Deriving four generalizations about nominals in three classifier languages

TRANG PHAN
Email: chengnn85@gmail.com

TUE TRINH
Email: trinhhuutue@tdtu.edu.vn

HUNG PHAN
Email: hungphanluong@gmail.com

Abstract

This note 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 nouns 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

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**.[1] 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).[2]

(1) John có một *(con) chó
John have one CL dog
'John has one 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 variations exist among languages of both types. For example, bare nouns can be definite in Chinese but not in Vietnamese, while classifier-noun combinations

[1] In line object language expressions will be **boldfaced**.

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

can be definite in Vietnamese but not in Chinese (cf. 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 (cf. Barker, 1992; Krifka, 1999; Link, 1983; Pelletier and Schubert, 1989; Schwarzschild, 1992, 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 (cf. Chierchia, 1998, 2010; Dayal, 2004; Krifka, 1995). 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 (cf. Saito et al., 2008; Cheng and Sybesma, 1999, 2005; Watanabe, 2010; Wu and Bodomo, 2009, among others). This paper 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 "principle and parameter" theory (cf. 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 languages can vary, i.e. the parameters. A particularly influential view of parametric variation, which is sometimes called the "Borer-Chomsky conjecture" (Borer, 1984), holds that all such variation is to be reduced to the functional lexicon. In other words, syntactic differences between languages, with the exception of those resulting from Saussurean arbitrariness, are assumed to derive from differences with respect to functional categories. For example, it has been proposed that whether wh-movement exists depends on C (cf. Huang, 1981, 1982), whether V-raising exists depends on T (cf. Pollock, 1989; Chomsky, 1991), and whether N rasing exists depends on D (cf. Longobardi, 2001; Cinque, 2005). Of course, there is no a priori reason to assume that functional items are restricted to those of categories C, T, and D, or that variation is restricted to the morphological ability to trigger movement. The term "functional category" is not definitional, and its extension is to be determined based on considerations of empirical adequacy, as well as theoretical economy and elegance. In the same way, the possibility must be kept open that functional categories may differ not only with respect to their morphological make-up, but also with respect to their availability: the functional lexicon of one language may contain a certain item which is absent from the functional lexicon of another language (cf. Manzini and Wexler, 1987; Bošković and Gajewski, 2011). In this paper, we will argue that our three-way comparison of Bahnar, Chinese, and Vietnamese shows that differences of precisely this kind exist.

Before we end this introduction and get to the main discussion, we will address the question of the grammatical status of classifiers. While it is quite uncontroversial to assume that morphemes such as definite articles, demonstratives, or silent type shifting operators are functional items, it is less so with classifiers. As classifiers seem to indicate the "class" of the nouns, which is a cognitive notion, one might feel that classifiers should be considered subtantives. Our assumption in this paper will be that they are functional items. We will now provide some justification for this assumption, using, without loss of generalization, examples from Vietnamese for illustration.[3]

One criterion for some lexical item to be considered "functional" is that it can be omitted without affecting the intended (truth-conditional) meaning: (2) will be understood as being logically equivalent, i.e. expressing the same proposition, as (1), even though it will be perceived as an ungrammatical sentence, e.g. one which would be spoken by a foreigner who has not mastered the language.

(2) *John có một chó
John have one dog

Related to this criterion is the phenomenon of informative redundancy. Although classifiers do "classify" nouns to some extent, their co-occurence with the nouns they combine with is, in principle, not predictable from the conceptual content of the nouns and must essentially be learned by heart. Take the classifiers **con** and **cái** in Vietnamese, for example. The first typically combines with nouns which denote animals, and the second typically combines with nouns which denote inanimate objects. However, the noun **thuyền** 'boat' may combine with both **con** and **cái**, and there is a strong preference, in standard Hanoi dialect at least, for combining the noun **dao** 'knife' with **con** rather than with **cái**.

(3) a. John nhìn thấy một con/cái thuyền
John see one CL boat
'John sees a boat'
b. John cầm một con/??cái dao
John hold one CL knife
'John is holding a knife'

Thus, the dependency between a noun and its accompanying classifier is similar to that between a noun and its grammatical gender in such languages as German and French: it is only partly semantically motivated.

[3]Note, importantly, that we are talking about classifiers and not measure words such as **herd**, **cup**, or **kilogram**. Phrases such as **a herd of cows**, **a cup of milk**, or **a kilogram of meat** express measurements which are more or less purpose related and exist in both classifier languages and number marking languages. Classifier phrases such as **con chó** 'CL dog' express "natural units" (cf. Krifka, 2003) and constitute the basis for a typological distinction. For more discussion on the differences between classifiers and measure words see Her (2012b,a).

