

Abstract: This paper reports Vietnamese results from the Genericity Toolkit. It summarizes how Vietnamese expresses kind, subkind, generic, and token reference using bare nouns, classifier-marked definites, and indefinites. The questionnaire shows that bare nouns are the default strategy for kind reference and generic statements, while definites and indefinites are largely restricted to object-level reference. The same patterns are found across natural kinds, artifacts, and social kinds. The report provides a descriptive overview of genericity in Vietnamese and contributes comparable data for cross-linguistic research.

1. Introduction

Vietnamese is a language of the Austroasiatic family, belonging to the Mon-Khmer branch and more specifically to the Vietic subgroup. It is spoken primarily in Vietnam, where it serves as the official and national language, and it is also used by substantial diaspora communities in countries such as the United States, Australia, France, Canada, and Germany. Vietnamese has approximately 85 to 90 million native speakers, with an additional 10 to 15 million people speaking it as a second language, making it one of the most widely spoken languages in Southeast Asia. Morphologically it is a monosyllabic and isolating language, which means morpheme and syllable boundaries co-incide and there is no inflection. There are thus no affixes, words are formally constant across all environments, and grammatical information is expressed by functional free morphemes. Syntactically it is quite consistently head-initial: objects follow verbs, adjectives follow nouns, sentences follow complementizers, etc. (cf. Nguyễn Tài Cẩn 1975).

1.1. Classifiers

With respect to the nominal domain, Vietnamese is a 'classifier' language of the East Asian variety (Bisang 1999): the combination of a noun and a numeral regularly requires mediation by a word which may to some degree depend on the 'class' of the objects described by the noun (hence the term 'classifier'). For example, the combination of **hai** 'two' and **chó** 'dog' must be mediated by the classifier **con**, which is generally selected by nouns denoting animals. If the noun denotes an inanimate object, the classifier would generally be the word **cái**.

- (1) a. **Nam có hai con chó**
Nam have two con dog
'Nam has two dogs'
- b. **Nam mua ba cái bàn**
Nam buy three cái table
'Nam bought three tables'

There are exceptions, however. In the standard Hanoi dialect, the noun **dao** 'knife' selects **con** while the noun **thuyền** can go with both **con** and **cái**. Thus, animacy does not make a clean cut. Many other semantic features are also at play, and they seem to crisscross unsystematically. One would think that the classifier **quả**, for example, is selected by nouns denoting things that are spherical such as **cam** 'orange', **bóng** 'ball', **đất** 'earth', and **cầu** 'sphere' itself, and that the classifier **quyển** is selected by nouns denoting things with pages that are bound at the long edge such as **sách** 'book', **sổ** 'diary', and **vở** 'cahier'. However, **quả** turns out to be selected by all nouns denoting fruits, whether or not the fruit in question is

spherical: 'four bananas' is **bốn quả chuối** (four quả banana). In addition, **quả** is selected by **núi** 'mountain', which is neither spherical nor a fruit. Another example showing the unpredictability of classifier selection is the noun **thư** 'letter' which selects the classifier **lá**, not **quyển**, even in cases where the letter in question has many pages which are bound at the long edge. Also, the nouns **báo** and **tạp chí** can be used interchangeably to mean 'journal', but **báo** requires **tờ** as classifier while **tạp chí** allows for both **tờ** and **quyển**.

I believe we can safely say that the relationship between a noun and the classifier it selects is essentially formal. The fact that in Vietnamese **dao** 'knife' and **chó** 'dog' select **con** while **bóng** 'ball', **chuối** 'banana' and **núi** 'mountain' select **quả** as classifier seems to be similar in kind to the fact that in German the word **Auto** requires **das** and the word **Wagen** requires **der** as definite article. The functional nature of classifiers is also corroborated by the intuition that they affect acceptability, not interpretation: **hai cái núi** is perceived as the wrong way to say 'two mountains', not as the right way to say something other than 'two mountains'. The expression is wrong in essentially the same way **das Wagen** and **der Auto** are wrong in German. I will henceforth gloss all classifiers as CL.

Do all nouns require classifiers for combination with numerals? No. There are 'special nouns' in Vietnamese which combine directly with numerals. They include **ngày** 'day', **người** 'person', **số** 'number', **loài** 'species', **lực lượng** 'forces', **đảng** '(political) party', **nhà nước** 'government', and many others.

