

Notes on nominals in Bahnar, Mandarin and Vietnamese

Tue Trinh

University of Wisconsin at Milwaukee
tuetrinh@alum.mit.edu

Trang Phan

Ton Duc Thang University
phanthihuyentrang@tdtu.edu.vn

Hung Phan

Vietnam Institute of Linguistics
hungphanluong@gmail.com

Abstract

In this paper, we present a set of facts concerning nominal structures which distinguishes between three classifier languages of the East Asian variety, namely Bahnar, Mandarin, and Vietnamese. We then propose an account of these facts. Our account reduces the observations to cross-linguistic differences with respect to the availability of certain syntactic configurations, which in turn are derived from cross-linguistic differences with respect to semantic interpretation and lexical resources.

Keywords: classifiers, demonstratives, argumenthood, definiteness

ISO 639-3 codes: bdq, cmn, vie

1 Observations

The principle and parameter theory seeks to find out (i) what is common to all languages (principles), and (ii) what are the ways in which languages vary (parameters). With respect to (ii), some recent developments can be noted. First, there is an attempt to reduce parametric variation to the morphology of functional items, e.g. to the relative strength of the EPP feature on T or C. Second, there is an attempt to pay more attention to macro-variations between typologically diverse languages, e.g. those between number marking and classifier languages.

Needless to say, this list should not be exhaustive. In principle, we expect languages to differ in several other respects, and there is no reason for the study of linguistic variation to be so constrained as just described. In particular, languages can vary on the basis of (i) differences in the *semantics* of functional items and/or (ii) differences in *lexical resource*, meaning one language can have a lexeme realizing some particular meaning with another language having no lexeme realizing that same meaning. Thus, the study of linguistic variation can benefit from looking at micro variations within one language type. This paper is an instance of that enterprise.

Bahnar, Mandarin, and Vietnamese are all “classifier” languages of the East Asian variety: nouns can only combine with numerals through the mediation of a functional, or quasi-functional, morpheme which is called the “classifier.” Here is an example from Vietnamese to illustrate.

- (1) a. *John có một con chó*
John have one CL dog.
'John has one dog.'
- b. **John có một chó*
John have one dog.
('John has one dog.')

Regarding demonstratives, in Mandarin and Vietnamese, a demonstrative requires a classifier but does not require a numeral, as shown in (2b) and (2c), respectively. In Bahnar, on the other hand, a demonstrative requires both a classifier and a numeral, as shown in (2a).

- (2) a. **(ʔbal) tɔ kɔ ʔnej*
two CLF dog DEM
b. *Nei (liang) zhi gou*
DEM two CLF dog
c. *(Hai) con chó đố*
two CLF dog DEM

Regarding argumenthood, in Vietnamese, bare classifier phrases, i.e. those of the form CLP-NP, can be arguments, as shown in (3c). In contrast, this does not hold for Bahnar and Mandarin, as shown in (3a) and (3b), respectively.¹

- (3) a. **tɔʔ kɔʔ waʔ kwa tʃɔŋ*
CLF dog want cross road.
b. **Zhi gou yao guo malu*
CLF dog want cross road.
c. *Con chó muốn sang đường*
CLF dog want cross road.

Regarding definiteness, in Bahnar and Vietnamese, bare numeral phrases, i.e. those of the form Num-CLF-NP, can be definite, as shown in (4a) and (4c), respectively. This does not hold for Mandarin, as shown in (4b).

- (4) a. *ʔbal tɔʔ kɔʔ waʔ kwa tʃɔŋ*
Two CLF dog want cross road.
'The two dogs want to cross the road'
b. **Liang zhi gou yao guo malu*
Two CLF dog want cross road.
(‘The two dogs want to cross the road’)
c. *Hai con chó muốn sang đường*
Two CLF dog want cross road.
'The two dogs want to cross the road'

Also regarding definiteness, in Bahnar and Mandarin, bare nouns can be definite, as shown in (5a) and (5b), respectively. In Vietnamese, bare nouns cannot be definite, as shown in (5c).