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

(4) a. *John có một cái chó
John have one CL dog
b. *John hat einen Haus
John have one.masc house

Last but not least, classifiers share with other uncontroversial cases of functional items in having a "world independent" semantics: their denotation stays constant across states of affairs. This will become clear in the discussion below. At this point, we take these considerations to be sufficient reasons for assuming that classifiers are functional items.

2. Four generalizations about Bahnar, Mandarin, and Vietnamese

Bahnar and Mandarin are similar to Vietnamese in being "classifier languages" of the East Asian variety: nouns can only combine with numerals through the mediation of a classifier, as has been illustrated for Vietnamese in the previous section. Let us now turn to the discussion of demonstratives, argumenthood, and definiteness in these three languages.[4]

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

(5) a. nei (liang) zhi gou
DEM two CL dog
b. (hai) con chó đó
two CL dog DEM

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

(6) *(ʔbal) tɔʔ kɔʔ nej
two CL dog DEM

[4]Observations on Vietnamese are based on intuitions of all three authors who are native speakers of this language. Observations on Bahnar are based on field work done by Hung Phan, the third author. Observations on Mandarin Chinese are based on Cheng and Sybesma (1999).

Let us state the first generalization.

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

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

(8) con chó muốn sang đường
CL dog want cross road

In contrast, this does not hold for Bahnar and Mandarin, as shown in (9-a) and (9-b), respectively.[5]

(9) a. *tɔʔ kɔʔ waʔ kwa tɤrɔŋ
CL dog want cross road
b. *zhi gou yao guo malu
CL dog want cross road

Let us state the second generalization.

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

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

(11) a. ʔbal tɔʔ kɔʔ waʔ kwa tɤrɔŋ
two CL dog want cross road
'Two dogs/the two dogs want to cross the road.'
b. hai con chó muốn sang đường
two CL dog want cross road
'Two dogs/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

Let us state the third generalization.

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

[5]We will disregard phrases in the object position due to complications that we think are orthorgonal to the questions at hand (cf. Trinh and Sudo, 2009).

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

(14) a. kɔʔ waʔ kwa tɤrɔŋ
dog want cross road
'Dogs want to cross the road.' / 'The dog(s) want(s) to cross the road.'
b. gou yao guo malu
dog want cross road
'Dogs want to cross the 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
'Dogs want to cross the road.' / *'The dog(s) want(s) to cross the road.'

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.

	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

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 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). We will first lay some theoretical groundwork in a brief and concise manner, introducing claims and their accompanying definitions of terms with the minimal amount of discussion. We would note

that most, if not all, of these claims have been motivated elsewhere, in particular in Chierchia (1998) and Trinh (2011) as well as in works cited in the text at the relevant places. After that we will show how the generalizations stated in the previous section are derived from the theory.

3.1 Theoretical groundwork

3.1.1 Bare nouns

We start with the semantics of bare nouns. Following Chierchia (1998) and several others, we hypothesize that bare nouns in number marking languages denote atomic predicates, i.e. sets of singularities, while bare nouns in classifier languages denote cumulative predicates, i.e. sets of both singularities and pluralities. 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 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.[6] The lexical entries for **dog** and **chó** are given in (17).[7]

(17) a. $[\![\textbf{dog}]\!]^w = [\lambda x.x \text{ is a singular dog}] = \{a,b,c\}$
b. $[\![\textbf{chó}]\!]^w = [\lambda x.x \text{ is a singular dog or a plurality of dogs}] = \{a,b,c, a{\oplus}b, a{\oplus}c, b{\oplus}c, a{\oplus}b{\oplus}c\}$

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 $a{\oplus}b \not\sqsubset a{\oplus}b$.

3.1.2 Numerals

Let us now turn to the semantics of numerals. First, 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 (proper or non-proper) part.

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

Suppose $P = \{a,b,c,a{\oplus}b\}$, then $sup(P) = a{\oplus}b{\oplus}c$.[8] Now, following Ionin and Matushansky (2006), we assume that counting requires "uniformity": only individuals with the same number of atomic parts can be counted.[9] We define uniformity as in

[6] We assume that what we say about the denotation of bare nouns in Vietnamese will hold for the other two classifier languages Bahnar and Mandarin also.

[7] Our usage of the λ-notation is one proposed in Heim and Kratzer (1998: 34–35) which has become standard: "$[\lambda\alpha : \phi.\gamma]$" represents the smallest function which maps every α such that ϕ to γ, where α is the argument variable, ϕ the domain condition, and γ 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 are omitted when there is no need to make it explicit.

[8] Note that the supremum of P does not have to be a member of P.

[9] 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**.

(19), where n is a variable ranging over natural numbers and $|x|_P$ is the number of parts of x that are P.[10]

(19) P 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. $\wp(P) = \{Q|Q \subseteq P\}$.

