Chapter 2

(In)definiteness and Vietnamese classifiers

Walter Bisang
Johannes Gutenberg University, Mainz & Zhejiang University

Kim Ngoc Quang
Johannes Gutenberg University, Mainz & University of Social Sciences and Humanities, Ho Chi Minh City

Vietnamese numeral classifiers (CL) in the bare classifier construction [CL+N] can be interpreted as definite and as indefinite. Based on a corpus of written and oral texts with a broad range of different contexts for the potential use of classifiers, this paper aims at a better understanding of the factors and linguistic contexts which determine the use of the classifier in [CL+N] and its specific functions. The following results will be presented: (a) Even though classifiers tend to be interpreted as definite, they are also used as indefinites, irrespective of word order (subject/preverbal or object/postverbal). (b) There is a strong tendency to use the [CL+N] construction with definite animate nouns in the subject position, while bare nouns [N] preferably occur with indefinite inanimate nouns in the object position. (c) The vast majority of nouns occurring with a classifier are sortal nouns with the features [–unique, –relational]. (d) Discourse and information structure are the most prominent factors which determine the grammar of Vietnamese classifiers. The influence of discourse is reflected in the pragmatic definiteness expressed by the classifier. Moreover, information structure enhances the use of a classifier in contexts of contrastive topic, contrastive focus and focus particles. Finally, thetic statements and some special constructions (existential clauses, verbs and situations of appearance) provide the environment for the indefinite interpretation of classifiers.

Walter Bisang & Kim Ngoc Quang. 2020. (In)definiteness and Vietnamese classifiers. In Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr. (eds.), *Nominal anchoring: Specificity, definiteness and article systems across languages*, 15–49. Berlin: Language Science Press. DOI: 10.5281/zenodo.4049679

1 Introduction

Numeral classifiers are an areal characteristic of East and mainland Southeast Asian languages in the context of counting. This fact is well known and has been frequently discussed in the literature since the 1970s (Greenberg 1972). What is less well known and has been discussed only in more recent times is the use of the same classifiers in the contexts of definiteness and indefiniteness, when they occur in the [CL+N] construction (bare classifier construction; cf. Bisang 1999; Cheng & Sybesma 1999; Simpson 2005; Wu & Bodomo 2009; Li & Bisang 2012; Jiang 2015; Simpson 2017; Bisang & Wu 2017). In Vietnamese, classifiers in [CL+N] are clearly associated with reference. What is controversial in the literature is the question of whether they are used only in the context of definiteness or in contexts of definiteness and indefiniteness. Tran (2011) claims that classifiers only have a definite interpretation, while Nguyen (2004) argues for both interpretations (see also Trinh 2011). A look at an example from Nguyen (2004) in (1) shows that both interpretations are possible. In this respect, it differs significantly from many Sinitic languages with [CL+N] constructions. While the definiteness/indefiniteness interpretation of classifiers depends on the preverbal or postverbal position of the [CL+N] construction in most of these languages,[1] Vietnamese classifiers can have both interpretations in both positions. In (1a), *con bò* [CL cow] is in the subject position and is open to both interpretations ('the cow'/'a cow'). Similarly, *cuốn sách* [CL book] in the object position of (1b) can be definite as well as indefinite ('the book'/'a book'):

(1) Nguyen (2004):

a. Con bò ăn lúa kìa!
CL cow eat paddy SFP
'Look! A/the cow is eating your paddy!'

b. Mang cuốn sách ra đây!
bring CL book out here
'Get a/the book!'

As can be seen from the following example, nouns without a classifier (bare nouns) can also be interpreted in both ways in both positions. In Nguyen's (2004) analysis, the only difference between the bare noun construction and the [CL+N]

[1] In Wang's (2015) survey of Sinitic classifiers as markers of reference, the definiteness/indefiniteness distinction is independent of word order relative to the verb in only 10 out of his 120 sample languages (cf. Type I classifiers in his terminology).

construction is that the former can be interpreted as singular or plural, while the latter can only have a singular reading:

(2) Nguyen (2004):

a. Bò ăn lúa kìa!
cow eat paddy SFP
'Look! A/the cow(s) is/are eating your paddy!'

b. Mang sách ra đây!
bring book out here
'Get a/the book(s), will you?'

Even though these examples show that classifier use is not obligatory and that classifiers can be interpreted as definite as well as indefinite, neither the conditions under which classifiers have these functions nor their specific referential meaning are well understood. Analyzing Vietnamese classifiers in the [CL+N] construction as variables whose interpretation depends on semantic, syntactic and discourse-pragmatic contexts, it is the aim of this paper to define the contexts which determine their use in terms of obligatoriness and their interpretation as definite and indefinite. Since the use of the classifier in the [CL+N] construction and its interpretation in terms of (in)definiteness in Vietnamese strongly depends on discourse and information structure, as in many other East and mainland Southeast Asian languages, looking at individual examples in isolation is not sufficient for modeling the function and the use of the [CL+N] construction. What is needed are texts, both written and oral. For that reason, we decided to set up our own corpus of Vietnamese, which is based on written and oral reports, by native speakers of Vietnamese, on the content of two silent movies (for details, cf. Section 2).

The analysis of the data from our Vietnamese corpus confirms the general observation that classifiers in [CL+N] can be interpreted as definite as well as indefinite, irrespective of word order. It also shows that the interpretation of numeral classifiers in terms of definiteness and indefiniteness in [CL+N] depends on semantic and syntactic (preverbal/postverbal or subject/object) factors, as well as on discourse and information structure. In addition to that, it turns out that the definite function is much more frequent than the indefinite function. Instances of indefinite [CL+N] constructions mainly occur in special contexts and constructions, such as thetic statements, existential clauses, and constructions with verbs which introduce previously unidentified referents into discourse (i. e., verbs of appearance). Given the relative rareness and the functional specifics of indefinite

classifiers, it may not come as a surprise that the indefinite interpretations of classifiers remained unnoticed in a number of studies of Vietnamese classifiers.

