational sentence, inverted specifical sentence, inverted identifying sentence, inverted identity statement, and definitional sentence. On the other hand, another copula construction *A ga B da* 'A is B' (NB: A is marked by nominative *ga*) is classified into four semantic types: specifical sentence, identifying sentence, identity statement, and presentational sentence.

Frajzyngier et al. (2002) is an important contribution to the literature of the copula construction, and presents problematic assumptions which will be discussed further in this paper. Focusing on modifying function (e.g. the man's dog) and predicative function (e.g. this is a dog), they examine which of these functions is coded as juxtaposition of two noun phrases in different languages. Their hypothesis is that, since employing the juxtaposition strategy in both functions would result in ambiguity, a language must employ the juxtaposition strategy for only one of the two functions. They first examine eight Chadic languages (the underlined languages below), then go on to examine twenty five languages which are geographically and genetically not homogenous. In conclusion, Frajzyngier et al. (2002) propose a dichotomic typology in which all languages examined in their study are argued to fall into either of the following two types (asterisk indicates that further examination of examples is required):

- Languages in which juxtaposition is only used for modification, and which have a

¹ In Japanese linguistics, there has been a traditional tripartite classification of copula constructions: specifical sentence, predicational sentence, and the so-called 'eel' sentence (a typical example of which is *boku wa unagi da* (1sg top eel cop) '(Speaking of) me, (I would choose) eel.'). In Western linguistics, on the other hand, a bipartite classification of copula constructions are common (see Section 2.3)

copula word for the present tense equational clauses:

French, English, Mupun, Gidar, Lele, Drehu, Mongolian, Quechua*, Siroi, and Limbu

• Languages in which juxtaposition is only used for predication, and which do not have a copula word for the present tense equational clauses:

Hausa*, Miya, Mina, Hdi, East Dangla, Kanuri, Lango, Kisi, Athpare, Chinese*, Classical Chinese*, Japanese, Spoken Turkish, Yidiny*, Gooniyandi*, Kayardild, Tümpisa*, Lavukaleve, Written Polish, Spoken Polish*, Kilivila, and Pipil.

Drawing on the fact that Chadic languages belong to both types and that both types are found in diverse geographic areas, they claim that this typology is not geographically and genetically biased.

2.2. Problems of previous works and claims of the present study

The present study argues against several claims and assumptions made by Frajzyngier et al. (2002).

First, the 'modifying function' in their sense is more precisely possessive function. Although this may be intended to focus on two NPs and their relationship, it must be pointed out that possessive relation and predicative relation are substantially different in semantic nature. If, on the other hand, we take into account adjectival modification, the two functions are reasonably similar:

[<i>the [white flower]</i>]	vs.	[<i>the flower</i>] [<i>is white</i>]
(modifying function)		(predicative function).

These two have been traditionally commonly referred to as endocentric and exocentric structures in Bloomfield's sense, or nexus and junction in Jespersen's sense. And it is this kind of semantic contrast that has been a central issue of linguistics. For this reason, the present paper focuses on adjectival modification and predication and their formal encodings².

The second problem of Frajzyngier et al. (2002) is their assumption that there is no language which employs a juxtaposition strategy for both modifying and predicative functions. In fact, there are such languages, as will be discussed later in this paper. Frajzyngier et al. (2002) argue that employing juxtaposition for both functions results in ambiguity and is therefore avoided in language. However, ambiguity can be avoided, if not completely, by several means to be noted in the following paragraph.

First, juxtaposing two NPs does not necessarily fail to mark grammatical relation of the two. For example, fixed word order may mark grammatical relation, as is the case in many SVO languages where word order marks grammatical relation of the NPs which are simply juxtaposed. Prosody may also help differentiate the functions of an otherwise identical structure.

Second, restrictiveness (or 'referentiality' in Nishiyama's 2003 terms) may play a role in disambiguation. Restrictiveness here refers to a notion similar to definiteness. As an illustration, let us examine the following two examples of English copula sentences:

² It is admitted that this method also has its problem: in languages where adjectival predication is expressed with a verbal rather than nominal strategy, the predication has nothing to do with a copula construction. However, the present study deals with such languages as well as languages in which adjectival predication involves a copula construction.

