

# *Squib Notule*

---

## *Deriving four generalizations about nominals in three classifier languages*

TRANG PHAN 

VNU University of Languages and International Studies, Hanoi

[trangphan@vnu.edu.vn](mailto:trangphan@vnu.edu.vn)

TUE TRINH 

Leibniz-Zentrum Allgemeiner Sprachwissenschaft, Berlin

[tuetrinh@alum.mit.edu](mailto:tuetrinh@alum.mit.edu)

and

HUNG PHAN

Vietnam Institute of Linguistics, Hanoi

[hungphanluong@gmail.com](mailto:hungphanluong@gmail.com)

Q1

### *Abstract*

This squib presents a set of facts concerning nominal structures in Bahnar, Mandarin, and Vietnamese. It proposes an account of these facts which reduces them to cross-linguistic differences with respect to the availability of particular syntactic configurations involving the bare noun and its extended projection. These differences, in turn, are derived from cross-linguistic variations with respect to the availability of items in the functional lexicon.

**Keywords:** classifiers, demonstratives, argumenthood, definiteness

### *Résumé*

Cette notule présente un ensemble de faits concernant les structures nominales en bahnar, mandarin et vietnamien. Il propose une explication de ces faits qui les réduit à des différences interlinguistiques par rapport à la disponibilité de configurations syntaxiques particulières impliquant le nom nu et sa projection étendue. Ces différences, à leur tour, sont dérivées de variations interlinguistiques en ce qui concerne la disponibilité d'éléments du lexique fonctionnel.

**Mots-clés:** classificateurs, démonstratifs, statut argument/modificateur, définition

---

This research was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 602.02-2018.300 to the first author, Trang Phan. The second author, Tue Trinh, also acknowledges the support of the ERC Advanced Grant ERC-2017-ADG 787929.

## 1. CLASSIFIER LANGUAGES AND PARAMETRIC VARIATION

One fact about linguistic variation is that nouns which intuitively denote the same concept can have different combinatorial properties in different languages. In English, the noun **dog** can combine directly with the numeral **one**, as in **John has one dog**.<sup>1</sup> In Vietnamese, on the other hand, the noun **chó** ‘dog’ cannot combine directly with the numeral **một** ‘one,’ but requires the mediation of a “classifier” (CL).<sup>2</sup>

- (1) John có    một \*(con) chó  
 John have one    CL    dog  
 ‘John has one/a dog’

This difference between English and Vietnamese is representative of the contrast between “number-marking” languages such as English, French, and German, and “classifier” languages such as Chinese, Vietnamese, and Japanese. It has been noted that variation exists among languages of both types. For example, bare nouns can be definite in Chinese but not in Vietnamese, while classifier-noun combinations can be definite in Vietnamese but not in Chinese (see Cheng and Sybesma 1999, Trinh 2011). Among number-marking languages, some, such as English and German, allow bare plurals to be arguments but others, such as French and Italian, do not. Such facts call for an account of the variation at both the macro level between classifier and number-marking languages, and at the micro level between languages within each group. This requires analysis and comparison of particular languages of both types. A fair amount of work has been devoted to the semantics of nominals in number-marking languages (see Link 1983, Pelletier and Schubert 1989, Barker 1992, Schwarzschild 1992, Krifka 1999, among others). Also, concrete proposals have been made to account for the macro-variation between classifier and number-marking languages, as well as for the micro-variation among the latter (see Krifka 1995; Chierchia 1998, 2010; Dayal 2004). Analyses of classifier languages, however, have been fewer and less explicit, and this is true to an even greater extent for the micro-variation between them. Works in this direction, to the best of our knowledge, tend to be heavily syntactic in nature, with semantic considerations playing a secondary role (see Cheng and Sybesma 1999, 2005; Saito et al. 2008; Wu and Bodomo 2009; Watanabe 2010, among others). This squib is an attempt at balancing the situation. Our objective is to show that given appropriate formalization of certain concepts, several facts about the syntax and semantics of nominals in three classifier languages – Bahnar, Chinese, and Vietnamese – can be made to follow from independently motivated assumptions about the building blocks of semantic representations, as well as plausible hypotheses about linguistic variation.

The general framework we adopt will be the “principles and parameters” theory (see Chomsky 1993, 1995, 1998, 2004). This theory seeks to find out what is common to all languages (i.e. the principles), and what are the ways in which

<sup>1</sup>In text, object language expressions are **boldfaced**.

<sup>2</sup>We follow the standard, albeit quite confusing, practice of using parentheses in examples: ( $\alpha$ ) means the expression is acceptable with or without  $\alpha$ , \*( $\alpha$ ) means it is only acceptable with  $\alpha$ , and (\* $\alpha$ ) means it is only acceptable without  $\alpha$ .

91 languages can vary (i.e. the parameters). An influential view, which is sometimes  
 92 called the “Borer-Chomsky conjecture,” holds that parametric variation is to be  
 93 reduced to the lexicon, in particular the functional lexicon.<sup>3</sup>

94 For example, it has been proposed that whether wh-movement exists depends on  
 95 C (see Huang 1981, 1982), whether V-raising exists depends on T (see Pollock 1989,  
 96 Chomsky 1991), and whether N-raising exists depends on D (see Longobardi 2001,  
 97 Cinque 2005). Of course, there is no a priori reason to assume that functional items  
 98 are restricted to those of categories C, T, and D, or that variation is restricted to the  
 99 ability to trigger movement. The term “functional category” is not definitional, and its  
 100 extension is to be determined based on considerations of empirical adequacy as well  
 101 as theoretical economy and elegance. In the same way, the possibility must be kept  
 102 open that functional categories may differ not only with respect to their featural  
 103 make-up, but also with respect to their availability: the functional lexicon of one  
 104 language may contain a certain item which is absent from the functional lexicon of  
 105 another language (see Manzini and Wexler 1987, Bۆsković and Gajewski 2011).  
 106 In this squib, we will argue that our three-way comparison of Bahnar, Chinese,  
 107 and Vietnamese shows that differences of precisely this kind exist.