To find out more about the function of classifiers and the factors that determine their use, the following criteria will be studied in more detail:

- Definiteness and indefiniteness of the nominal expression
- The semantic feature of [±animate] of the noun
- The semantic features of [±unique] and [±relational] in terms of Löbner's (1985; 2011) four basic types of nouns
- The syntactic criterion of word order (position of the noun in the subject/preverbal or object/postverbal position)
- The role of discourse and the relevance of identifiability
- Information structure and the use of a classifier (contrastive topics, certain types of focus) as well as its function (i. e., theticity and indefiniteness)
- The combination with specific verbs (i. e., existential verbs and verbs of appearance).

The structure of the paper is as follows: after the discussion of methodological issues in Section 2, Section 3 will describe classifiers in their definite functions and the criteria that determine their use. Section 4 will do the same with classifiers in their indefinite function. The conclusion in Section 5 will briefly summarize our findings and situate them with regard to other languages with numeral classifiers that are used in contexts of definiteness as well as indefiniteness.

2 Methodology

Our analysis of the function and the use of classifiers in the [CL+N] construction is based on a Vietnamese corpus of 30 written texts and 30 oral texts produced by native speakers of Vietnamese who were asked to report on the content of two films which were previously presented to them on the screen of a personal computer. One of the films was used to create a written corpus, the other an oral corpus. The total number of informants involved was 46 (25 female and 21 male informants). Fourteen informants (five female and nine male) from among these 46 informants participated in both experiments and thus produced a written and an oral text.[2] In total, there were 15 male and 15 female informants, as well as 15

[2]Since it was more difficult to find male informants, we had to ask more males to take part in both experiments. Six of the remaining 12 male informants only produced a written text, while the other six only were involved in the oral experiment.

graduate and 15 undergraduate informants for each experiment. The reason for this arrangement was to check for potential effects originating from differences in gender or modality (written vs. oral). Since we did not find any significant differences, we will not address this issue in the present paper.

The experiments were carried out by Kim Ngoc Quang in Ho Chi Minh city (Southern Vietnam) with the support of assistants who played the role of addressees (readers/listeners). This arrangement was necessary to avoid speaker assumptions about information shared with the addressee. Thus, the informants reported their stories in a situation in which it was clear that the addressee did not know the story.

For the purpose of our study, we needed two films with multiple protagonists, frequently changing scenes with different perspectives and a large number of animate and inanimate objects involved in a variety of actions expressed by transitive and intransitive verbs. The first film, with the title 'Cook, Papa, Cook', is a silent movie of nine minutes and 38 seconds in length.[3] This very lively film, which was used to create the written corpus, has three protagonists: a husband, a wife and their son. The story is characterized by intense quarrels between the husband and his wife. Because of this, the wife decides that she is no longer prepared to make breakfast for her husband. His attempts to make it himself are met by a number of obstacles and end up turning the kitchen into a total mess. When he finally manages to make his own kind of breakfast, his wife refuses to eat it.

The second film, which was used to set up the oral corpus, is from the 'Pear Stories' (Chafe 1980).[4] It is five minutes and 54 seconds long. It has two protagonists: a farmer and a young boy, who steals the farmer's pears from some baskets, while the farmer is up a tree picking the rest of the pears. When cycling away from the farmer, he inadvertently rides over a stone because he is distracted by a girl cycling in the opposite direction. As a consequence, the pears roll out of the basket and scatter all over the road. Three other boys arrive and help the boy to pick up the pears. As a reward for their help, the boy offers them each a pear. Later on, the three boys walk past the farmer while eating their pears. The film ends with the farmer trying to understand what has happened.

The length of the 30 written texts varies between 491 and 1,944 words. The written corpus as a whole consists of 31,663 words. The total length of the oral corpus is 17,777 words, after transcription. The length of the 30 oral texts varies between 321 and 1,061 words.

In this paper, the two corpora are employed as sources of examples of a broad range of different classifier functions and different conditions responsible for

[3]The film can be seen on YouTube at https:://www.youtube.com/watch?v=OITJxh51z3Q.

[4]The film can be seen on YouTube at https:://www.youtube.com/watch?v=bRNSTxTpG7U.

their occurrence. Moreover, the data from these corpora are used for some generalizations about frequency, as far as that is possible on the basis of calculating simple percentages.

3 Classifiers and definiteness

This section examines the correlation between classifiers with a definite interpretation in the [CL+N] construction[5] from various perspectives. §3.1 discusses the semantic feature of animacy and its interaction with definiteness. An examination of the semantic features of uniqueness and relationality in §3.2 shows that the vast majority of nouns occurring with a classifier are sortal nouns, defined by their features of [–unique]/[–relational]. The interaction of word order (preverbal/subject and postverbal/object) with animacy and definiteness is explored in §3.3. Finally, the roles of discourse (identifiability) and information structure (contrastive topics, focus particles and contrastive focus) are discussed in §3.4.

3.1 Animacy and definiteness

Animacy plays an important role in grammar. This can be clearly seen from the animacy hierarchy as introduced by Silverstein (1976) and Dixon (1979), which is involved in such divergent domains of grammar as alignment, differential object marking, direct/inverse marking and number marking on nouns (to name just a few). An examination of this hierarchy in its full form, as it is presented in Croft (2003: 130), shows that it is not only concerned with animacy but also with person and referentiality.