Mary's husband is a teacher (proper inclusion) and *Mary's husband is John* (equation). Here, transposition of the copula subject and the copula complement results in an ungrammatical sentence in the case of proper inclusion (**A teacher is Mary's husband*), but not in the case of equation (*John is Mary's husband*). This ungrammaticality is explained as follows (see also Nishiyama 2003): A and B in the copula sentence A is B must have a semantic relationship shown in either Figure 1 or 2.

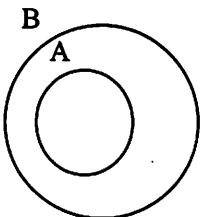


Figure 1

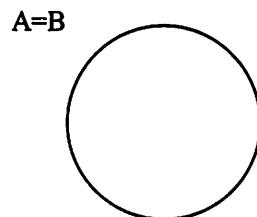


Figure 2

Now, let us consider two juxtaposed structures in an hypothetical language in which a copula word is not employed: (a) *the white flower* and (b) *the flower white*. These structures are not ambiguous if the language has a general syntactic rule whereby the preceding element modifies the following one, allowing us to interpret that (a) is modification and (b) is predication. It may still be argued that (a) can be interpreted as predication like '*White is the flower*' (emphasis) and (b) as modification like '*flower's whiteness*'. However, these interpretations should be marked, and the default interpretations must be that (a) is modifying and (b) predicative. In (b), the definite article *the* plays an important role in controlling the restrictiveness of the word *flower*.

In what follows, I will show how word order and restrictiveness play a role in several languages.

3. A cross-linguistic examination of the form of the copula sentence and the modifying structure

Here I introduce two cover-terms for further discussion below: linker, which is used to construct an endocentric structure, and copula, which is used to construct an exocentric structure. Cross-linguistically, these formatives may be independent words but may also be affixes or even a certain change in word form. Four types of languages can be predicted in terms of the presence (+) or absence (-) of these two formatives.

- | | | |
|------|-------------|------------|
| I. | Linker (-), | Copula (-) |
| II. | Linker (+), | Copula (-) |
| III. | Linker (-), | Copula (+) |
| IV. | Linker (+), | Copula (+) |

Particularly interesting are Type II and III languages. Our prediction is that in an absolute Type II language, where the presence of a linker marks an endocentric

structure (thus encoding modification), simple juxtaposition will automatically mark an exocentric structure (thus encoding predication). By contrast, in an absolute Type III language, where the presence of a copula marks an exocentric structure, simple juxtaposition will automatically mark an endocentric structure. These should be, of course, two extremes in a continuum on which intermediate cases exist, where the presence of a linker and/or a copula is conditioned by complex factors in the same language.

In what follows, each type of language listed above will be examined in detail. As noted earlier, we will focus on the relationship between an adjectival word and a noun.

3.1. Type I language (- copula, - linker)

- Mongolian: This language has a constituent order of modifier-modified, which means that simple juxtaposition can be interpreted as an endocentric structure. The adjectival word behaves like a nominal in this language.

(1) a. cənxər təngər
blue sky

b. təngər cənxər (baj-dag) .
sky blue COP-HABIT

In (1b), the only interpretation is that of predication and not of modification like ‘blueness of the sky’, due to the intonation superimposed on the words in question.

The following pair of examples shows that, whereas (2a) is ambiguous because of the word order and the problem of word class system (demonstratives and adjectives are subclass of the nominals), (2b) is not due to the fact that a demonstrative word is directly followed by a subject noun.

Mongolian thus supports my hypothesis argued for in Section 2. This language does have a copula³ which appears in the past tense and/or in the negative. A nominal modifier is marked with the genitive case. According to a native speaker of Mongolian who I consulted with, intonation plays a role in interpreting examples like (2).

Frajzyngier et al. (2002) categorizes Mongolian as a language which employs juxtaposition for modification but not for predication and which has a copula word. They treat quantifying constructions such as *xun buugədə* (man/person total number) 'all people' as examples of juxtaposition, and analyze *yum* (thing) and *məŋ* (just like that) as copula words. Unfortunately, these analyses are based on atypical structures, so that these

³ There are many languages which do not have a copula word. For example, Pustet (2003) points out that 41 out of 131 languages (31.5 %) do not have a copula word.

analyses cannot be argued to reflect the linguistic facts of Mongolian.

It is admitted, however, that Mongolian has several forms that function like a copula, such as *bol*, which derived from *bol-bol* (become-CONDITIONAL), and *n'*, which derived from the third person pronoun. These copula-like words are similar to Japanese *wa* (Kazama 2003).