- (2) a. **Nam có một số sách**
Nam have one number book
'Nam has a number of books'
- b. **Nam nuôi hai loài chó**
Nam breed two species dog
'Nam breeds two species of dogs'
- c. **Nam gặp ba người**
Nam meet three person
'Nam met three persons'

I know of no (successful) attempt to characterize these nouns as a natural class. For present purposes I will assume that it is a formal fact about a noun whether it requires a classifier. In other words, I will take the fact that **người** 'person' can, but **mèo** 'cat' cannot, combine directly with a numeral to be of essentially the same kind as the fact that Italian **mobile** can, but English **furniture** cannot, combine directly with a numeral. In most of what follows I will leave special nouns aside.¹

Note that the word 'noun' in what I have said refers to what is usually called 'count nouns', i.e. those whose denotation can be modelled as sets: 'dog', 'cat', 'person'. Unsurprisingly, Vietnamese also has 'mass nouns', i.e. those which denote substances. Examples are **nước** 'water', **sữa** 'milk', **cát** 'sand', **sắt** 'iron', etc. As these do not come in natural units, there are no classifiers for them. Counting regarding these nouns is effected by measure words such as **cốc** 'cup', **hộp** 'box', **cân** 'kilogram'. Thus, they

¹ Note that these special nouns can combine directly with numerals but they do not have to. Classifiers can optionally intervene, and in such cases, the choice of the classifier does affect the semantics of the combination. Take **người** 'person', for example. The phrase **một thằng người** means a male person due to the classifier **thằng**, while **một chủng người** means an ethnic group, i.e. a kind of persons, due to the classifier **chủng**. I thank Matthew Husband for drawing my attention to this fact.

behave like their counterparts in English. I will also leave mass nouns and measure words aside.

1.2. Definiteness and demonstratives

The profile of definite nominals in Vietnamese seems to be an eclectic mix. We see from the English translations above that expressions of the form [numeral + CL + NP] can be indefinite. Can they be definite? The answer is that they can be, provided the numeral is not **một** 'one'.

- (3) a. **Nam mua một quyển sách**
Nam buy one CL book
i. 'Nam bought a book'
ii. # 'Nam bought the book'
- b. **Nam mua hai quyển sách**
Nam buy two CL book
i. 'Nam bought two books'
ii. 'Nam bought the two books'

When the numeral is not present, [CL + NP] expresses a singular definite. For plural definites, the word **những** or the word **các**, both of which I will gloss as PL, is added to the left of [CL + NP].

- (4) a. **Nam mua quyển sách**
Nam buy CL book
'Nam bought the book'
- b. **Nam mua những/các quyển sách**
Nam buy PL CL book
'Nam bought the books'

As for plural indefinites, the standard way to express 'some Ns' in Vietnamese is by way of using **một số** 'a number of', as exemplified by (2a).

Definite nominals can combine with the demonstrative articles **này** 'this'/'these' and **đó** 'that'/'those'.

- (5) a. **Nam thích quyển sách này**
Nam like CL book this
'Nam likes this book'
- b. **Nam thích những/các quyển sách đó**
Nam like PL CL book those
'Nam likes those books'

1.4. Bare nouns

NP can occur without CL and numerals.

- (6) **Nam có chó**
Nam have dog
'Nam has dogs'

Note the translation of the bare noun **chó** as 'dogs'. It has been claimed that bare plurals in English are semantically 'number-neutral', which means that the basic, literal meaning of **Nam has dogs** is 'Nam has one or more than one dogs', and the inference that he has more than one dog comes about by way of implicatures. In this respect bare nouns in Vietnamese resemble bare plurals in English: they are number-neutral (Chierchia 1998, Dayal 2004). Thus, (6) is true if Nam has exactly one or more than one dog. The difference between Vietnamese and English is that the plurality inference is much weaker in Vietnamese than it is in English. An English speaker might agree that **Nam has dogs** is not false in case Nam has exactly one dog,

but he would likely judge the sentence to be highly misleading (Sauerland 2003; Sauerland, Andersen & Yatsushiro 2005; Spector 2007; Zweig 2009). A Vietnamese speaker, I believe, would more likely than not infer from (6) that Nam has more than one dog, but would not say that the sentence is misleading if it turns out that he has exactly one dog. For this discussion, I think we can safely claim that bare nouns in Vietnamese are number-neutral.

The translation I provided for (6) shows that bare nouns can be indefinite: the sentence is false, which means it is felicitous, in a context where there are no dogs. Is there a reading of this sentence under which it is infelicitous if there are no dogs? In other words, can **chó** 'dog' mean 'the dog' or 'the dogs'? The answer is no. Bare nouns cannot be definite in Vietnamese. We predict, then, that bare nouns cannot combine with demonstratives.