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

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

3.1.3 Classifiers

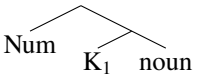
From what we just said it follows that numerals cannot combine with bare nouns in classifier languages, since these nouns denote cumulative predicates which are not uniform. This is why mediation of the classifier is required. Let us first define a function *at* which maps any cumulative predicate P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

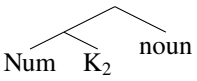
(21) $x \in at(P) \Leftrightarrow_{\text{def}} x \in P \wedge \neg\exists y(y \in P \wedge y \sqsubset x)$

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.

(22) a. $[\![K_1]\!]^w = [\lambda P.at(P)]$
b. $[\![K_2]\!]^w = [\lambda n : n \in D_{\langle\langle e,t\rangle,\langle e,t\rangle\rangle}[\lambda P.n(at(P))]]$

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].

(23) a.


b.


[10] 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, 2018) for discussion).

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

3.1.4 Definiteness and kind reference

We define a silent morpheme THE which has roughly the same meaning as English definite article **the**. Specifically, $[\![\text{THE}]\!]^w$ 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).

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

Suppose $P = \{a\}$, then $[\![\text{THE}]\!]^w(P) = a$. If $P = \{a, b, a \oplus b\}$, then $[\![\text{THE}]\!]^w(P) = a \oplus b$. However, if $P = \{a, b\}$ or $P = \emptyset$, then $[\![\text{THE}]\!]^w(P)$ will be undefined.

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 in (25), where $\oplus P$ is the $\oplus$-closure of P, i.e. $\oplus P = \{sup(Q) | Q \subseteq P\}$.[11]

(25) $[\![\text{KIND NP}]\!]^w = [\lambda w.sup([\![\text{NP}]\!]^w)]$ if $[\![\text{NP}]\!]^w = \oplus[\![\text{NP}]\!]^w$, undefined otherwise

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

In addition to the operator KIND, we define its inverse, EXT, which is also a silent morpheme and which maps kinds into the plurality which instantiate them in each world.

(26) $[\![\text{EXT KIND NP}]\!]^w = sup([\![\text{KIND NP}]\!]^w(w))$

In addition, we propose the following preference principle. At this point we will have to assume that this is a primitive of natural language grammar.

(27) The KIND-over-THE principle
If neither [KIND α] nor [THE α] is a type mismatch, use [KIND α] instead of [THE α]

In other words, when it is possible to use KIND, using THE will lead to formal deviance.

[11] The reason for this requirement is that concepts that are true of singularities only, for example 'being Noam Chomsky' or 'being the shoe on my left foot,' should not be able to serve as names of kinds (cf. Chierchia, 1998).

3.2 Accounting for the facts

We are now in the position to derive the generalizations established in section 2. Our proposal concerns only the functional lexicon, and is quite simple. Specifically, we propose that Bahnar, Mandarin and Vietnamese differ 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. Table 2 summarizes this cross-linguistic distribution of the functional morphemes K_1, K_2, THE, KIND and EXT.

	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

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

First, consider generalization 1, repeated below.

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

Proof – Logically, there are two possible parses for the DEM-CL-NP string: either [DEM [CL NP]] or [[DEM CL] NP]. 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 $\langle e,t\rangle$. 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 τ 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.

(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, i.e. 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 among the three languages under discussion, only Vietnamese fits the description of such a language: it is the only language that have both K_1 and THE in its functional lexicon. QED.

Next, consider generalization 3, repeated below.

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

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

Finally, consider generalization 4, repeated below.

(31) Generalization 4
Bare NP can be definite in Bahnar and Mandarin, but not in Vietnamese

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

4. Conclusion

We have established four generalizations about the syntax and semantics of nominal structures in three classifier languages – Bahnar, Mandarin, and Vietnamese – which show an intricate pattern of cross-linguistic variation. We developed an analysis which derives these generalization purely in terms of differences among the three languages with respect to their functional lexicon. Specifically, we defined pieces of formal meaning which have been given empirical motivation in other works and advanced a proposal as to which piece is realized as a functional item in which language. We then show that syntactic and semantic consequences of our proposal match the four generalizations we established in a precise manner. The set of facts we

discussed is admittedly rather compact, but its small size allows a fully explicit account to be formulated which forces puzzling stipulations such as the KIND-over-THE principle to be manifest and which makes it possible to execute exact computations of the meaning of syntactic structures. In addition, it invites expansion of the data base which we hope to pursue in future work.

REFERENCES

Barker, Chris. 1992. Group terms in English: Representing groups as atoms. *Journal of Semantics* 9(1): 69–93.

Borer, Hagit. 1984. *Parametric Syntax*. Dordrecht: Foris.

Bošković, Željko and Jon Gajewski. 2011. Semantic correlates of the NP/DP parameter. *Proceedings of NELS* 39: 121–134.