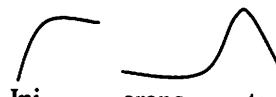
• Indonesian: In contrast to Mongolian, this language has the constituent order of modified-modifier. Hence, juxtaposition can mark modification. The following examples illustrate this (Ushie 1975).

- (3) Hal penting
thing significant
'significant thing'
- (4) Orang muda itu
person young that
'That young person'

Demonstrative or personal pronoun comes at the outmost of the modifying structure, thus marking the edge of this structure. This means that, if a noun phrase follows these pronouns, then the adjacent two noun phrases form an exocentric (i.e. predicative) structure. This Indonesian fact shows how a demonstrative pronoun serves as a historical source of the copula word. In addition to the edge marking function of pronouns, intonation serves as an important marker for this purpose.

- (5) Hal ini penting
thing this significant
'This is a significant thing' (NOT 'This significant thing')
- (6) Orang itu muda
person that young
'That person is young.' (NOT 'That young person')

- (7) a. Pulpen ini
fountain.pen this
'This fountain pen'
- b. Ini pulpen.
this fountain.pen
'This is a fountain pen.' (NOT 'This fountain pen')

- (8) 
Ini orang utan.
this Orangutan
'This is an orangutan.' (Intonation based on Furihata (2005: 36))

As mentioned above, Indonesian has the constituent order of modified-modifier,

so that any number of modifiers can be put after the noun modified. Thus, it might appear that this language needs some kind of edge marking that marks the end of the noun phrase, such as a copula word. However, it is now clear that such a device is not necessarily crucial: that is, the restrictiveness of a noun phrase fed by a demonstrative pronoun does serve as an edge marker in relation to another noun phrase that follows.

If an adjective is preceded by a pronoun, these two do not form a single noun phrase, as shown in (9b).

- (9) a. Ini penting.
this significant
'This is significant.'
- b. Penting ini.
significant this
'This is significant.'

If a noun is followed by an adverb, they do not form a single phrase, thus resulting in an exocentric structure. Indeed, an adverb is a common historical source of the copula word cross-linguistically.

- (10) Air sangat jernih
water very be.clear
'The water is very clear.'
- (11) Besi amat keras
Iron very hard
'Iron is very hard.'

Thus, the part of speech of a word that is adjacent to a noun is also an important aid in the distinction between endocentric and exocentric structuring.

In Indonesian the copula is absent in principle, even though there are emphatic copular forms like *adalah* ('have' + the emphatic morpheme *-lah*) and *ialah* ('3SG' + the emphatic morpheme *-lah*). On the other hand, a relative word *yang* connects two words belonging to different phrases into one noun phrase, thus marking an endocentric structure. Indonesian does not have articles; however, it has a third person suffix *-nya*, which functions crucially in forming phrases and clauses.

In summary, languages with no copula or linker have a fixed constituent order for the modifying structure, and in these languages restrictiveness and intonation play a significant role in marking a copula sentence.

3.2. Type II language (+ linker, - copula)

• Tagalog: In an endocentric structure the linker connects two words. Word order here is free. There is no copula word in this language. In a predicative structure the word order is VS. The following examples (taken from Moriguchi 1985: 40) illustrate these statements.

- (12) magandá-ng babae
beautiful-LK woman
'beautiful woman'

Since word order is free for the modifying structure, a linker must be used to mark which is modifier and which is modified. Also, it is noted that the article is crucial in forming an exocentric structure given that there is no copula in this language.

- (Egyptian) Arabic: the constituent order for the modifying structure is modified-modifier. Also, the definite article within the modifying structure agrees with the head noun. The article is pronounced with the preceding word (the following examples come from Koike 1998).

- (15) wálad kíbí'r
 boy big
 'big boy'

(16) el -wálad _el -kíbí'r
 the boy the big
 'the big boy'

(17) el -wálad _el -kíbí'r kaslā'n
 the boy the big lazy
 'The big boy is a lazy person.'

(18) el -wálad _el -kíbí'r _el -kaslā'n kaddā'b
 the boy the big the lazy.person liar
 'The big lazy boy is a liar'

Unlike Indonesian and Tagalog, where an article marks an exocentric structure, Arabic employs an article to mark an endocentric (modifying) structure, and the absence of an article marks discontinuity of the phrase (thus an exocentric structure). In this language third person pronouns function like a copula (Matsuda 1988: 471). The modifying structure in this kind of language is appositional in nature.