- (7) ***Nam thích chó này**
 Nam like dog this
 ('Nam likes this dog' / 'Nam likes these dogs')

1.5. Quantification

1.5.1. Nominal quantification

We have considered existential quantification: 'a book' is expressed as **một quyển sách** (one CL book) and 'some books' as **một số sách** (one number book). Majority quantification is expressed using the compound **phần lớn** which decomposes into **phần** 'part' and **lớn** 'large' and means 'most' or 'the majority of'. The complement of **phần lớn** can be a bare noun or a plural definite.

- (8) **phần lớn (các con) chó thích ăn thịt**
 most PL CL dog like eat meat
 'most (of the) dogs like to eat meat'

Universal quantification is expressed in a variety of ways. One involve a fronted NPI, which is morphologically identical to a which-phrase, in conjunction with the particle **cũng** 'also' at the auxiliary position. The meaning of this strategy is close to 'every' in English.

- (9) **con chó nào cũng thích ăn thịt**
 CL dog which also like eat meat
 'every dog likes to eat meat'

Another strategy is to use the word **tất cả**, which combines with a bare noun or a plural definite, in conjunction with the word **đều** at the auxiliary position. The meaning of **tất cả** is close to 'all' in English. I will not attempt to translate **đều**, which I will gloss as DEU.

- (10) **tất cả (các con) chó đều thích ăn thịt**
 all PL CL dog DEU like eat meat
 'all (of the) dogs like to eat meat'

Finally, the meaning of 'each' is expressed by **mỗi**, which takes a singular definite as complement.

- (11) **mỗi con chó cắn hai người**
 each CL dog bite two person
 'each dog bit two persons'

1.5.2. Adverbial quantification

Quantification can also be effected by adverbs in Vietnamese. The domain is expressed by a bare noun. Thus, we observe quantificational variability effects similar to those in English, as indicated by the translations.

- (12) a. **chó luôn sủa ban đêm**
dog always bark at night
'dogs always bark at night' (i.e. all dogs bark at night)
- b. **chó thường sủa ban đêm**
dog often bark at night
'dogs often bark at night' (i.e. most dogs bark at night)
- c. **chó không bao giờ sủa ban đêm**
dog never bark at night
'dogs never bark at night' (i.e. no dogs bark at night)

2. Methods

I am a native speaker of Vietnamese, and completed the toolkit myself. I used all five story boards: (i) 'Banana', (ii) 'Lion', (iii) 'Canadian', (iv) 'Car', and (v) 'Water', in that order. I completed all five story boards in one session which lasted about an hour. There were no stories which posed any difficulty for me. I did not consult other speakers, as I felt my judgements were robust enough.

3. Results

3.1. Kind reference

Reference to kinds can be made by using either a bare noun or a singular definite. These options are naturally translated with bare plurals or singular definites in English (Carlson 1977, Chierchia 1998).

- (13) a. **chuối được thuần hoá đầu tiên tại Papua New Guinea**
banana passive cultivated first in Papua New Guinea
'bananas were first cultivated in Papua New Guinea'
- b. **quả chuối được thuần hoá đầu tiên tại Papua New Guinea**
CL banana passive cultivated first in Papua New Guinea
'the banana was first cultivated in Papua New Guinea'
- (14) a. **sư tử là một loài mèo lớn**
lion be one species cat large
'lions are a large cat'
- b. **con sư tử là một loài mèo lớn**
CL lion be one species cat large
'the lion is a large cat'
- (15) a. **Người Trung Quốc phát minh ra la bàn**
person China invent compass
'the Chinese invented compasses'
- b. **Người Trung Quốc phát minh ra cái la bàn**
person China invent CL compass
'the Chinese invented the compass'

Note that some nouns, i.e. those that are not only syntactically but also semantically mass, do not select for classifiers. In these cases kind reference is made by using the bare noun only.

- (16) **nước xuất hiện trên Trái Đất khoảng 4.4 tỷ năm trước**
water appear on Earth around 4.4 billion year ago
'water appeared on Earth around 4.4 billion years ago'

Nationalities are expressed as **người** 'person' + name of country, as evidenced above by (15). As **người** is a count noun which does not select for classifiers, kind reference with respect to nationalities is made by using this same phrase.

- (17) **người Nhật có họ với người Trung Quốc**
 person Japan related to person China
 'The Japanese are related to the Chinese'

Singular and plural indefinites, as well as plural definites cannot be used to refer to kinds, and neither can plural definites.