108 Before we end this introduction and get to the main discussion, we will  
 109 address the question of the grammatical status of classifiers. While it is quite uncon-  
 110 controversial to assume that elements such as definite articles, demonstratives, or silent  
 111 type-shifting operators are functional items, it is less so with classifiers. As classifiers  
 112 seem to indicate the “class” of the nouns, which is a cognitive notion, one might feel  
 113 that classifiers should be considered substantives. Our assumption in this squib will be  
 114 that they are functional items. We will now provide some justification for this  
 115 assumption, using, without loss of generalization, examples from Vietnamese for  
 116 illustration.<sup>4</sup>

---

118 <sup>3</sup>As far as we know, the term “Borer-Chomsky conjecture” was coined in Baker (2008:  
 119 156), who formulates it thus: “All parameters of variation are attributable to differences in  
 120 the features of particular items (e.g., the functional heads) in the lexicon.” Borer (1984: 3) pro-  
 121 poses a theory which “restricts the availability of variation to the possibilities which are offered  
 122 by one single component: the inflectional component.” In conjunction with the assumption that  
 123 inflection is effected by functional heads, Borer’s thesis amounts to saying that variation is to  
 124 be explained via the functional lexicon. And to quote from Chomsky (2001: 2): “Parametric  
 125 variation is restricted to the lexicon, and insofar as syntactic computation is concerned, to a  
 126 narrow category of morphological properties, primarily inflectional.” This is in the same  
 127 spirit as Borer’s thesis.

128 <sup>4</sup>Note, importantly, that we are talking about classifiers and not measure words such as **herd**,  
 129 **cup**, or **kilogram**. Phrases such as **a herd of cows**, **a cup of milk**, or **a kilogram of meat** express  
 130 measurements which are more or less purpose-related and exist in both classifier languages and  
 131 number-marking languages. Classifier phrases such as **con chó** ‘CL dog’ express “natural  
 132 units” (see Krifka 2003) and constitute the basis for a typological distinction. For more discussion  
 133 on the differences between classifiers and measure words see Her (2012b, a). Note, also, that the  
 134 view that classifiers and measure words should be clearly distinguished, while popular, is not one  
 135 that has not been challenged. For an argument that these two categories are more similar than it  
 appears, see Borer (2005).

136 One criterion for some lexical item to be considered “functional” is that it can be  
 137 omitted without affecting the intended meaning: (2) will be understood as saying the  
 138 same thing as (1), even though it will be perceived as an ungrammatical sentence.<sup>5</sup>

- 139 (2) \*John có một chó  
 140 John have one dog  
 141 (‘John has one/a dog’)

142 Thus, classifiers do not add semantic content to the sentence.<sup>6</sup> Related to this obser-  
 143 vation is the fact that there is a degree of arbitrariness in the relation between a noun  
 144 and the classifier that it requires. Take the classifiers **con** and **cái** in Vietnamese, for  
 145 example. The first typically combines with nouns which denote animals, and the  
 146 second typically combines with nouns which denote inanimate objects. However,  
 147 the noun **thuyền** ‘boat’ may combine with both **con** and **cái**, and there is a strong  
 148 preference, in standard Hanoi dialect at least, for combining the noun **dao** ‘knife’  
 149 with **con** rather than with **cái**.

- 150  
 151 (3) a. John nhìn thấy một con/cái thuyền  
 152 John see one CL boat  
 153 ‘John sees a boat’

---

154  
 155  
 156  
 157 <sup>5</sup>We did not conduct an experiment to test this intuition systematically, but a cursory infor-  
 158 mal survey of ten native speakers – five living in Hanoi and five living in Berlin – shows  
 159 across-the-board agreement that (2) does not sound right and should be corrected to (1). We  
 160 take this to be evidence that (2) is perceived as deviant, while expressing the same content  
 161 as (1).

162 <sup>6</sup>An anonymous reviewer pointed out that this claim needed to be qualified given cases of  
 163 words changing their meaning in accordance with the classifier they combine with. An  
 164 example from Vietnamese is the word **sáo** which means ‘flute’ when combined with the clas-  
 165 sifier **cái** and means ‘starling’ when combined with the classifier **con**. We agree that the claim  
 166 needs to be qualified, or more precisely, clarified. Here is what we say. We take a “word” to be  
 167 a bundle of phonological, syntactic, and semantic properties. This, we believe, is an uncontro-  
 168 versial position. Equally uncontroversial, in our opinion, is the agreement that two different  
 169 words, whether by sheer chance or by historical accident, might have the same pronunciation.  
 170 This means that in the case of **cái sáo** ‘flute’ vs. **con sáo** ‘starling,’ we really have two different  
 171 words with the same pronunciation, each of which requires a different classifier. The argument  
 172 generalizes to homophonous words whose meanings are more closely related. Thus, while clas-  
 173 sifiers do not add semantic content to the sentence, they can disambiguate in the same way that  
 174 pointing at a flute vs. pointing at a starling while uttering the word **sáo** can disambiguate. Now,  
 175 having said this, we admit that reality is more complicated than the picture we just painted. The  
 176 same reviewer brought to our attention cases in Bangla and Assamese where the classifier  
 177 encodes the speaker’s attitude towards the referent of the sister NP. Obviously, appealing to  
 178 homophony to defend the view that classifiers are functional would be unintuitive. Prima  
 179 facie the move would then have to be to say that encoding speaker’s attitude is not “adding  
 180 semantic content” in the sense required of a non-functional item. But we admit to having no  
 satisfactory response to this critique, and agree with the reviewer that the issue is “best left  
 for another venue.”