(3) Animacy hierarchy (Croft 2003: 130):
first/second person pronouns > third person pronoun > proper names > human common noun > non-human animate common noun > inanimate common noun.

The role of animacy in a strict sense is limited to the animacy scale, which goes from human to animate to inanimate. Animacy generally contributes to prominence (for a good survey, cf. Bornkessel-Schlesewsky & Schlesewsky 2009). Another important scale that contributes to prominence is the definiteness scale that runs from personal pronoun to proper name, to definite NP, to indefinite specific NP, to non-specific NP (cf. Aissen 2003, on the relevance of these two

[5]Notice that we do not discuss instances of [NUM CL N] with numerals > 1 because we do not have enough data in our corpus.

scales for differential object marking). As will be shown in this subsection, based on the Vietnamese data from our experiments, both scales have their impact in the use of classifiers inasmuch as there is a strong tendency for classifiers to be used with definite animate nouns.

As for animacy, Table 1 below shows a clear correlation between the feature of [±animate] and classifier use. Out of 1,698 instances with animate nouns, 1,571 instances[6] (92.5%) take a classifier, while only 127 instances[7] (7.5%) occur without a classifier. In contrast, only 742 instances[8] of [–animate] nouns (27.6%) occur with a classifier, while 1,948 instances[9] (72.4%) are bare nouns.[10]

Table 1: Token frequency of classifier use with [±animate] nouns in written texts and oral texts (in our Vietnamese corpus)

Nouns in narratives	**[+animate]**	**[–animate]**
[CL+N]	1,571 instances (92.5%)	742 instances (27.6%)
[N]	127 instances (7.5%)	1,948 instances (72.4%)

Our Vietnamese data also show that classifiers can be interpreted as definite as well as indefinite but that there is a strong tendency towards definite interpretation in our written and in our oral corpus. This can be seen from Table 2, in which 1,444 instances of [CL+N] in the written corpus are definite (92.0%; 1,154 + 290), while only 125 instances are indefinite (8.0%; 22 + 103). Similarly, the oral corpus shows 680 instances of classifiers in their definite function (91.4%; 395 + 285), which contrast with only 64 classifiers with an indefinite reading (8.6%; 0 + 64). The same table additionally shows that definiteness clusters with animacy. In the written corpus, 1,154 animate definite nouns with a classifier (90.4%) contrast with only 122 animate definite nouns with no classifier (9.6%). In the case of oral texts, animate definite nouns reach an even higher percentage: 100% of these

[6] 1,571 is the result of all [+animate] nouns with a classifier in the written corpus (978 + 19 + 176 + 3) plus all [+animate] nouns with a classifier in the oral corpus (262 + 0 + 133 + 0) in Table 6.

[7] 127 is the result of all [+animate] nouns with no classifier in the written corpus (8 + 1 + 114 + 0) plus all [+animate] nouns with no classifier in the oral corpus (0 + 1 + 3 + 0) in Table 6.

[8] 742 is the result of all [–animate] nouns with a classifier in the written corpus (34 + 9 + 256 + 94) plus all [–animate] nouns with a classifier in the oral corpus (55 + 2 + 230 + 62) in Table 6.

[9] 1,948 is the result of all [–animate] nouns with no classifier in the written corpus (78 + 31 + 1,092 + 365) plus all [–animate] nouns with no classifier in the oral corpus (12 + 0 + 324 + 46) in Table 6.

[10] The frequencies of classifier use in the tables in this paper are for those occurrences in [CL+N] constructions; hence sequences such as *hai cuốn sách* [two CL book] 'two books' would not be counted in these tables.

nouns take a classifier. As for inanimate definite nouns, only 19.9% of the written corpus (290 out of a total of 1,460) and 45.9% of the oral corpus (285 out of 621) take a classifier.

Table 2: Token frequency of [±animate] nouns and their interpretation as definite and indefinite in written texts and oral texts (in our Vietnamese corpus)

	Written texts				**Oral texts**			
	[+animate]		**[–animate]**		**[+animate]**		**[–animate]**	
	[CL+N]	**[N]**	**[CL+N]**	**[N]**	**[CL+N]**	**[N]**	**[CL+N]**	**[N]**
Definite	**1,154**	**122**	**290**	**1,170**	**395**	**3**	**285**	**336**
Indefinite	22	1	103	396	0	1	64	46
Total	1,176	123	393	1,566	395	4	349	382

The following two examples illustrate the use of animate nouns with a classifier. In (4), the classifier occurs with one of the human protagonists of the story, who is clearly identifiable and definite at the point at which he is mentioned in that example. In example (5), the classifier is interpreted as indefinite. The animate noun *dê* 'goat' is introduced into the story.[11] As will be seen later in §4.2, the co-occurrence with the copula verb *là* 'to be' is one of the typical contexts in which [CL+N] is interpreted as indefinite (cf. example 31):

(4) [+animate, +CL, +DEF] (Oral text 19, sentence 12)

Cậu bé thấy thế tặng mỗi người một trái lê …
CL boy see that present each person one CL pear
'The boy saw that he gave each of them one pear.'

(5) [+animate, +CL, –DEF] (Oral text 16, sentence 7)

Có một người dẫn con, con đó chắc là **con dê**, đi ngang qua.
have one person lead CL CL DEM maybe CL goat go pass over
'There was a man who led a, a, it may be a goat, passing by.'

[11]Notice, however, that in the continuation of this text, the goat is further specified as a *dê núi* [goat mountain] 'wild goat' and does not take a classifier. With this type of compound, classifiers are often omitted.