- (18) a. ***một quả chuối được thuần hoá đầu tiên tại PNG**
 one CL banana passive cultivated first in PNG
 ('bananas were first cultivated in PNG')
 b. ***một số chuối được thuần hoá đầu tiên tại PNG**
 one number banana passive cultivated first in PNG
 ('bananas were first cultivated in PNG')
 b. ***những quả chuối được thuần hoá đầu tiên tại PNG**
 PL CL banana passive cultivated first in PNG
 ('bananas were first cultivated in PNG')

3.2. Subkind reference

Reference to subkinds can be made using the nouns **loại** 'type' for inanimate objects and **loài** or **giống** 'species' for animate objects.

- (19) a. **một số loài sư tử đã tuyệt chủng**
 one number species lion PERF extinct
 'some species of lions have gone extinct'
 b. **mỗi quốc gia trồng một giống chuối**
 each country grow one species banana
 'each country grows a species of banana'
 c. **John phát minh ra hai loại máy tính**
 John invent two type computer
 'John invented two kinds of computers'

When the nominal contains a demonstrative article, subkind reference is possible with both such nouns and without.

- (20) a. **(loài) sư tử này đã tuyệt chủng**
 (species) lion this PERF extinct
 'this species of lions have gone extinct'
 b. **(giống) chuối này chỉ có ở Châu Phi**
 (species) banana this only have in Africa
 'this kind of bananas exists only in Africa'

Note that the sequence bare noun + demonstrative must refer to a subkind. Under the definite reading it is ungrammatical. Bare nouns can also receive subkind interpretation in quantified structures, as evidenced by the sentences in (21).

- (21) a. **chó nào cũng thích ăn thịt**
 dog which also like eat meat
 'all species of dogs like to eat meat'
 b. **phần lớn chó thích ăn thịt**
 most dog like eat meat
 'most species of dogs like to eat meat'

Singular and plural indefinites, as well as plural definites, cannot refer to subkinds.

- (22) a. ***Apple và IBM là hai cái máy tính**
 Apple and IBM be two CL computer
 ('Apple and IBM are two kinds of computers')
- b. ***Apple và IBM là một số máy tính**
 Apple and IBM be one number computer
 ('Apple and IBM are kinds of computers')
- c. ***Apple và IBM là những cái máy tính**
 Apple and IBM be PL CL computer
 ('Apple and IBM are kinds of computers')

3.3. Generic reference

Similar to English, Vietnamese uses bare nouns express generics (Krifka et al. 1995; Chierchia 1998).

- (23) a. **sư tử có bờm**
 lion have mane
 'lions have manes'
- b. **nước luôn cần thiết cho cơ thể**
 water always necessary for body
 'water is always necessary for the body'
- b. **người Nhật thường chăm chỉ**
 person Japan usually diligent
 'Japanese are usually diligent'
- c. **chuối không bao giờ chua**
 banana never sour
 'bananas are never sour'

Singular and plural indefinites, as well as definites, cannot be generic.

- (24) a. ***một con sư tử có bờm**
 one CL lion have mane
 ('a lion has manes')
- b. ***một số sư tử có bờm**
 one number lion have mane
 ('lions have manes')
- c. ***con sư tử có bờm**
 CL lion have mane
 ('lions have manes')
- d. ***những con sư tử có bờm**
 PL CL lion have mane
 ('lions have manes')

3.4. Token/object reference in habituals

In habitual sentences reference to tokens/objects is made by using bare nouns or indefinites.

- (25) a. **sáng nào Nam cũng ăn chuối**
 morning which Nam also eat banana
 'Nam eats bananas every morning'
- b. **sáng nào Nam cũng ăn một quả chuối**
 morning which Nam also eat one CL banana
 'Nam eats a banana every morning'
- c. **sáng nào Nam cũng ăn một số chuối**
 morning which Nam also eat one number banana
 'Nam eats some bananas every morning'

If definites are used in these sentence, the sentences would convey the odd message that that Nam eats the same banana or the same set of bananas every morning.

3.5. Token/object reference in episodic sentences

In episodic sentences reference to tokens/objects is made by using any kind of nominals.