- Russian: Agreement in gender, number, and case is observed in an endocentric structure. Word order is remarkably flexible, so that it is impossible for word order to mark which is modifier and which is modified. There is no article, which means that it is impossible to have recourse to the manipulation of restrictiveness in order to mark (dis)continuity of constituents. The following examples are from Nakazawa (1991, transcription SK).

- | | | |
|------------------|-------------|---------------|
| (19) <i>odin</i> | <i>moj</i> | <i>dobryj</i> |
| one.man.SG.M.NOM | my.SG.M.NOM | kind.SG.M.NOM |

drug	uzna-l	eto.
male.friend.SG.NOM	realized-SG.M.	that
(20) odna	moja	dobraja
one.SG.F.NOM	my.SG.F.NOM	kind.SG.F.NOM
podruga	uzna-la	eto.
female.friend.SG.NOM	realized-SG.F.NOM	that

Agreement thus helps mark the modifying relationship being established between constituents, which may not necessarily be contiguous.

No copula is used in the present tense. Adjectives, however, have a predicative form (short form), which indicates the predicative function of the adjective thus marked. Intonation also plays an important role in the copula construction.

In summary, compensating for the lack of articles (and other devices to control restrictiveness) or a copula or a linker, agreement plays a crucial role in marking the modifying structure. As in the case of Arabic, Russian is a language in which each word shows mutual independence and juxtaposition of words mark an exocentric structure. This kind of language, however, has a system whereby agreement marks apposition and thus an endocentric structure.

• Chinese: This language has the constituent order of modifier-modified. In the endocentric structure the word *de* is used. In a copula clause the copula *shi* may be used but optional⁴. Copula-less juxtaposition is usually used in the casual spoken language. According to a native speaker of Chinese whom I consulted with, the copula may be omitted unless ambiguity (with possessive constructions, etc.) arises.

It might first appear that Chinese is categorized as a language with a linker but not with a copula. However, *de* in the endocentric structure does not mark the modifying word but nominalizes the preceding word, with the two words linked by *de* being appositional. Toudou and Aihara (1985) explain this as follows:

Consider the following examples: 鲜红的花 *xian1 hong2 de hua* 最新的书 *zui4 xin1 de shu1* The first expression denotes that among various flowers with various colors, the particular flower that the speaker is looking at belongs to a class of flowers which are red. Likewise, the second expression denotes that among various books which may be old or new, the particular book that the speaker is looking at belongs to a class of books which are newest. As we explained earlier, the word *de* is a nominalizer. Thus, the expressions “鲜红的 *xian1 hong2 de*” and “最新的 *zui4 xin1 de*” are nominalized expressions, which denote the membership of the larger groups denoted by the nouns that follow (i.e. “花 *hua1*” and “书 *shu1*”). This convention is not uncommon in Chinese. For example, “松 *song1*” (pine) followed by “树 *shu4*” (tree) becomes “松树 *song1 shu4*” (pine tree), and “鲤 *li3*” (carp) followed by “鱼 *yu2*” (fish) becomes “鲤鱼 *li3 yu2*” (carp). Each of these examples comprises the smaller group and the larger group. It is generally agreed that the word “的 *de*” designates proper inclusion, but this word only contributes to nominalization, and it is the general convention of the modifying relationship mentioned above that marks the modification between the two elements.

⁴ This morpheme historically derived from an anaphoric pronoun, as pointed out by a number of studies (see, for example, Li and Thompson 1977 and Hengeveld 1992).

Hence, word order is actually crucial in the modifying structure, in which the more restricted element precedes the less restricted one.

In this language, a numeral does not function as a modifier, as is expected from the above-mentioned account.

- (21) 一个人 yi2 ge ren *一个的人 yi2 ge de ren
 one CLF person
 ‘one person’

(22) 很多学生 hen3 duo1 xue2sheng *很多的学生 hen3 duo1 de xue2sheng
 HEN many student
 ‘many students’

This is parallel to the restriction found in Japanese, whereby **ooi hito* (be.numerous person) 'intended meaning) many people' is ungrammatical. Unlike, say, *ookii hito* (be.big person), which denotes a particular person belonging to a class of people who are big, the ungrammatical expression does not denote particular people belonging to a group of people who are numerous. Linguistic typology tends to assume that the combination of a numeral and a noun is modification by the former of the latter just like the combination of an adjective and a noun; however, this is not necessarily valid cross-linguistically.