- (5) a. *kɔʔ waʔ kwa tʃɔŋ*
Dog want cross road.
'The dog(s) want(s) to cross the road'
b. *Gou yao guo malu*
Dog want cross road
(‘The dog(s) want(s) to cross the road’)
c. *Chó muốn sang đường*
Dog want cross road.
'Dogs want to cross roads' / *‘The dog(s) want(s) to cross the road’

Table (6) summarizes the cross-linguistic variations between Bahnar, Mandarin and Vietnamese which we have just discussed. Note that the observations about Mandarin which we presented above are taken from Cheng and Sybesma (1999).

(6)	Bahnar	Mandarin	Vietnamese
Whether DEM can combine directly with CL-NP	No	Yes	Yes
Whether CLF-NP can be arguments	No	No	Yes
Whether NUM-CLF-NP can be definite	Yes	No	Yes
Whether bare NP can be definite	Yes	Yes	No

¹ We will disregard phrases in the object position due to complications that we think are orthogonal to the questions at hand (cf. Trinh and Sudo 2009.).

2 Analysis

Our analysis of the facts presented above follows in large part the account proposed by Trinh (2011) for the differences between Mandarin and Vietnamese, which is itself based on the proposal made in Chierchia (1998).

2.1 The meaning of nouns, numerals, and classifiers

We assume that bare nouns in number marking languages denote atomic predicates, i.e. sets of singularities. In contrast, bare nouns in classifier languages denote cumulative predicates, sets of both singularities and pluralities. Suppose a, b and c are the only dogs in world w , then the English word **dog** has the meaning in (7a), while its Vietnamese counterpart, **chó**, has the meaning in (7b).

- (7) a. $[[\mathbf{dog}]^w = [\lambda x [x \text{ is a dog}]] = \{a, b, c\}$
 b. $[[\mathbf{chó}]^w = [\lambda x [x \text{ is a dog or a plurality of dogs}]] = \{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$

Let ' $x \sqsubset y$ ' mean that x is a proper part of y and ' $x \sqsubseteq y$ ' mean that x is a proper part of y or identical to y . We define a function sup which maps a predicate P to the "supremum" of P , i.e. that entity which has all and only members of P as part, and a function at which maps a predicate to the set of singularities, i.e. 'atoms,' in it.

- (8) a. $\forall x [x \in P \leftrightarrow x \sqsubseteq sup(P)]$
 b. $\forall x [x \in at(P) \leftrightarrow [x \in P \wedge \neg \exists y [y \in P \wedge y \sqsubset x]]]$

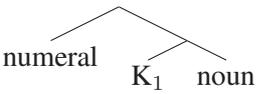
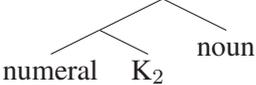
Following Ionin and Matushansky (2006), we assume that counting requires "uniformity": only individuals with the same number of atomic parts can be counted.²

- (9) $[[\mathbf{two}]^w(P) = \frac{[\lambda x [\exists y [y \in \wp(P) \wedge |y| = 2 \wedge x = sup(y)]]]}{\exists n [\forall x [P(x) \rightarrow |x| = n]]}]$

This means numerals cannot combine with bare nouns in classifier languages. Mediation of the classifier is required. We define two types of classifiers, K_1 and K_2 .

- (10) a. $[[K_1]^w = [\lambda P [at(P)]]]$
 b. $[[K_2]^w = [\lambda n [\lambda P [n(at(P))]]]]$

As we can see, the meaning of K_2 include that of K_1 . Specifically, $[[K_2]]$ maps a numeral n and a predicate P to $n([[K_1]](P))$. This means we have two different bracketings for numeral phrases.

- (11) a. 
 b. 

2.2 Three lexemes and an preference principle

We define a silent morpheme **THE** which has roughly the same meaning as English **the**. Specifically, $[[\mathbf{THE}]^w$ maps a predicate P to the "maximal" entity in P if there is one, undefined otherwise. This captures both the existence and uniqueness presuppositions of definiteness (cf. Heim 1991).

- (12) $[[\mathbf{THE}]^w(P) = \frac{sup(P)}{sup(P) \in P}$

² We write " $\wp(P)$ " for the power set of P . Note that we use the "fraction notation" in our semantic definitions, with the numerator representing the assertive and the denominator representing the presuppositional content of the relevant expression.