- 181 b. John cầm một con/?? cái dao  
 182 John hold one CL knife  
 183 ‘John is holding a knife’

184 Thus, the dependency between a noun and its accompanying classifier is, to some  
 185 degree, similar to that between a noun and its grammatical gender in such  
 186 languages as German and French.<sup>7</sup>

187 Another way in which noun-classifier dependency resembles grammatical  
 188 gender is that combining a noun with a “wrong” classifier, just like inflecting a  
 189 noun with a “wrong” gender, results in a grammatical error, not in a different  
 190 meaning. Thus, the classifier to go with **chó** ‘dog’ is **con**, not **cái**, and the grammatical  
 191 gender of **Haus** ‘house’ is neuter, not masculine, but combining **chó** ‘dog’ with the  
 192 classifier **cái**, and **Haus** ‘house’ with a masculine determiner, will still convey the  
 193 intended meaning, even though the sentence is perceived as formally deviant.

- 194 (4) a. \*John có một cái chó  
 195 John have one CL dog  
 196 (‘John has one/a dog’)  
 197 b. \*John hat einen Haus  
 198 John have one.masc house  
 199 (‘John has one/a house’)  
 200

201 Last but not least, classifiers show two properties which have been considered dis-  
 202 tinctive of functional items. First, they are a closed class: while it is imaginable a  
 203 company might invent a new noun to name a new product, it is unimaginable for  
 204 it to invent a new classifier to count items of that new product. Second, classifiers  
 205 has a “world independent” semantics: their denotation stays constant across different  
 206 states of affairs. This will become clear in the discussion below. At this point, we take  
 207 these considerations to be sufficient reasons for assuming that classifiers are func-  
 208 tional items.

## 211 2. FOUR GENERALIZATIONS ABOUT BAHNAR, MANDARIN, AND VIETNAMESE

212 Bahnar and Mandarin are similar to Vietnamese in being “classifier languages” of the  
 213 East Asian variety: nouns can only combine with numerals through the mediation of a  
 214 classifier, as has been illustrated for Vietnamese in the previous section. Let us now  
 215 turn to the discussion of demonstratives, argumenthood, and definiteness in these  
 216 three languages.<sup>8</sup>  
 217

218 <sup>7</sup>Note that functional items indicating grammatical gender can disambiguate in the same  
 219 way classifiers do, as discussed in footnote 6. An example from German is the masculine  
 220 noun **Gehalt**, which means ‘content,’ and the neuter noun **Gehalt**, which means ‘salary.’  
 221 The homophony of these two historically related but synchronically distinct words can be dis-  
 222 ambiguated by the masculine definite article **der** and the neuter definite article **das**.

223 <sup>8</sup>Observations on Vietnamese are based on intuitions of all three authors, who are native  
 224 speakers of this language. Observations on Bahnar are based on field work done by the  
 225 third author. Observations on Mandarin Chinese are based on Cheng and Sybesma (1999).

226 In Mandarin and Vietnamese, a demonstrative requires a classifier but does not  
 227 require a numeral, as shown in (5a) and (5b), respectively.

- 228 (5) a. nei (liang) zhi gou  
 229 DEM two CL dog  
 230 ‘those two dogs’  
 231 b. (hai) con chó đó  
 232 two CL dog DEM  
 233 ‘those two dogs’  
 234

235 In Bahnar, on the other hand, a demonstrative requires both a classifier and a numeral,  
 236 as shown in (6).

- 237 (6) \*(ʔbal) tɔʔ kɔʔ nej  
 238 two CL dog DEM  
 239 ‘those two dogs’  
 240

241 Let us state the first generalization.

- 242 (7) Generalization 1  
 243 DEM can combine with CL-NP in Mandarin and Vietnamese, but not in Bahnar

244 Regarding argumenthood, bare classifier phrases (i.e. those of the form CL-NP)  
 245 can be verbal arguments in Vietnamese, as shown in (8).

- 246 (8) con chó muốn sang đường  
 247 CL dog want cross road  
 248 ‘The dog wants to cross the street’  
 249

250 In contrast, this does not hold for Bahnar and Mandarin, as shown in (9a) and (9b),  
 251 respectively.

- 252 (9) a. \*tɔʔ kɔʔ waʔ kwa tɔrɔŋ  
 253 CL dog want cross road  
 254 b. \*zhi gou yao guo malu  
 255 CL dog want cross road  
 256

257 Let us state the second generalization.

- 258 (10) Generalization 2  
 259 CL-NP can be verbal arguments in Vietnamese but not in Bahnar or Mandarin

260 We take the semantic type of intransitive and transitive verbs to be  $\langle e, t \rangle$  and  $\langle e, \langle e, t \rangle \rangle$ ,  
 261 respectively. For example,  $[[\text{smokes}]] = [\lambda x: x \in D_e \ x \text{ smokes}]$   
 262 and  $[[\text{loves}]] = [\lambda y: y \in D_e \cdot [\lambda x: x \in D_e \cdot x \text{ loves } y]]$ . (see Heim and Kratzer 1998,  
 263 and see footnote 11 below for an explanation of the lambda notation). When we  
 264 say that a nominal is an argument of a verb, what we mean is that the nominal or  
 265 its trace is interpreted as an argument to the function denoted by the verb, which  
 266 entails that only nominals which are of type  $e$  or have traces of type  $e$  can be argu-  
 267 ments. This is the sense in which Chierchia (1998) uses the term “argumental” in  
 268 classifying nominals (see e.g. Chierchia 1998:344). Now, it has been proposed that  
 269 nominals of type  $\langle e, t \rangle$ , when they are in object position, can compose with the  
 270 verb via the rule of Restrict (see Chung and Ladusaw 2004, Trinh and Sudo 2009,



Trinh 2011). What Restrict does can be described informally as making a new verb out of a verb and a nominal. This means that nominals which compose with verbs via Restrict are not verbal arguments, in our terminology. Since subjects cannot compose with verbs via Restrict, the fact that a nominal cannot be subject can be considered evidence that it is not of type *e* (i.e. that it cannot be a verbal argument). Thus, the subject position provides a more reliable diagnostic for argumenthood than the object position. For this reason, we will disregard the object position in our discussion on argumenthood. Note that in other discussions, say one on whether numerals can combine directly with nouns, the position of the relevant nominal will play no role. Thus, examples (1), (2) and (3) all have the nominal in object position. This does not affect anything which we have just said regarding argumenthood.<sup>9</sup>

Regarding definiteness, bare numeral phrases, that is, those of the form Num-CL-NP, can be definite in Bahnar and Vietnamese, as shown in (11a) and (11b), respectively.