There is severe restriction on the predicative use of adjectives in Chinese. That is, a degree adverb is usually required before the predicative adjective. The most common adverb of this kind is *hen3*, which does not really function as a degree adverb in this context.

- (23) 这 个 孩子 很 可爱。
 zhei4 ge hai2zi hen3 ke3ai4
 this CLF child HEN be.lovely
 'This child is lovely.'

Without a degree adverb, the statement will be interpreted as a general statement or as a contrastive statement. Chinese is often referred to as a topic-prominent language (Li and Thompson 1976), which means that the sentence-initial phrase is interpreted as the topic of the sentence. Thus, unless this phrase is not actualized in some way or another, the statement is interpreted as a general or contrastive statement.

- (24) 这 本 书 好, 还是 那 本 书 好?
 zhei4 ben3 shu1 hao3 hai2shi nei4 ben3 shu1 hao3
 this CLF book be.good or that CLF book be.good
 'Which is better, this book or that book?'

这 本 书 好。
this CLF book good
'This book is better.'

Hence, the adverb *hen* functions like a copula, in that it breaks up two successive

noun phrases. It is a diachronic possibility, then, that a degree adverb will be reanalyzed as a copula.

In general, Chinese has a strict system of marking restrictiveness, in such a way that a concrete notion is marked by demonstratives or quantifiers. Therefore, the distinction between the endocentric and exocentric structures is made through the marking of restrictiveness and word order.

3.3. Type III languages (- linker, + copula)

• English: This language mainly has the constituent order of modifier-modified, and uses a copula word in the exocentric structure.

(25) An/The interesting book ⇔ The book is interesting.

However, if the modifier consists of two or more words, the constituent order is of modified-modifier. In this language, number agreement is partially observed (e.g. the alternation of demonstratives such as *this/these*).

Three devices play an important role in marking the distinction between the endocentric and exocentric structures: the article, which comes to phrase-initial position, the obligatory copula, and the constituent order of modifier-modified.

3.4. Type IV languages (+ linker, + copula)

Languages such as many Native North Amerindian languages, Korean, Nivkh, and Ainu have an inflecting verb-like adjective⁵. In such languages the adjective is usually inflected for grammatical categories like tense and/or person. These languages have a copula by our definition of the copula. The endocentric structure involves a kind of relativization, which means that these languages have a linker as well.

In this section I first examine Persian, in which the adjective is nominal, then go on to examine Japanese, which has a verbal adjective. Finally I will examine Eskimo and Koryak, two polysynthetic languages in which the adjectival morpheme is expressed as a bound morpheme.

• Persian: In the endocentric structure, *ezafe*, a kind of linker, is used. Modified-modifier is the unmarked constituent order. The following examples come from Kamei, Kono, and Chino (eds., 1996: 116) and Okazaki (1982: 31).

(26) xāne-ye bozorg-e qadīmī-ye qermez
house-EZF big-EZF old-EZF red
'A big, old, and red house'

(27) īn gol sorx ast.
this flower red be.3SG
'This flower is red.'

(28) īn gol-e sorx ast.

⁵ See Matsumoto (2007), which is a detailed cross-linguistic survey of whether adjectival words are nominal or verbal.

this	flower-EZF	red	be.3SG
‘This is a red flower.’			

Comparison between (27) and (28) shows that the absence of *ezafe* is a positive marker of discontinuity between a noun and the adjective that follows. This language, however, does have a copula word.

- Japanese: In the endocentric structure the unmarked constituent order is modifier-modified. The adjective as a modifier inflects to mark whether it is followed by a noun or another adjective or nothing.

(29) <i>huru-kute</i>	<i>aka-kute</i>	<i>ooki-na</i>	<i>ie</i>
be.old-SEQ	be.red-SEQ	be.big-ADN	house
‘An old, red, and big house’			

In the exocentric structure, the topic marker *wa* and the auxiliary verb *desu* also function like a copula together with the *-i* inflection of the predicative adjective

(30) <i>sono</i>	<i>ie</i>	<i>wa</i>	<i>huru-kute</i>	<i>ooki-i.</i>
that	house	TOP	be.old-SEQ	be.big-NPST
‘That house is old and big.’				

Thus, adjectives in Japanese inflect, and are subsumed under the class called “yougen” (inflecting words) together with verbs.