Among the individuals in the universe of discourse, there are kinds, which are functions from properties to individual concepts (cf. Chierchia and Turner 1988, Chierchia 1998). We define a silent morpheme KIND which turns nouns into names of kinds.³

$$(13) \quad \llbracket \text{KIND NP} \rrbracket^w = \frac{[\lambda w \text{ sup}(\llbracket \text{NP} \rrbracket^w)]}{\llbracket \text{NP} \rrbracket^w = \oplus \llbracket \text{NP} \rrbracket^w}$$

Thus, $\llbracket \text{KIND} \rrbracket^w$ maps each cumulative predicate P into the function from each world w to $\llbracket \text{THE NP} \rrbracket^w$. Finally, we define a silent morpheme EXT which maps kinds into the plurality which instantiate them in each world.

$$(14) \quad \llbracket \text{EXT KIND NP} \rrbracket^w = \text{sup}(\llbracket \text{KIND NP} \rrbracket^w(w))$$

We now propose that Bahnar, Mandarin, and Vietnamese differ with respect to lexical resource in the following way: (i) Bahnar has K_2 but not K_1 , while the opposite holds for Mandarin and Vietnamese; (ii) Bahnar and Vietnamese have THE but Mandarin does not; (iii) all three languages have KIND; (iv) Bahnar and Mandarin have EXT but Vietnamese does not. The table in (15) summarizes this cross-linguistic distribution of the relevant lexemes.

(15)	Bahnar	Mandarin	Vietnamese
Whether the lexicon contains K_1	No	Yes	Yes
Whether the lexicon contains K_2	Yes	No	No
Whether the lexicon contains THE	Yes	No	Yes
Whether the lexicon contains KIND	Yes	Yes	Yes
Whether the lexicon contains EXT	Yes	Yes	No

In addition, we propose the following preference principle. We have no explanation for this principle at this time.

- (16) Preference principle
Prefer KIND to THE!

2.3 Deriving the facts

Given (15), the syntactic structures available in argument positions will be distributed among Bahnar, Mandarin, and Vietnamese in the following way.

(17) Structures available in argument positions	Bahnar	Mandarin	Vietnamese
	No	Yes	Yes
	No	No	Yes
	Yes	No	Yes
	Yes	Yes	No

³ We write “ $\oplus P$ ” to denote the \oplus -closure of P . This means that $\oplus P$ is always a cumulative predicate.

Structure A is not available in Bahnar because Bahnar has K_2 , not K_1 , and K_2 cannot combine directly with a noun. On the other hand, Mandarin and Vietnamese has K_1 , so A is available in these languages. The cross-linguistic distribution of A explains the first line of (6): demonstratives can combine directly with CL-NP in Mandarin and Vietnamese but not in Bahnar.

Structure B is available only in Vietnamese for the simple reason that only Vietnamese has THE and allows classifiers to combine with nouns. This is not true of Bahnar, which does not allow classifiers to combine directly with nouns, and not true of Mandarin, which does not have THE. The cross-linguistic distribution of B explains the second line of (6): only Vietnamese permits CL-NP in argument positions.

Structure C is available in Bahnar and Vietnamese, but not in Mandarin, because only Mandarin lacks THE among these three languages. The cross-linguistic distribution of C explains the third line of (6): NUM-CL-NP can be definite in Bahnar and Vietnamese but not in Mandarin.

Structure D is available in Bahnar and Mandarin but not in Vietnamese, because Bahnar and Mandarin have EXT but Vietnamese does not. The cross-linguistic distribution of D explains the last line of (6): NP can be definite in Bahnar and Mandarin but not in Vietnamese.

References

- Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30: 509–542.
- Chierchia, Genaro & Raymond Turner. 1988. Semantics and property theory. *Linguistics and Philosophy* 11: 261–302.
- Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6: 339–405.
- Heim, Irene. 1991. Artikel und Definitheit. In *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, ed. Arnim von Stechow and Dieter Wunderlich, 487–535. De Gruyter.
- Ionin, Tanya & Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23: 315–360.
- Trinh, Tue. 2011. Nominal reference in two classifier languages. *Proceedings of Sinn und Bedeutung* 15: 629–644.
- Trinh, Tue & Yasutada Sudo. 2009. The Indefiniteness Effect in Mandarin and Vietnamese. Talk given at the Zentrum für Allgemeine Sprachwissenschaft, Berlin .