The following example shows how inanimate nouns tend to be realized as bare nouns, even if they are definite. The referents expressed by *thang* 'ladder' and *cây* 'tree' have already been mentioned but do not have classifier marking:[12]

(6) [–animate, –CL, +DEF] (Oral text 27, sentence 3)

Sau đó, ông ấy lại leo lên **thang** và leo lên
after that 3.SG again climb PREP ladder CONJ climb PREP

cây hái tiếp.
tree pluck continue

'After that, he [the farmer] climbed up the ladder and climbed onto the tree again to continue picking [pears].'

The comparatively less frequent combination of inanimate nouns with classifiers is illustrated by the following two examples:

(7) [–animate, +CL, +DEF] (Written text 2, sentence 8)

Cậu ta đã đặt **cái xô** ngay giữa bố và mẹ.
3.SG PERF place CL bucket right between father and mother

'He put the bucket right between his father and mother.'

(8) [–animate, +CL, –DEF] (Written text 24, sentence 14)

Lúc này, người đàn ông thức dậy, lấy **cái bình** rót nước vào ly,
time DEM CL man wake up take CL bottle pour water PREP glass

'At this time, the man woke up, he took a bottle and poured water into a glass,'

In (7), the inanimate noun *xô* 'bucket' was previously introduced into the scene by one of the protagonists (the boy). Given that the bucket is activated in the hearer's mind, the classifier marks definiteness in this example. In (8), the noun *bình* 'bottle' refers to a newly introduced concept. Thus, the classifier *cái* marks indefiniteness in this context.

The relationship between animacy/definiteness and word order (the position of the [CL+N] construction relative to the preverbal and postverbal positions) will be discussed in §3.3.

[12] One of our reviewers asks if *thang* 'ladder' and *cây* 'tree' may be analyzed as instances of incorporation into the verb plus preposition. Given that both referents represented by these nouns can be clearly identified from their previous mention as individuated countable concepts in the text, such an analysis does not seem to be very likely.

3.2 The semantic features of uniqueness and relationality

The distinction between ±relational[13] and ±unique[14] nouns as discussed by Löbner (1985; 2011) is of crucial importance for describing the use of classifiers in Vietnamese. The combination of these features with their two values yields the following four basic types of nouns, which correspond to four types of concepts or four logical types: sortal nouns ([–relational]/[–unique]; ⟨e,t⟩), individual nouns ([–relational]/[+unique]; ⟨e⟩), relational nouns ([+relational]/[–unique]; ⟨e,⟨e,t⟩⟩) and functional nouns ([+relational]/[+unique]; ⟨e,e⟩).

Table 3 presents our data on the presence or absence of classifiers in the context of Löbner's (1985; 2011) basic types of nouns. As can be seen, the vast majority of nouns occurring with a classifier are sortal nouns ([–unique]/[–relational]): out of a total of 2,313 nouns with a classifier, 2,309 (99.8%) belong to this type. Moreover, only three [+unique] nouns (marked in bold) out of 108 (2+83+1+22) take a classifier (2.8%), while 105 of them are realized as bare nouns (97.2%). In a similar way, relational nouns ([–unique]/[+relational]) have a strong tendency to occur without a classifier. Only one out of a total of 57 instances of this type (1.8%) takes a classifier.

Table 3: Token frequency of classifier with [±relational], [±unique] nouns in written texts and oral texts (in our Vietnamese corpus)

[±relational],	**[+relational]**				**[–relational]**			
[±unique] nouns	**[+unique] functional**		**[–unique] relational**		**[+unique] individual**		**[–unique] sortal**	
	[CL+N]	**[N]**	**[CL+N]**	**[N]**	**[CL+N]**	**[N]**	**[CL+N]**	**[N]**
Written texts	2	76	0	48	0	2	1,567	1,563
Oral texts	0	7	1	8	1	20	742	351
Total	**2**	83	**1**	56	**1**	22	**2,309**	**1,914**

From the four non-sortal nouns with a classifier, two are used in anaphoric situations. In example (9), the [+unique/+relational] noun *môn*g 'buttocks' is first introduced into the story by a bare noun. The second time it is mentioned, the same noun occurs with the general classifier *cái*, its interpretation being definite because the object it denotes is now activated in the hearer's mind:

[13]Relational nouns have not only a referential argument, but also an additional relational argument (cf. the relational noun *daughter [of someone]* in contrast to the absolute noun *girl*).

[14]Unique nouns denote concepts which are uniquely determined in a given situation (e. g., *the sun, the pope*). Notice that the default use of uniqueness is singular definite. Plural, indefinite and quantificational uses require special marking.

(9) (Written text 1, sentence 45)

Bị nóng **mông**, anh ta mở vòi-nước xịt mát cho **cái mông**,
PASS hot buttock 3.SG open water-tap spray cool for CL buttock

thì lúc đó, bạn anh ta chồm từ ngoài cửa sổ vào hối
CONJ time DEM friend 3.SG prance from outside window in urge

anh ta nhanh-lên kẻo trễ giờ.
3.SG hurry-up otherwise late

'[His] buttocks were burnt, he turned on the tap and sprayed cool water onto the buttocks, at that time, his friend gesticulated from outside of the window to urge him to hurry up as otherwise he would be late.'

A similar pattern is found in example (10) with the [–unique/+relational] noun *chân* 'leg', which is expressed by a bare noun when it is first mentioned. Later on, it is taken up together with the general classifier *cái* expressing definiteness in this context:

(10) (Oral text 4, sentence 21)

Lê đổ ra tung toé, hình như nó bị đau **chân** nữa, thấy nó
pear pour out everywhere seems 3.SG PASS hurt leg more see 3.SG

sờ sờ **cái chân**.
touch touch CL leg

'The pears rolled out everywhere, it seemed that his leg was hurt, (because I saw) he touched [his] leg.'