- (11) a. ʔbal tɔʔ kɔʔ waʔ kwa tɻɻɔŋ  
 two CL dog want cross road  
 ‘The two dogs want to cross the road.’  
 b. hai con chó muốn sang đường  
 two CL dog want cross road  
 ‘The two dogs want to cross the road.’

This does not hold for Mandarin, as shown in (12).

- (12) \*liang zhi gou yao guo malu  
 two CL dog want cross road  
 (‘The two dogs want to cross the road.’)

Let us state the third generalization.

- (13) Generalization 3  
 Num-CL-NP can be definite in Bahnar and Vietnamese, but not in Mandarin

Also regarding definiteness, bare nouns can be definite in Bahnar and Mandarin, as shown in (14a) and (14b), respectively.

- (14) a. kɔʔ waʔ kwa tɻɻɔŋ  
 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.’

In Vietnamese, however, bare nouns cannot be definite, as shown in (15).

- (15) chó muốn sang đường  
 dog want cross road  
 \*‘The dog(s) want(s) to cross the road.’

---

<sup>9</sup>We thank an anonymous reviewer for drawing our attention to this issue.

	Bahnar	Mandarin	Vietnamese	
DEM can combine directly with CL-NP	No	Yes	Yes	Generalization 1
CL-NP can be arguments	No	No	Yes	Generalization 2
NUM-CL-NP can be definite	Yes	No	Yes	Generalization 3
Bare NP can be definite	Yes	Yes	No	Generalization 4

**Table 1:** Four generalizations about Bahnar, Mandarin, and Vietnamese

Let us state the fourth generalization.

(16) Generalization 4

Bare NP can be definite in Bahnar and Mandarin, but not in Vietnamese

[Table 1](#) summarizes the facts about Bahnar, Mandarin and Vietnamese which we have just discussed.

As we can see, three typologically similar languages can display subtle, intricate and quite puzzling distinctions in distribution and interpretation with respect to the nominal domain. We will propose an account of the four generalizations established above which derives these distinctions from the sort of parametric variations mentioned in the introduction, namely differences in terms of availability of items in the functional lexicon.

### 3. DERIVING THE GENERALIZATIONS

Our analysis of the facts just presented extends the proposal, made in Trinh (2011) for the differences between Mandarin and Vietnamese, to include Bahnar. The definitions below, save that of  $K_2$  in (22b), are taken from that work, barring notational differences which are not substantial. The section will start with some theoretical groundwork, and ends with the derivation of the four generalizations stated above, in the form of informal proofs.

#### 3.1 Theoretical groundwork

This section lays out some assumptions and terms that underlie our analysis. It should be noted that most, if not all, of these assumptions have been motivated elsewhere, in particular in Chierchia (1998) and Trinh (2011) as well as in works cited therein.

##### 3.1.1 Bare nouns

Following Chierchia (1998) and several others, bare nouns will be assumed to denote atomic predicates (i.e. sets of singularities) in number-marking languages, and to denote cumulative predicates (i.e. sets of both singularities and pluralities) in classifier languages. Thus, suppose  $a$ ,  $b$  and  $c$  are the only dogs in world  $w$ , then the English word **dog** denotes, in  $w$ , the set  $\{a, b, c\}$ , while the denotation in  $w$  of its



Vietnamese counterpart, **chó**, is the set  $\{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$ , where  $x \oplus y$  is the plurality consisting of  $x$  and  $y$ .<sup>10</sup> The lexical entries for **dog** and **chó** are given in (17b).<sup>11</sup>

$$(17) \text{ a. } \llbracket \mathbf{dog} \rrbracket^w = [\lambda x. x \text{ is a singular dog}] = \{ a, b, c \}$$

$$\text{ b. } \llbracket \mathbf{chó} \rrbracket^w = [\lambda x. x \text{ is a singular dog or a plurality of dogs}] = \{ a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c \}$$

Let ' $x \sqsubset y$ ' mean  $x$  is a proper part of  $y$  and ' $x \sqsubseteq y$ ' mean that  $x$  is a part of  $y$ , i.e. is a proper part of or identical to  $y$ . Thus,  $a \sqsubset a \oplus b$  and  $a \oplus b \sqsubseteq a \oplus b$ , but  $a \oplus b \not\sqsubseteq a \oplus b$ .

### 3.1.2 Numerals

For the semantics of numerals, the function *sup* is defined as one which maps a predicate  $P$  to the “supremum” of  $P$ , which is to say, that entity which has all and only members of  $P$  as (proper or non-proper) part.

$$(18) x \in \text{sup}(P) \Leftrightarrow_{\text{def}} \forall y(y \in P \leftrightarrow y \sqsubseteq x)$$

Suppose  $P = \{a, b, c, a \oplus b\}$ , then  $\text{sup}(P) = a \oplus b \oplus c$ .<sup>12</sup> Counting requires uniformity: only individuals with the same number of atomic parts can be counted (Ionin and Matushansky 2006).<sup>13</sup> Uniformity is defined in (19), where  $n$  is a variable ranging over natural numbers and  $x_P$  is the number of parts of  $x$  that are  $P$ .<sup>14</sup>

$$(19) P \text{ is uniform} \Leftrightarrow_{\text{def}} \exists n(\forall x(P(x) \rightarrow |x|_P = n))$$

To illustrate, the lexical entry for the numeral **two** is given in (20), where  $\wp(P)$  is the power set of  $P$ , i.e.  $\wp(P) = \{Q \mid Q \subseteq P\}$ .

$$(20) \llbracket \mathbf{two} \rrbracket^w(P) = [\lambda x. \exists y(y \in \wp(P) \wedge |y|_P = 2 \wedge x = \text{sup}(y))] \text{ if } P \text{ is uniform, undefined otherwise}$$

Thus, numerals are of type  $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$ , that is, the type of restrictive modifiers. Suppose  $P = \{a, b, c\}$ , then  $\llbracket \mathbf{two} \rrbracket^w(P) = \{a \oplus b, a \oplus c, b \oplus c\}$ . However, if  $P = \{a, b, a \oplus b\}$ , then  $\llbracket \mathbf{two} \rrbracket^w(P)$  will be undefined, as  $P$  is not uniform.