In languages like Japanese, dependency (i.e. whether the predicate is in a main clause or in a dependent clause) as well as tense and mood are inflectional categories. Together with the rigid constituent order of modifier-modified, this inflectional marking helps distinguish between the endocentric and exocentric structure.

The copulas *da/desu* (or their variant forms) show dialectal variation in terms of its obligatoriliness and co-occurrence restriction. Thus in the Kyushu dialects the copula is not used at all, whereas in dialects of central Japan such as the Nagano dialect, the copula may follow a verb. On the other hand, it is safe to say that the topic marker *wa* or other forms with the same function exist in all dialects. In Japanese linguistics, it has often been pointed out that the topic marker *wa* functions to divide and combine elements (Nishiyama 2003: 357), and this dual function is evident in the copulas in different languages examined so far. To be more precise, the function of the copula is to syntactically divide elements and to combine elements in information-structure.

Japanese has other forms that function like the topic marker *wa*: *nara* ‘if, speaking of’ and *tte* ‘speaking of’. The functional difference between these forms is a topic of further research. It is noted that *nara* is historically derived from the conditional form of the verb *naru* ‘become’.

- Eskimo: The following description and examples come from Miyaoka (1978: 64-75), with an orthographic adjustment. In Eskimo, the adjectival word is basically encoded as a verb. A noun is modified by an adjective in the following two ways.

A. Using lexical suffixes: Lexical suffixes are etymologically unrelated to independent

words that have a similar meaning. Thus, in the following examples, the adjectival concept 'big' is expressed as a lexical affix in (31) or as an independent word in (34), but these forms are not similar. The same holds true for the pair (32) and (35) both of which express the concept 'small'.

B. Using analytic expressions with an appositional structure. Here, the appositional head noun and modifying adjective agree in case and number. The adjective is transformed into a noun (with the affix *-lria*), except for 'new (one)' and 'old (one)', which are not subject to nominalization.

[A]

- (31) qayar-pak-ø
kayak-be.big-ABS.SG
'a big kayak'

- (32) qayaa-cuarr-ø
kayak-be.small-ABS.SG
'a small kayak'

[B]

- (33) nutaaraq-ø qayaq-ø
new.one-ABS.SG kayak-ABS.SG
'a new kayak'

According to native speakers of Eskimo, if strategy B is used where strategy A is appropriate, the expression will sound childish or foreign.

- (34) ??ange-lria-ø qayaq-ø
be.big-NLZ-ABS.SG kayak-ABS.SG

- (35) ??mike-lria-ø qayaq-ø
be.small-NLZ-ABS.SG kayak-ABS.SG

In the case of the exocentric structure, the adjective inflects for tense and person just like verbs.

- (36) una qayaq-ø mikku-ø.
this kayak-ABS.SG be.small-3SG.PRS
'This kayak is small.'

A noun is verbalized when it is used as a predicate.

In many languages, some property concepts like 'big' and 'small', 'close' and 'estranged' may be expressed by bound morphemes (e.g. diminutive affixes and augmentative affixes) rather than independent words. These concepts are usually more compatible with nouns than other property concepts.

- Koryak: The following description and examples are based on Kurebito (1997: 10, 16)

and Kurebito (p.c.). This language has two ways of noun modification.

- A. One or more modifying elements (which may be a noun stem, adjective stem, verb stem, or adverb stem) are incorporated as part of the modified noun word.
 - B. An analytic expression in which the modifier and the modified noun (which agree in number) occur in a noun phrase.

[A]

- (37) *mejn-ə-wejem-ti*
big-EPN-river-ABS.DU
'two big rivers'

[B]

- (38) n-ə-mejə-ŋ-qine-t wejem-ti
PRED-EPN-big-EPN-PRED-3DU river-DU
‘two big rivers’

Strategy A might first appear to be similar to the Strategy A of Eskimo. However, whereas in Eskimo the adjectival morpheme is a lexical affix which is etymologically distinct from the corresponding independent word, in Koryak it is an independent word that is compounded with a noun. When an adjective occurs without a modified noun, it is used as a predicate, inflecting for number and person.

3.5. Summary of this section

In this section I will summarize formal characteristics of copulas and linkers in the diverse languages examined above.