In the other two instances of the [CL+N] construction with a non-sortal noun, the use of the classifier is due to information structure (focus). For that reason, the relevant examples will be discussed in §3.4.3 (cf. (23) and (25)).

3.3 Word order, definiteness and animacy

In many Sinitic numeral classifier systems, the referential status associated with the classifier in [CL+N] constructions depends on word order relative to the verb (see Wang 2015 for a survey). The following examples in (11) and (12) from Li & Bisang (2012) show how the preverbal subject position and the postverbal object position are associated with definiteness and indefiniteness in Mandarin, in the Wu dialect of Fuyang and in Cantonese.

While the [CL+N] construction in the subject position is ungrammatical in Mandarin Chinese (11a), it is interpreted in terms of definiteness in the Wu dialect of Fuyang (11b) and in Cantonese (11c).

(11) [CL+N] in the subject position (Li & Bisang 2012: 338)
Context: Where is the book?

a. Mandarin:

nà běn shū, **(*ge) xuéshēng** mǎi-zǒu le.
that CL book CL student buy-away PF
'The book, the student(s) has/have bought it.'

b. Wu Chinese:

pen cy ke ia?sn ma le tçhi die.
CL book, CL student buy PFV go SFP
'The book, the student bought (it).'

c. Cantonese:

bun syu, go hoksaang maai-jo la.
CL book CL student buy-PFV SFP
'The book, the student bought (it).'

In the object position, the classifier in [CL+N] is associated with indefiniteness in Mandarin (12a) and the Wu dialect of Fuyang (12b). In Cantonese, it goes with definiteness and indefiniteness (12c):

(12) [CL+N] in the object position (Li & Bisang 2012: 338-339)

a. Mandarin:

wǒ mǎi-le **liàng chē**.
I buy-PFV CL car
'I bought a car.'

b. Wu Chinese:

Nge ma le **bu ts^{h}ots^{h}i**.
I buy PFV CL car
'I bought a car.'

c. Cantonese:

Keuih maai-zo **gaa ce**.
he sell-PFV CL car
'I sold a car/the car.'

As can be seen from Table 4, the situation is different in Vietnamese. The [CL+N] construction occurs preverbally and postverbally and the classifier can be associated with definiteness as well as indefiniteness in both positions. A closer look reveals that the definite interpretation of the classifier is generally

preferred. The overall percentage of definite [CL+N] constructions is 91.8% in contrast to only 8.2% of classifiers with an indefinite function.[15] The dominance of the definite interpretation is even stronger in the subject position (cf. the figures printed in bold). If the written and oral texts are combined, 1,329 out of 1,359 [CL+N] constructions, or 97.8%, are definite.[16] In the object position, the asymmetry between the definite and the indefinite interpretation is not as strong as in the subject position. In spite of this, the definite interpretation still clearly dominates, with 795 (432 + 363) instances (83.3%), compared with only 159 (97 + 62) instances (16.7%) with an indefinite interpretation.[17]

Table 4: Token frequency of the presence/absence of a classifier in subject and object positions in relation to definite vs. indefinite function (in our Vietnamese corpus)

Nouns in	**Written texts**				**Oral texts**			
narrative	**Subject**		**Object**		**Subject**		**Object**	
	[CL+N]	**[N]**	**[CL+N]**	**[N]**	**[CL+N]**	**[N]**	**[CL+N]**	**[N]**
Definite	**1012**	86	432	1206	**317**	12	363	327
Indefinite	**28**	32	97	365	**2**	1	62	46
Total	1040	118	529	1571	319	13	425	373

The two examples in (13) and (14) illustrate the definite function of the classifier in [CL+N]. In (13), *con lừa* [CL donkey] 'the donkey' is in the subject position. Because it is mentioned in the previous context, the classifier *con* has a definite reading. In (14), *cô vợ* [CL wife] 'the wife' is in the object position. Since it is mentioned in the preceding text, it is also interpreted as definite:

(13) Definite subject (Oral text 26, sentence 9)

Con lừa cứ nhìn vào các
CL donkey always look inside PL

cần xé lê như muốn đứng lại và ăn lê.
CL pear like want stop CONJ eat pear

'The donkey kept on looking into the baskets as if it wanted to stand by and eat them.'

[15]The total number of definite [CL+N] constructions is 2,124 (1,012 + 432 + 317 + 363); the total number of indefinite [CL+N] constructions is 189 (28 + 97 + 2 + 62).

[16]The total number of definite [CL+N] constructions in subject position is 1,329 (1,012 + 317); the total number of indefinite [CL+N] constructions is 30 (28 + 2).

[17]Recall that bare nouns in Vietnamese can also occur in both subject positions and object positions and be interpreted as either definite or indefinite.

(14) Definite object (Written text 9, sentence 14)

Bực mình, anh chồng đóng sầm cửa khiến **cô vợ** giật mình,
angry CL husband slam door cause CL wife startled

rồi bỏ vào nhà tắm.
CONJ leave enter bathroom

'Annoyed, the husband slammed the door. This upset [his] wife, then he went to the bathroom.'

The following two examples focus on the object position and indefiniteness (for indefinite [CL+N] constructions in the subject position, cf. §4.1). At the same time, they also illustrate how classifiers in the same syntactic position can be interpreted as indefinite or definite, depending on context. In example (15) from our data on written texts, we find the same expression (*chiếc xe* [CL car] 'a/the car') in both functions.