<sup>10</sup>What holds for bare nouns in Vietnamese is assumed to hold for bare nouns in Bahnar and Mandarin also.

<sup>11</sup>The  $\lambda$ -notation is used here as proposed in Heim and Kratzer (1998: 34–35) which has become standard: “[ $\lambda\alpha:\phi.\gamma$ ]” represents the smallest function which maps every  $\alpha$  such that  $\phi$  to  $\gamma$ , where  $\alpha$  is the argument variable,  $\phi$  the domain condition, and  $\gamma$  the value description. Following standard practice, we use lower case “ $x$ ,” “ $y$ ” for variables of type  $e$ , and upper case “ $P$ ,” “ $Q$ ” for variables of type  $\langle e, t \rangle$ . Note that the domain condition is omitted when there is no need to make it explicit.

<sup>12</sup>Note that the supremum of  $P$  does not have to be a member of  $P$ .

<sup>13</sup>The reason for this requirement is obvious: if individuals of different numerosity, say  $a$  and  $b \oplus c$ , can be considered units in counting, we would not know how many dogs there are when we hear ‘**there are two dogs**’.

<sup>14</sup>Limiting  $n$ 's range to natural numbers serves to simplify the exposition and suffices for present purposes, but will obviously raise questions about such sentences as **John read 2.5 Russian novels**. We leave such issues for other occasions (see Haida and Trinh 2016, 2021 for discussion).

### 3.1.3 Classifiers

From what has just been said, it follows that numerals cannot combine with bare nouns in classifier languages, since these nouns denote cumulative predicates which are not uniform. This is why mediation of the classifier is required. The function *at* is defined as one which maps any cumulative predicate *P* to a subset of *P* whose members have no proper parts that are *P*.

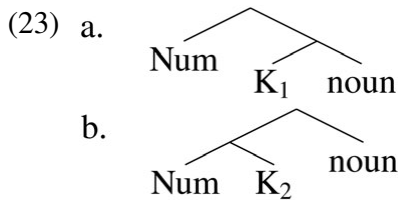
$$(21) x \in at(P) \Leftrightarrow_{\text{def}} x \in P \wedge \neg \exists y(y \in P \wedge yx)$$

We are now ready to propose meanings for the classifier. Anticipating the discussion which will come presently, we define two types of classifiers,  $K_1$  and  $K_2$ <sup>15</sup>

$$(22) \text{ a. } \llbracket [K_1] \rrbracket^w = [\lambda P.at(P)] \llbracket [KIND NP] \rrbracket^w = [\lambda w$$

$$\text{ b. } \llbracket [K_2] \rrbracket^w = [\lambda n:n \in D\langle\langle e, t \rangle, \langle e, t \rangle\rangle][\lambda P.n(at(P))]$$

As we can see,  $K_1$  maps a predicate to a predicate, while  $K_2$  maps a numeral and a predicate *P* to a predicate. This means we have two different bracketings for numeral phrases of the surface profile [Num K noun].



Both of these structures have been argued to exist. Specifically, it has been proposed that Chinese opts for (23a) and Japanese for (23b) (see Saito et al. 2008). We will argue that Chinese and Vietnamese opt for (23a) while Bahnar opts for (23b).

### 3.1.4 Definiteness and kind reference

A silent morpheme *the* is defined which has roughly the same meaning as the English definite article **the**. Specifically,  $\llbracket [THE] \rrbracket^w$  maps a *P* to the “maximal” entity in *P* if there is one, undefined otherwise. This captures both the existence and uniqueness presuppositions of definiteness (see Heim 1991).

$$(24) \llbracket [THE] \rrbracket^w(P) = sup(P) \text{ if } sup(P) \in P, \text{ undefined otherwise}$$

Suppose  $P = \{a\}$ , then  $\llbracket [THE] \rrbracket^w(P) = a$ . If  $P = \{a, b, a \oplus b\}$ , then  $\llbracket [THE] \rrbracket^w(P) = a \oplus b$ . However, if  $P = \{a, b\}$  or  $P = \emptyset$ , then  $\llbracket [THE] \rrbracket^w(P)$  will be undefined. This is the same THE as in Trinh (2011).

Among the individuals in the universe of discourse, there are kinds, which are functions from properties to individual concepts (see Chierchia and Turner 1988; Chierchia 1998). A silent morpheme *kind* is defined which turns nouns into names of kinds in (23), where  $\oplus P$  is the  $\oplus$ -closure of *P*, i.e.  $\oplus P = \{sup(Q) | Q \subseteq P\}$ .<sup>16</sup>

<sup>15</sup> $K_1$  is Trinh’s (2011) CL.  $K_2$  does not feature in that work, and is motivated here by the observations about Bahnar.

<sup>16</sup>Note that *kind* is *K* in Trinh (2011). The definition imposes on *kind* the requirement that its argument be a cumulative predicate, and that its extension contain more than one element in

- 451 (25)  $[[\text{KIND NP}]^w = [\lambda w. \text{sup}([\text{NP}]^w)]$  if  $[[\text{NP}]^w = \oplus[\text{NP}]^w$  and  $\exists w: |[\text{NP}]^w| > 1$ ,  
 452 undefined otherwise

453 Thus,  $[[\text{KIND}]]^w$  maps each cumulative predicate  $P$  into the function from each world  $w$   
 454 to  $[[\text{THE NP}]^w$ . Note that this definition of kind entails that neither  $[\text{CL NP}]$  nor  $[\text{Num}$   
 455  $\text{CL NP}]$  can combine with kind, as these are not cumulative predicates.