- (26) a. **sáng hôm qua Nam ăn chuối**
morning which Nam eat banana
'Nam ate bananas yesterday morning'
- b. **sáng hôm qua Nam ăn một quả chuối**
morning which Nam eat one CL banana
'Nam ate a banana yesterday morning'
- c. **sáng hôm qua Nam ăn một số chuối**
morning yesterday Nam eat one number banana
'Nam ate some bananas yesterday'
- d. **sáng hôm qua Nam ăn quả chuối**
morning which Nam eat CL banana
'Nam ate the banana yesterday morning'
- e. **sáng hôm qua Nam ăn những quả chuối**
morning which Nam eat PL CL banana
'Nam ate the bananas yesterday morning'

3.6. Summary

		kinds	subkinds	generics	tokens habitual	tokens episodic
bare nouns		yes	yes (with demonstrative)	yes	yes	yes
indefinite	singular	no	no	no	yes	yes
	plural	no	no	no	yes	yes
definite	singular	yes	no	no	no	yes
	plural	no	no	no	no	yes

4. Conclusion

This study has examined how Vietnamese encodes kind, subkind, generic, and token reference across a range of nominal forms. The central empirical result is the strong functional specialization of bare nouns. Bare nouns are the unmarked strategy for kind reference, generic statements, and habitual token reference, and they remain fully available for episodic token reference as well. In contrast, indefinite and definite nominals are sharply restricted in their ability to express kinds and generics. Singular definites can denote kinds, but they cannot express generics or habitual token reference, while plural definites are confined to episodic token reference. Indefinites pattern similarly: they are excluded from kind and generic readings but freely available for token interpretations.

These restrictions suggest that Vietnamese grammar draws a principled distinction between nominals that introduce discourse referents and nominals that denote properties or kinds. Bare nouns appear to be semantically number-neutral and type-flexible, allowing them to shift between property, kind, and existential interpretations depending on clausal aspect and operator structure. Classifier-marked nominals, by contrast, are more tightly anchored to individuated object reference. The impossibility of using indefinites or plural definites for generics indicates that genericity in Vietnamese is not merely a pragmatic inference

layered on existential structure, but is grammatically encoded through a dedicated nominal form: the bare noun. Note that inspection of the individual storyboards revealed no systematic differences between the semantic classes represented by the nouns. The patterns observed for natural kinds (e.g. banana, lion), artifacts (e.g. car), and social kinds (e.g. Canadian) were uniform with respect to the availability of bare nouns, classifier-marked definites, and indefinites across the various interpretation types.

More broadly, the Vietnamese system supports the view that classifier languages can maintain a rich distinction between kind-level and object-level reference without articles in the Indo-European sense. The interaction between classifiers, definiteness marking, and number neutrality yields a tightly constrained mapping between form and interpretation. Future work should test the robustness of these patterns across dialects, speakers, and discourse contexts, and should investigate how Vietnamese fits into cross-linguistic typologies of genericity and kind reference. The present results provide a descriptive baseline and suggest that Vietnamese offers particularly clear evidence for a grammar in which bare nominals occupy the semantic default position for reference to kinds and genericity.

References

- Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages Counting and beyond. In Numeral Types and Changes Worldwide, edited by Jadranka Gvozdanović, pp. 113–186. New York: De Gruyter Mouton.
- Carlson, Greg. 1977. Reference to kinds in English. PhD dissertation, University of Massachusetts, Amherst.
- Chierchia, Gennaro. 1998. Reference to Kinds across Language. Natural Language Semantics 6: 339–405.
- Dayal, Veneeta. 2004. Number Marking and (in)Definiteness in Kind Terms. Linguistics and Philosophy 27: 393–450.
- Manfred Krifka, Francis Jeffry Pelletier, Gregory Carlson, Alice ter Meulen, Gennaro Chierchia & Godehard Link. 1995. Genericity: An Introduction. In The Generic Book, edited by Greg N. Carlson & Francis Jeffry Pelletier, pp. 1–124. Chicago: University of Chicago Press.
- Nguyễn Tài Cẩn. 1975. Ngữ pháp tiếng Việt. Hà Nội: Nhà xuất bản Đại học và Trung học chuyên nghiệp.
- Sauerland, Uli. 2003 A new semantics for number. Proceedings of SALT 13:258–275.
- Sauerland, Uli, Anderssen, Jan and Yatsushiro, Kazuko. 2005. The Plural is Semantically Unmarked. In Linguistic Evidence: Empirical, Theoretical and Computational Perspectives, edited by Stephan Kepser and Marga Reis, pp. 413–434. Berlin, New York: De Gruyter Mouton.
- Spector, Benjamin. 2007. Aspects of the Pragmatics of Plural Morphology: On Higher-Order Implicatures. In Presupposition and Implicature in Compositional Semantics, edited by Uli Sauerland & Penka Stateva, pp. 243–281. London: Palgrave Macmillan.
- Zweig, Eytan. 2009. Number-neutral bare plurals and the multiplicity implicature. Linguistics and Philosophy 32: 353–407.