(15) [Indefinite object, ±DEF] (Written text 1, sentence 26)

Anh ta bước vào nhà thì lại bị đứa con chơi **chiếc xe**$_1$
3.SG step PREP house CONJ EMPH PASS CL son play CL car

đẩy trúng vào chân khiến anh ta ngã ngửa vào **chiếc xe**$_2$.
push RES PREP leg cause 3.SG fall.back PREP CL car

'When he entered the house, he ran into his son who was playing and he got hit by a car [a toy car] into [one of his legs]. [This] made him fall down onto the car.'

In the first line, the noun *xe* 'car' in *chiếc xe* is not activated by previous context. Thus, the classifier must be interpreted as indefinite. In the second line, the same car is taken up again with the same classifier (*chiếc*), which now has a definite interpretation. The next example is from our oral corpus:

(16) [Indefinite object, ±DEF] (Oral text 26, sentence 1)

Có một người đàn ông đang ở trên **cái thang** bắc lên **cây**
exist one CL man PROG PREP top CL ladder connect PREP CL

lê và đang hái **trái lê**.
pear CONJ PROG pluck CL pear

'There was a man on [a] ladder which was propped up against [a] pear tree. He was picking [its] pears.'

In this example, we find three [CL+N] constructions, i. e., *cái thang* [$CL_{general}$ ladder], *cây lê* [CL_{tree} pear] and *trái lê* [CL_{fruit} pear]. Since the first two nomi-

nal concepts are newly introduced, the corresponding [CL+N] constructions are interpreted as indefinite ('a ladder' and 'a pear tree'). The third [CL+N] construction is associated with the previously mentioned pear tree. For that reason, the classifier *trái* for fruits can be interpreted as definite through bridging ('its pears [i. e., the pears of the previously mentioned tree]').

If the data on classifier use in the subject and in the object position is combined with the semantic feature of animacy as in Table 5, it can be seen that there is a clear preference for animate nouns in the subject position. There are 1,269 instances (85.2%) of [+animate] nouns in the subject position, which contrast with only 221 instances (14.8%) of [–animate] nouns. Similarly, the object position is characterized by its clear preference for [–animate] nouns. There are 2,469 [–animate] object nouns (85.2%) and only 429 [+animate] object nouns (14.8%). Thus, the data in Table 5 reflect the well-known preference of animate subjects and inanimate objects (cf. Givón 1979, Du Bois 1987 and many later publications).

Table 5: Distribution of instances of [±animate] nouns in the positions of subject and object (in our Vietnamese corpus)

	Subject		**Object**	
	[+animate]	**[–animate]**	**[+animate]**	**[–animate]**
Written texts	1,006	152	293	1,807
Oral texts	263	69	136	662
Total	1,269 (85.2%)	221 (14.8%)	429 (14.8%)	2,469 (85.2%)

Finally, the combination of the three parameters of word order (subject vs. object), reference (definite vs. indefinite) and animacy (animate vs. inanimate) yields the following results for the presence/absence of the classifier ([CL+N] vs. [N]):

Table 6 reveals that, of the 1,012 definite [CL+N] constructions in the subject position of the written text corpus, 978 (96.6%) are [+animate] nouns. Only 34 definite [CL+N] constructions in the subject position (3.4%) are [–animate]. Similarly in oral texts, 262 animate definite subject [CL+N] constructions (82.6%) contrast with only 55 inanimate definite subject [CL+N] constructions (17.4%). In the object position, the percentage of animate nouns with definite subject [CL+N] constructions is much lower: 40.7% (176 vs. 256) in the corpus of written texts and 36.6% (133 vs. 230) in the corpus of oral texts. The results from Table 6 combined with the results from Table 4 (general preference of definite classifier

Table 6: Presence/absence of classifiers depending on the features of [±animate], subject vs. object and definite vs. indefinite (in our Vietnamese corpus)

[+def] vs. [–def]	**Written texts**							
	Subject				**Object**			
	[CL+N] [±ani]		**[N] [±ani]**		**[CL+N] [±ani]**		**[N] [±ani]**	
	+	**–**	**+**	**–**	**+**	**–**	**+**	**–**
+def	**978**	**34**	8	78	176	256	114	1,092
–def	**19**	**9**	1	31	3	94	0	365
Total	**997**	**43**	9	109	179	350	114	1,457

[+def] vs. [–def]	**Oral texts**							
	Subject				**Object**			
	[CL+N] [±ani]		**[N] [±ani]**		**[CL+N] [±ani]**		**[N] [±ani]**	
	+	**–**	**+**	**–**	**+**	**–**	**+**	**–**
+def	**262**	**55**	0	12	133	230	3	324
–def	**0**	**2**	1	0	0	62	0	46
Total	**262**	**57**	1	12	133	292	3	370

interpretation, particularly with [CL+N] constructions in the subject position) plus Table 5 (preference of animate subjects) show that the classifier prototypically occurs with definite animate nouns in the subject position.

These observations can be visualized more clearly by means of the bar chart in Figure 1. The blue columns represent definiteness, while the green ones stand for indefiniteness:

In accordance with the data in Table 6, the blue columns representing definiteness are generally higher than the green columns, reflecting again the overall dominance of the definite function of Vietnamese classifiers. Moreover, the blue column in Figure 1 clearly dominates over the green column at the leftmost pole representing animate subjects with classifiers [Subj, +CL, +ani]. The preference