456 In addition to the operator kind, the inverse of kind is also defined. It is *ext*,  
 457 which is also a silent morpheme and which maps kinds into the plurality which  
 458 instantiate them in each world.<sup>17</sup>

- 459 (26)  $[[\text{EXT} [\text{KIND NP}]]^w = [[\text{KIND NP}]^w(w)$   
 460

461 In addition, we propose the following preference principle. At this point we will have  
 462 to assume that this is a primitive of natural language grammar.<sup>18</sup>

- 463 (27) The KIND-OVER-THE principle  
 464 If both  $[[[\text{KIND } \alpha]]^w$  and  $[[[\text{THE } \alpha]]^w$  are defined, use  $[\text{KIND } \alpha]$  instead of  $[\text{THE } \alpha]$

465 In other words, when it is possible to use kind, it is not possible to use THE.

### 466 3.2 Accounting for the facts

467 We are now in the position to derive the generalizations established in section 2. Our  
 468 proposal concerns only the functional lexicon, and is quite simple. Specifically, we  
 469 assume that Mandarin and Vietnamese differ in the way proposed by Trinh (2011),  
 470 and add Bahnar to the list. Our addition results in the following: (i) Bahnar has  $K_2$   
 471 but not  $K_1$ , while the opposite holds for Mandarin and Vietnamese; (ii) Bahnar  
 472 and Vietnamese have THE but Mandarin does not; (iii) all three languages have  
 473 KIND; (iv) Bahnar and Mandarin have EXT but Vietnamese does not. Table 2 sum-  
 474 marizes this cross-linguistic distribution of the functional morphemes  $K_1$ ,  $K_2$ , the,  
 475 kind and EXT.  
 476

477 This distribution of functional items across Bahnar, Mandarin, and Vietnamese  
 478 have consequences for the availability of syntactic structures among these three lan-  
 479 guages. It turns out that these consequences match the generalizations established in  
 480 section 2 precisely. Let us now derive these.

481 First, consider generalization 1, repeated below as (26).

- 482 (28) Generalization 1  
 483 DEM can combine with CL-NP in Mandarin and Vietnamese, but not in Bahnar  
 484

---

487  
 488 at least some possible world ( $X$  is the cardinality of set  $X$ ). The first requirement prevents the  
 489 kind reading for singular nouns in English and CL-NP combinations in Vietnamese. The  
 490 second requirement prevents concepts that are necessarily true of singularities only, for  
 491 example ‘being Noam Chomsky’ or ‘being the shoe on my left foot,’ from serving as names  
 492 of kinds (see Chierchia 1998). Both requirements are empirically motivated.

493 <sup>17</sup>Trinh (2011) uses the same name for the inverse of the kind operator, which is called K  
 494 there. Note that *ext* is of type  $\langle\langle s, e \rangle, e \rangle$ .

495 <sup>18</sup>This is the ‘‘Preference Principle’’ proposed in Trinh (2011), with the if-clause added.  
 Thus, it makes more explicit what is assumed in Trinh (2011).

	Bahnar	Mandarin	Vietnamese
$K_1$	No	Yes	Yes
$K_2$	Yes	No	No
the	Yes	No	Yes
kind	Yes	Yes	Yes
ext	Yes	Yes	No

**Table 2:** Functional elements in nominal structures

Proof – There are two possible parses for the DEM-CL-NP string: either [DEM [CL NP]] or [[DEM CL] NP].<sup>19</sup> Under the standard assumption that demonstratives, just like definite and indefinite articles, take predicates (i.e. expressions of type  $\langle e, t \rangle$ ) as arguments, [[DEM CL] NP] is excluded, since neither  $K_1$  nor  $K_2$ , our options for CL, is of type  $\langle e, t \rangle$ .<sup>20</sup> Thus, [DEM [CL NP]] is the only possible parse. Given that NP is a predicate, hence of type  $\langle e, t \rangle$ , CL in the DEM-CL-NP string must be of type  $\langle \langle e, t \rangle, \tau \rangle$  where  $\tau$  is some arbitrary type. As  $K_1$  is of type  $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$  and  $K_2$  of type  $\langle \langle \langle e, t \rangle, \langle e, t \rangle \rangle, \langle \langle e, t \rangle, \langle e, t \rangle \rangle \rangle$ , CL in the DEM-CL-NP string must be  $K_1$  and cannot be  $K_2$ . Since Bahnar has  $K_2$ , not  $K_1$ , while Mandarin and Vietnamese have  $K_1$ , not  $K_2$ , the DEM-CL-NP string can be generated in Mandarin and Vietnamese but not in Bahnar. QED.

Next, consider generalization 2, repeated below as (27).

(29) Generalization 2

CL-NP can be verbal arguments in Vietnamese, but not in Bahnar or Mandarin

Proof – Verbal arguments are of type  $e$  (Heim and Kratzer 1998). By hypothesis, [ $K_1$  NP] is of type  $\langle e, t \rangle$  and [ $K_2$  NP] is a type mismatch, that is, uninterpretable. This means that for a language to have CL-NP as verbal argument, it must have  $K_1$  and it must have a silent operator which maps [ $K_1$  NP] into an expression of type  $e$ . From the inventory of silent operators postulated above, only the fits the description of such an operator, which means that among the three languages under discussion, only Vietnamese fits the description of such a language: it is the only language to have both  $K_1$  and the in its functional lexicon. QED.

Next, consider generalization 3, repeated below as (28).

(30) Generalization 3

Num-CL-NP can be definite in Bahnar and Vietnamese, but not in Mandarin

<sup>19</sup>We assume that DEM, CL, and NP are adjacent and together make up one constituent, that is to say, there is no movement nor intervening empty categories. We thanks an anonymous reviewer for pointing out the necessity of this qualification.

<sup>20</sup>An anonymous reviewer raises the question of whether it is correct to claim that DEM takes predicates as arguments, given the fact that in languages like Portuguese and Greek exhibit DEM + definite article + NP sequences. We have not investigated the relevant phenomena and can only say, at this point, that what we analyze as DEM here might be split into the definite article plus another element in those languages.