Cheng, Lisa Lai Shen and Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4): 509–542.

Cheng, Lisa Lai Shen and Rint Sybesma. 2005. Classifiers in four varieties of Chinese. In *The Oxford Handbook of Comparative Syntax*, ed. Guglielmo Cinque and Richard Kayne, 259–292. Oxford: Oxford University Press.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6(4): 339–405.

Chierchia, Gennaro. 2010. Meaning as an inferential system: Polarity and free choice phenomena. Harvard University.

Chierchia, Gennaro and Raymond Turner. 1988. Semantics and property theory. *Linguistics and Philosophy* 11(3): 261–302.

Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In *Principles and Parameters in Comparative Grammar*, ed. Robert Freidin, 417–454. Cambridge: MIT Press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from building 20: Essays in linguistics in honor of sylvain bromberger*, ed. Keneth Hale and Samuel Jay Keyser, 1–52. Cambridge: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, Noam. 1998. Minimalist inquiries. MIT Working Papers in Linguistics 15.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond: The cartography of syntactic structures*, ed. Adriana Belletti, vol. 3, 104–131. Oxford: Oxford University Press.

Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36(3): 315–332.

Dayal, Veneeta. 2004. The universal force of free choice "any". *Linguistic Variation Yearbook* 4(1): 5–40.

Haida, Andreas and Tue Trinh. 2016. Beyond density. Talk given at the Exh Workshop, MIT.

Haida, Andreas and Tue Trinh. 2018. Splitting atoms in natural language. Unpublished manuscript.

Heim, Irene. 1991. Artikel und Definitheit. In *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, ed. Arnim von Stechow and Dieter Wunderlich, 487–535. Berlin: De Gruyter.

Heim, Irene and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Her, One-Soon. 2012a. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122: 1668–1691.

Her, One-Soon. 2012b. Structure of classifiers and measure words: A lexical functional account. *Language and Linguistics* 13(6): 1211–1251.

Huang, C.-T. James. 1981. Move *Wh* in a language without *Wh* movement. *The Linguistic Review* 1: 369–416.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, Massachusetts Institute of Technology.

Ionin, Tanya and Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23(4): 315–360.

Krifka, Manfred. 1995. Common nouns: a contrastive analysis of English and Chinese. In *The Generic Book*, ed. Greg Carlson and Francis Jeffrey Pelletier, 398–411. Chicago: University of Chicago Press.

Krifka, Manfred. 1999. Mass expressions. In *Concise Encyclopedia of Grammatical Categories*, ed. Keith Brown and Jim Miller, 221–223. Amsterdam: Elsevier.

Krifka, Manfred. 2003. Bare NPs: Kind-referring, indefinites, both, or neither? *Proceedings of SALT* 13: 180–203.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Meaning, Use, and the Interpretation of Language*, ed. Rainer Baeuerle, Christoph Schwarze, and Arnim von Stechow, 302–323. Berlin: de Gruyter.

Longobardi, Giuseppe. 2001. The structure of DPs: Some principles, parameters, and problems. In *Handbook of Contemporary Syntactic Theory*, ed. Mark Baltin and Chris Collins, 562–603. Oxford: Blackwell.

Manzini, M. Rita and Kenneth Wexler. 1987. Parameters, binding theory, and learnability. *Linguistic Inquiry* 18(3): 413–444.

Pelletier, Francis Jeffrey and Lenhart K. Schubert. 1989. Mass expressions. In *Handbook of Philosophical Logic*, ed. Dov Gabbay and Franz Guenthner, vol. 4, 327–407. Dordrecht: Reidel.

Pollock, Jean-Yves. 1989. Verb movement, UG and the structure of IP. *Linguistic Inquiry* 20(3): 365–424.

Saito, Mamoru, Jonah T.-H. Lin, and Keiko Murasugi. 2008. N′-ellipsis and the structure of noun phrases in Chinese and Japanese. *Journal of East Asian Linguistics* 17(3): 247–271.

Schwarzschild, Roger. 1992. Types of plural individuals. *Linguistics and Philosophy* 15(6): 641–675.

Trinh, Tue. 2011. Nominal reference in two classifier languages. *Proceedings of Sinn und Bedeutung* 15: 629–644.

Trinh, Tue and Yasutada Sudo. 2009. The Indefiniteness Effect in Mandarin and Vietnamese. Talk given at Zentrum für Allgemeine Sprachwissenschaft Berlin.

Watanabe, Akira. 2010. Notes on nominal ellipsis and the nature of no and classifiers in japanese. *Journal of East Asian Linguistics* 19(1): 61–74.

Wu, Yicheng and Adams Bodomo. 2009. Classifiers ≠ determiners. *Linguistic Inquiry* 40(3): 487–503.