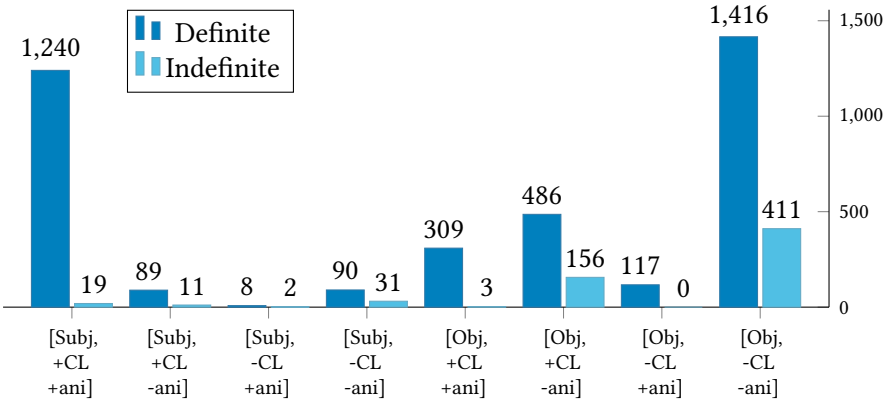

Figure 1: Token frequency of [±animate] nouns in subject and object function, marking definiteness or indefiniteness with or without a classifier (in our Vietnamese corpus)

for classifier use with animate subjects is further corroborated if the total number of tokens with the features [Subj, +CL, +ani] in the written and the oral corpus is compared with the total number of tokens with the features [Subj, −CL, +ani]. The figure for [Subj, +CL, +ani] is 1,259 (978 + 19 + 262 + 0), while the figure for [Subj, −CL, +ani] is just 10 (8 + 1 + 0 + 1). Thus, the use of the classifier with animate subjects overwhelmingly dominates over its absence with 99.2%. In addition to these results, the rightmost pole in Figure 1 with the features [Obj, −CL, −ani] demonstrates that inanimate object nouns tend to occur without a classifier. The overall number of tokens with the features [Obj, −CL, −ani] from the written and the oral texts is 1,827 (1,092 + 365 + 324 + 46), while the overall number of tokens with the features [Obj, +CL, −ani] is only 642 (256 + 94 + 230 + 62). Thus, the percentage of inanimate object nouns without a classifier is 74.0% against 26.0% with a classifier. Taken together, there is a clear preference for animate subjects to occur with a classifier and for inanimate objects to occur as bare nouns.

To conclude, the data presented in this subsection show that the (in)definiteness interpretation of the classifier is not rigidly determined by the position of the [CL+N] construction relative to the verb (subject vs. object position). In fact, there is an overall preference for interpreting classifiers in [CL+N] as definite even though indefinite [CL+N] constructions are found in both positions. In spite of this, there are other factors which operate against this general tendency as well as against the use of classifiers in definite contexts. The semantic factors were presented above in §3.2. §3.4 will discuss aspects of discourse and information structure.

3.4 Discourse and information structure

Discourse and information structure affect the meaning of Vietnamese classifiers as well as their presence or absence in a given context. As discussed in §3.4.1 on meaning, the definiteness expressed by the classifier is discourse-based. The same subsection also shows how discourse enhances the use of classifiers with [+unique] nouns which otherwise show a strong preference for occurring as bare nouns in our data (cf. §3.2). §3.4.2 and §3.4.3 illustrate how information structure determines the presence of a classifier. It will be shown that contrastive topics generally take a classifier (cf. §3.4.2). Similarly, focus, as it manifests itself in contrastive focus and focus particles, can support the use of a classifier, even with non-sortal nouns (§3.4.3).

3.4.1 Definiteness, identifiability and information structure

Classifiers in [CL+N] constructions very rarely occur with [+unique] nouns (cf. §3.2 on the strong preference for sortal nouns ([–unique]/[–relational])). Moreover, the majority of definite classifiers are used in anaphoric contexts, in which a previously introduced concept is taken up with a classifier in order to highlight the speaker's assumption that it can be identified by the hearer (cf. examples (4), (7), (13), (14) and (15)). Even two of the four non-sortal nouns with a classifier acquire their classifier in an anaphoric context (cf. (9) and (10); for the other two, cf. §3.4.3 on focus). Taken together, these facts are strong indicators that the definiteness expressed by the classifier marks pragmatic definiteness rather than semantic definiteness in terms of Löbner (1985). In Schwarz's (2009; 2013) framework, Vietnamese definite classifiers express anaphoric or "strong" definiteness rather than unique or "weak" definiteness.

With these properties, the definiteness associated with the classifier corresponds to the findings of Li & Bisang (2012: 17) on identifiability. As they show in example (17) from the Wu dialect of Fuyang, uniqueness is not a necessary condition for the definite interpretation of the [CL+N] construction. Unique concepts can be expressed either by bare nouns or by the [CL+N] construction. A [+unique] [–relational] noun like *thin* 'sky' in (17) occurs in its bare form if the sky is understood generically as the one and only one sky. Thus (17a) is a generic sentence expressing the fact that the sky is blue in general. In contrast, the classifier in *ban thin* [CL sky] (17b) indicates that the speaker means the sky as it is relevant for a given speech situation with its temporal or spatial index, and that s/he thinks that the hearer can identify it (Li & Bisang 2012: 17):

(17) Wu dialect of Fuyang (Li & Bisang 2012: 17):

a. Generic use:
Thin zi lan ko.
sky be blue SFP
'The sky is blue (in general).'

b. Episodic use:
Ban **thin** gintsɔ man lan.
CL_{piece} sky today very blue
'The sky is blue today.'

In Vietnamese, the situation seems to be similar. Since a much larger corpus than the two corpora used here would be needed to find examples like (17), we present another example from a Vietnamese dictionary in (18) (Nguyen et al. 2005: 116 and 1686). In (18a), we find *trời* 'sky' as a bare noun. In this form, the sky is understood generically as the endless outer space seen from the earth with its general property of being full of stars. In contrast, *bầu trời* [CL sky] 'the sky' in (18b) with a classifier denotes the inner space seen from the earth as it is currently relevant to the speech situation. The speaker employs the classifier to inform the hearer that s/he is referring to the sky as it currently matters and as it can be identified by the speaker and the hearer in a shared temporal or spatial environment.