541 Proof – By virtue of the definition of  $K_1$  and  $K_2$ , the Num-CL-NP string is parsed as  
 542 [[Num CL] NP] in Bahnar and as [Num [CL NP]] in Mandarin and Vietnamese. Both  
 543 of these structures, however, are expressions of type  $\langle e, t \rangle$ . Thus, the only way for a  
 544 language to have Num-CL-NP interpretable as definite is for it to have a silent oper-  
 545 ator which maps expressions of type  $\langle e, t \rangle$  into definite descriptions. Again, THE is the  
 546 only item among those postulated above which can do this. As it is available in  
 547 Bahnar and Vietnamese but not in Mandarin, we derive generalization 3. QED.<sup>21</sup>

548 Finally, consider generalization 4, repeated below as (31).

549 (31) Generalization 4

550 Bare NP can be definite in Bahnar and Mandarin, but not in Vietnamese

551 Proof – From the definition of THE, KIND and EXT, it follows that there are two parses of  
 552 NP, which results in a definite description: either [THE NP] or [ext [KIND NP]]. Given  
 553 the kind-over-the principle, [THE NP] is unavailable in Bahnar and Vietnamese, since  
 554 these languages have both the and kind. By hypothesis, Mandarin does not have the,  
 555 so [THE NP] is not available in Mandarin either. Thus, the only way for a bare NP to  
 556 be definite in Bahnar, Mandarin, or Vietnamese is to be parsed as [EXT [KIND NP]].  
 557 As Bahnar and Mandarin have ext, while Vietnamese does not, bare NPs can be def-  
 558 inite in the first two, but not in the last. QED.

#### 560 4. CONCLUSION

561 We have established four generalizations about the syntax and semantics of nominal  
 562 structures in three classifier languages – Bahnar, Mandarin, and Vietnamese – which  
 563 show an intricate pattern of cross-linguistic variation. We developed an analysis  
 564 which derives these generalization purely in terms of differences among the three lan-  
 565 guages with respect to their functional lexicon. Specifically, we defined pieces of  
 566 formal meaning which have been given empirical motivation in other works, and  
 567 advanced a proposal as to which piece is realized as a functional item in which lan-  
 568 guage. We then showed that syntactic and semantic consequences of our proposal  
 569 match the four generalizations we established in a precise manner.<sup>22</sup> The set of  
 570

---

572 <sup>21</sup>An anonymous reviewer pointed out that our conclusion disagrees with the proposal  
 573 made in Rullmann and You (2006), which provides an *e*-type analysis of CL-NP in  
 574 Mandarin. We acknowledge and thank the reviewer for drawing our attention to this fact.

575 <sup>22</sup>An anonymous reviewer raised the issue of “explanatory adequacy”: How do children  
 576 acquire grammars which involve such silent operators as THE and KIND from the primary linguis-  
 577 tic data? We admit that regarding this conceptually important issue we have nothing more con-  
 578 crete to say than the general statements made in the introduction, namely that acquisition is  
 579 setting of parameters and variation is confined to the functional lexicon. We would note, in  
 580 this connection, that the lack of a specific hypothesis about how the proposed syntactic and  
 581 semantic analysis squares with theories of language acquisition is, in our view, a feature  
 582 common to many, if not most, of the works in the technical literature. The same reviewer  
 583 pointed out that our account would be more explanatorily adequate if further facts were pre-  
 584 sented whose description required the elements of our analysis, and in addition to showing  
 585 that the data *can* be, we also show that they *have to* be, analyzed the way we propose. This  
 point is, of course, valid, and the criticism could be made that our present proposal is more



586 facts we discussed is admittedly compact, but its small size allows a fully explicit  
 587 account to be formulated which forces puzzling stipulations such as the kind-over-  
 588 the principle to be manifest and which makes it possible to execute exact computa-  
 589 tions of the meaning of syntactic structures. In addition, it invites expansion of the  
 590 data base which we hope to pursue in future work.

## 592 REFERENCES

- 593
- 594 Baker, Mark C. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge  
 595 University Press.
- 596 Barker, Chris. 1992. Group terms in English: Representing groups as atoms. *Journal of*  
 597 *Semantics* 9(1): 69–93.
- 598 Borer, Hagit. 1984. *Parametric syntax: Case studies in Semitic and Romance languages*.  
 599 Dordrecht: Foris.
- 600 Borer, Hagit. 2005. *Structuring sense, Volume 1: In name only*. Oxford: Oxford University  
 601 Press on Demand.
- 602 Bósković, Željko, and Jon Gajewski. 2011. Semantic correlates of the NP/DP parameter.  
 603 *Proceedings of NELS* 39: 121–134.
- 604 Cheng, Lisa Lai Shen, and Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure  
 605 of NP. *Linguistic Inquiry* 30(4): 509–542.
- 606 Cheng, Lisa Lai Shen, and Rint Sybesma. 2005. Classifiers in four varieties of Chinese. In *The*  
 607 *Oxford Handbook of Comparative Syntax*, ed. Guglielmo Cinque and Richard Kayne,  
 608 259–292. Oxford: Oxford University Press.
- 609 Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics*  
 610 6(4): 339–405.
- 611 Chierchia, Gennaro. 2010. *Meaning as an inferential system: Polarity and free choice phenom-*  
 612 *ena*. Cambridge, MA: Harvard University Press.
- 613 Chierchia, Gennaro, and Raymond Turner. 1988. Semantics and property theory. *Linguistics*  
 614 *and Philosophy* 11(3): 261–302.
- 615 Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In  
 616 *Principles and parameters in comparative grammar*, ed. Robert Freidin, 417–454.  
 617 Cambridge, MA: MIT Press.
- 618 Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from building*  
 619 *20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Keneth Hale and Samuel Jay  
 620 Keyser, 1–52. Cambridge, MA: MIT Press.
- 621 Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- 622 Chomsky, Noam. 1998. Minimalist inquiries. MIT Working Papers in Linguistics 15.
- 623 Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Linguistics*, ed. Michael  
 624 Kenstowicz, 1–52. Cambridge, Massachusetts: MIT Press.
- 625 Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond: The cartog-*  
 626 *raphy of syntactic structures*, ed. Adriana Belletti, vol. 3, 104–131. Oxford: Oxford  
 627 University Press.
- 628 Chung, Sandra, and William A. Ladusaw. 2004. *Restriction and saturation*. Cambridge, MA:  
 629 MIT Press.