In Type I languages, where there is no linker or copula, the constituent order is basically rigid, and intonation plays an important role. In the other types of languages, in which the copula and/or the linker exists, the following generalizations can be made

In languages where the copula is not extensively used, several devices to mark discontinuity are used: personal particle (e.g. Mongolian), demonstrative (e.g. Indonesian), degree adverb (e.g. Chinese), etc. These devices are potential sources for the copula. The concept of restrictiveness is crucial in analyzing why these devices help mark discontinuity.

On the other hand, the linker is found in many languages and in various forms: the linker in Tagalog, Ezafe in Persian, Adnominal in Japanese, agreement of articles in Arabic, gender-number-case agreement in Russian, lexical affixes in Eskimo, and incorporation in Koryak. These various ways of marking noun modification fall into those languages that employ agreement and apposition (Russian and Arabic) and others.

Indonesian and Persian are two extremes in terms of the use of juxtaposition. That is, in Indonesian juxtaposition marks modification, whereas in Persian it marks exocentric structure, and uses ezafe to mark endocentric structure.

4. Conclusion

Language is constrained by the principle of linearity, whereby linguistic elements must be arranged in the linear order. Therefore, every language has ways to mark continuity and discontinuity in terms of phrase structure. Many languages have a rigid constituent order whereby continuity and discontinuity can be distinguished. From the perspective of information structure, some languages have a system in which constituent order is rigid in terms of restrictiveness or in which articles play a role in marking continuity and discontinuity.

Whereas there are vast studies on verbal predicates and verbal categories, there are much less studies on copula sentences and their related phenomena. This is partly because the copula is a subclass of verb in well-known European languages such as English, which leads to overlooking other languages where the copula does not exist or is expressed in other forms. Focusing on the constraint of linearity and the marking of continuity and discontinuity, we need to re-examine this hitherto poorly explored area of linguistic typology. In doing so, it is necessary to pay attention to the modifying structure and the possessive structure, and the place of the copula construction within the entire system of the language.

We also need to pay attention to those languages which have two kinds of copula: Hawaiian (specific/non-specific, see Shibata 1992: 342), Spanish (based on aspectual differences, see Terasaki 1998: 115), Thai and Cambodian (based on referentiality, see Mikami 1985). It is also an important future research topic to examine the topicalization phenomena in various languages.

Abbreviations

1, 2, 3 (1 st person, 2 nd person, 3 rd person)	
ABS(OLUTIVE)	M(ASUCLINE)
ADN(OMINAL)	N(EUTER)
CLF(CLASSIFIER)	NLZ(NOMINALIZER)
COP(ULA)	NOM(NATIVE)
DU(AL)	NPST(NONPAST)
EPN(EPENTHETIC VOWEL)	PL(URAL)
EZF(EZAFE)	PRS(ENT)
F(EMININE)	SEQ(VENTIAL)
HABIT(UAL)	SG(SINGULAR)
LK(LINKER)	TOP(IC)

R e f e r e n c e s

- Frajzyngier, Z., H. Krech, and A. Mirzayan (2002) Motivation for copulas in equational clauses. *Linguistic Typology* 6, 155–198.
- Hengeveld, K. (1992) Non-verbal predication. Theory, typology, diachrony. functional Grammar Series 15. Berlin, New York: Mouton de gruyter.
- Hurihata, M. (2005) *Indonesia go no shikumi*. Tokyo: Hakusuisha.
- Kamei, T., R. Kono and E. Chino eds. (1996) *The Sanseido Encyclopaedia of Linguistics*. vol. 6, Terms. [in Japanese] Tokyo: Sanseido.
- Kazama, S. (2003) Arutai syogengo no 3 guruupu (tyuruku, mongoru, tsungusu) oyobi tyoosengo, nihongo no bunpou wa hontou ni nite irunoka -taisyou bunpou no kokoromi. [Do the three Groups of