630 like the work of an engineer whose goal is primarily to get things to work than that of a scientist  
 whose goal is primarily to explain why things work the way they do. To this we can only say:  
 “guilty as charged.”



- 631 Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic*  
632 *Inquiry* 36(3): 315–332.
- 633 Dayal, Veneeta. 2004. The universal force of free choice *any*. *Linguistic Variation Yearbook*  
634 4(1): 5–40.
- 635 Haida, Andreas, and Tue Trinh. 2016. Beyond density. Talk given at the Exh Workshop, MIT.
- 636 Haida, Andreas, and Tue Trinh. 2021. Splitting atoms in natural language. In *Formal*  
637 *approaches to number in Slavic and beyond*, ed. Mojmír Döcekal and Marcin Wągiel,  
638 277–296. Language Science Press.
- 639 Heim, Irene. 1991. (Artikel und Definitheit. In *Semantik: Ein internationales Handbuch der*  
640 *zeitgenössischen Forschung*), ed. Arnim von Stechow and Dieter Wunderlich, 487–  
641 535. Berlin: De Gruyter.
- 642 Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Oxford:  
643 Blackwell.
- 644 Her, One-Soon. 2012a. Distinguishing classifiers and measure words: A mathematical perspec-  
645 tive and implications. *Lingua* 122: 1668–1691.
- 646 Her, One-Soon. 2012b. Structure of classifiers and measure words: A lexical functional  
647 account. *Language and Linguistics* 13(6): 1211–1251.
- 648 Huang, C.-T. James. 1981. Move *Wh* in a language without *Wh* movement. *The Linguistic*  
649 *Review* 1: 369–416.
- 650 Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Doctoral  
651 dissertation, Massachusetts Institute of Technology.
- 652 Ionin, Tanya, and Ora Matushansky. 2006. The composition of complex cardinals. *Journal of*  
653 *Semantics* 23(4): 315–360.
- 654 Krifka, Manfred. 1995. Common nouns: a contrastive analysis of English and Chinese. In *The*  
655 *generic book*, ed. Greg Carlson and Francis Jeffrey Pelletier, 398–411. Chicago:  
656 University of Chicago Press.
- 657 Krifka, Manfred. 1999. Mass expressions. In *Concise encyclopedia of grammatical categories*,  
658 ed. Keith Brown and Jim Miller, 221–223. Amsterdam: Elsevier.
- 659 Krifka, Manfred. 2003. Bare NPs: Kind-referring, indefinites, both, or neither? *Proceedings of*  
660 *SALT* 13: 180–203.
- 661 Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical  
662 approach. In *Meaning, use, and the interpretation of language*, ed. Rainer Baeuerle,  
663 Christoph Schwarze, and Arnim von Stechow, 302–323. Berlin: de Gruyter.
- 664 Longobardi, Giuseppe. 2001. The structure of DPs: Some principles, parameters, and pro-  
665 blems. In *Handbook of contemporary syntactic theory*, ed. Mark Baltin and Chris  
666 Collins, 562–603. Oxford: Blackwell.
- 667 Manzini, M. Rita, and Kenneth Wexler. 1987. Parameters, binding theory, and learnability.  
668 *Linguistic Inquiry* 18(3): 413–444.
- 669 Pelletier, Francis Jeffrey, and Lenhart K. Schubert. 1989. Mass expressions. In *Handbook of*  
670 *philosophical logic*, ed. Dov Gabbay and Franz Guentner, vol. 4, 327–407. Dordrecht:  
671 Reidel: Springer.
- 672 Pollock, Jean-Yves. 1989. Verb movement, UG and the structure of IP. *Linguistic Inquiry* 20  
673 (3): 365–424.
- 674 Rullmann, Hotze, and Aili You. 2006. General number and the semantics and pragmatics of  
675 indefinite bare nouns in Mandarin Chinese. In *Where semantics meets pragmatics*, ed.  
676 Klaus von Stechow and Ken Turner, 175–196. Amsterdam: Elsevier.
- 677 Saito, Mamoru, Jonah T.-H. Lin, and Keiko Murasugi. 2008. N'ellipsis and the  
678 structure of noun phrases in Chinese and Japanese. *Journal of East Asian Linguistics*  
679 17(3): 247–271.

- 676 Schwarzschild, Roger. 1992. Types of plural individuals. *Linguistics and Philosophy* 15(6):  
677 641–675.
- 678 Trinh, Tue. 2011. Nominal reference in two classifier languages. *Proceedings of Sinn und*  
679 *Bedeutung* 15: 629–644.
- 680 Trinh, Tue, and Yasutada Sudo. 2009. The Indefiniteness Effect in Mandarin and Vietnamese.  
681 Talk given at Zentrum für Allgemeine Sprachwissenschaft Berlin.
- 682 Watanabe, Akira. 2010. Notes on nominal ellipsis and the nature of *no* and classifiers in japa-  
683 nese. *Journal of East Asian Linguistics* 19(1): 61–74.
- 684 Wu, Yicheng, and Adams Bodomo. 2009. Classifiers ≠ determiners. *Linguistic Inquiry* 40(3):  
685 487–503.
- 686
- 687
- 688
- 689
- 690
- 691
- 692
- 693
- 694
- 695
- 696
- 697
- 698
- 699
- 700
- 701
- 702
- 703
- 704
- 705
- 706
- 707
- 708
- 709
- 710
- 711
- 712
- 713
- 714
- 715
- 716
- 717
- 718
- 719
- 720