- the "Altaic" Languages (Turkic, Mongolic, and Tungusic), as well as Korean, really Resemble Japanese Grammatical Structure? An attempt at a Contrastive Grammar Analysis.] T. Osada and A. Vovin (eds.) *Perspectives on the Origins of the Japanese Language*. Nichibunken Japanese Studies Series, No. 31: 249-340. International Research Center for Japanese Studies.
- Koike, Y. (1998) *Mikka de oboeru arabia go*. Tokyo: Gakuseisya.
- Kurebito, M. (1997) *Koryaku go no meisi no gouseikei to bunsekikei*. [Incorporative Form and Analytical Form of Noun in Koryak.] Bulletin of the Hokkaido Museum of Northern People vol. 6: 9-30. Hokkaido Museum of Northern Peoples.
- Li, C. N. (ed.) (1976) Subject and topic. New York: Academic Press.
- Li, C. N. (ed.) (1977) Mechanisms of syntactic change. Austin: University of Texas Press.
- Li, C. N. and S. A. Thompson (1976) Subject and Topic: a new typology of languages. In Li (1976): 457-489.
- Li, C. N. and S. A. Thompson (1977) A mechanism for the development of copula morphemes, In Li (1977): 419-444.
- Matsuda, I. (1988) *Arabia-go*. [Arabian Language.] T. Kamei, R. Kono and E. Chino (eds.) *The Sanseido Encyclopaedia of Linguistics*. vol. 1: 462-470. [in Japanese] Tokyo: Sanseido.
- Matsumoto, K. (2007) *Sekai gengo no naka no nihongo*. - *Nihongo keitouron no aratana chihei* - [Japanese in the World's Languages. - A new perspective on the Origin of the Japanese Language -] Tokyo: Sanseido.
- Mikami, N. (1985) *Tai go, raosu go, kanbozia go no copula ni tsuite*. [On the copulas of Thai, Lao, and Khmer.] Keio gizyuku daigaku gengo bunka kenkyuzyo kiyou. [Reports of the Keio Institute of Cultural and Linguistic Studies.] vol. 17: 167-181.
- Miyaoka, O. (1978) *Esukimoo no gengo to bunka*. [The Language and Culture of Eskimo.] Tokyo: Kobundo.
- Moriguchi, K. (1985) *Piripino go bunpou*. [The Pilipino Grammar.] Tokyo: Daigaku-syorin.
- Nakazawa, H. (1991) *Hajimete no rosia go*. Tokyo: Kodansha.
- Okazaki, M. (1982) *Kiso perusia go*. [Fundamental Persian.] Tokyo: Daigaku-syorin.
- Pustet, R. (2003) Copulas. Universals in the Categorization of the Lexicon. Oxford Studies in Typology and Linguistic Theory. New York: Oxford University Press.
- Shibata, N. (1992) *Hawai-go*. [Hawaiian Language.] T. Kamei, R. Kono and E. Chino (eds.) *The Sanseido Encyclopaedia of Linguistics*. vol. 3: 341-344. [in Japanese] Tokyo: Sanseido.
- Tamura, S. (1988) *Ainu-go*. [Ainu Language.] T. Kamei, R. Kono and E. Chino (eds.) *The Sanseido Encyclopaedia of Linguistics*. vol. 1: 6-94. [in Japanese] Tokyo: Sanseido.
- Terasaki, H. (1998) *Supein go*. [Spanish Language.] The Institute of Language Research ed. *A Guide to the World's Languages Part I: Europe and America*. 107-122. [in Japanese] Tokyo: Sanseido.
- Toudou M. and S. Aihara (1985) *Shintei tyuugokugo gairon*. [in Japanese] Tokyo: Taishukan syoten.
- Ushie, K. (1975) *Indonesia go no nyuumon*. [Introduction to Indonesian.] Tokyo: Hakusuisha.

Linguistic Typology of the North

Volume 2

Edited by Tokusu Kurebito



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

Linguistic Typology of the North, Volume 2
(ed.) Tokusu Kurebito

ISBN 978-4-86337-085-2

Published: March 2011

Project director: Kurebito, Tokusu

**The project “Typological and Comparative Study on Polysynthesis
of the Languages in Northeastern Asia”**

**Research Institute for Languages and Cultures of Asia and Africa (ILCAA)
Tokyo University of Foreign Studies
3-11-1, Asahi-cho, Fuchu-shi, Tokyo, 183-8534, Japan
URL:<http://www.aa.tufts.ac.jp/>**

**(c) 2011 Anna Bugaeva, Chikago Ono, Charles Weinstein, Shijiro Kazama,
Megumi Kurebito, Fuyuki Ebata, Tokusu Kurebito**

Printed by Nihon Routo Pringting Co. Ltd.

**The research behind this publication was financially supported by Grant-in Aid for
Scientific Research (B) from the Japanese Minister of Education, Science, Sport
and Culture for the project “Typological and Comparative Study on Polysynthesis
of the Languages in Northeastern Asia” (#21401022)**