(18) a. Trời đầy sao.
sky full star
'The sky is full of stars.'

b. **Bầu** trời đêm nay đầy sao.
CL sky tonight full star
'THE sky tonight is full of stars.'

Further evidence for the discourse-dependency of classifier use with [+unique] nouns comes from the fact that the noun *trời* 'sky' can take several different classifiers, e. g., *bầu trời* [CL_{round} sky], *khung trời* [CL_{frame} sky] or *vùng trời* [CL_{area} sky], etc. The selection of a specific classifier out of a set of possible classifiers depends on the particular property of the sky the speaker wants to highlight to facilitate its identifiability to the hearer. In such a situation, selecting a particular classifier is even compulsory:

(19) *(**Khung/bầu/vùng**) trời mơ ước của hai chúng ta đây rồi!
CL sky dream POSS two 2.PL here SFP
'Our dream sky/world is here!'

In the above example, the speaker creates a specific notion of the sky as it is relevant for her/him and the hearer. This 'dream sky' is then anchored in space and time as relevant to the speech situation by a classifier.

In another of our four examples of non-sortal nouns with a classifier in (23), the [+unique, –relational] noun *đất* 'earth, ground' is marked by the classifier *mặt* 'face/surface' in a situation of contrastive focus. As in the case of the sky in (19), this noun is also compatible with other classifiers, among them *mảnh/miếng* 'piece' and *vùng* 'area'. The selection of a specific classifier depends again on the properties of the concept expressed by the noun as they are relevant to the speech situation.

3.4.2 Contrastive topics

There is an impressive body of literature on contrastive topics. For the purpose of this paper, Lambrecht's (1994: 183, 291, 195) discourse-based definition in terms of two activated topic referents which are contrasted will be sufficient. This type of topic is quite frequent in our Vietnamese corpus. A look at the statistics shows that classifier use is very strongly associated with contrastiveness. In fact, there is a classifier in each of the 84 instances of contrastive focus (66 in the written corpus and 18 in the oral corpus). Moreover, all nouns occurring in this function are [+animate].

In most examples, the action/state of one protagonist is contrasted with the action/state of another protagonist. As shown in (20), the actions of the son in the kitchen are contrasted with the actions of his mother in the bedroom (described as 'the wife' from the perspective of the husband). The son takes the classifier *đứa* for young boys, while the mother takes the classifier *bà* for women. The contrast between these two protagonists is supported by the adverbial subordinator *còn* 'while/whereas':

(20) (Written text 26, sentence 23)
Đứa con trai thì đứng lên kệ-bếp và vẽ bậy lên tường,
CL son TOP stand up kitchen-bar CONJ draw disorderly on wall
còn **bà vợ** thì nằm ăn đồ ăn nhanh với vẻ mặt khoái chí.
while CL wife TOP lie eat fast.food PRE expression delightful
'[His] son stood on the kitchen base (cabinet) and scribbled [something] onto the wall, while [his] wife was lying in bed, eating fast food with a facial expression of delight.'

In (21), the husband is contrasted with his wife. The husband's anger and his intention to make his wife eat some food is mirrored against his wife's reaction of refusing to give in. Both nouns take a classifier. The husband occurs with the classifier *ông* for men and the wife again takes the classifier *bà* for women. The contrast is explicitly expressed by the disjunctive conjunction *nhưng* 'but':

(21) (Written text 26, sentence 36)

Thấy thái độ của vợ-mình, **ông chồng** điên-máu-lên và
see attitude POSS wife-self CL husband get.crazy and

bắt ép ăn, **nhưng bà vợ** vẫn không ăn.
force eat CONJ CL wife still NEG eat

'Seeing the behaviour of his wife, the husband went crazy and [tried to] force her to eat, but [his] wife still did not eat.'

In the final example of this subsection, there is a contrast between a protagonist and a non-protagonist. The noun *bé* 'boy', as one of the two protagonists in the Pear Story, is contrasted with the children (*trẻ* 'child'). What is contrasted is the boy's action of leaving on a bike and the children's action of walking away. Again, both nouns occur with a classifier (*thằng* for the boy and *bọn* for the children) and there is a contrastive conjunction (*còn* 'while, whereas'):

(22) (Oral text 6, sentence 31)

Thằng bé tập tễnh dắt xe đi vài bước, còn **bọn trẻ** thì
CL boy limping lead bike go few step CONJ CL kid TOP

đi theo hướng ngược lại.
go toward direction opposite

'The boy led the bike limpingly, while the children walked in the opposite direction of the boy.'

3.4.3 Focus

Classifiers are also selected in various types of focus. This will be shown by the discussion of the two remaining non-sortal nouns with a classifier (cf. §3.2) plus two additional examples. The first example is on the [+unique, –relational] noun *đất* 'earth/ground'. In (23), this noun is interpreted as definite by the classifier *mặt*[18] for flat surfaces because it has the function of contrastive focus. The author

[18] *Mặt* has the meaning of 'face'. In this context, it is a classifier for objects with a flat surface. As a full noun, it can be interpreted as a [+relational] noun as in *mặt bàn* [surface table] 'the surface of the table'.

of this text starts her story from the perspective of the protagonist, a farmer, who is up 'on a tree' (*trên một cái cây*). Having described a series of the farmer's actions up there, her attention suddenly moves to the position of the baskets 'down on the ground' (*dưới mặt đất*), which is contrasted to the position up on the tree.[19]

(23) (Oral text 13, sentence 8)

Ổng leo lên một cái thang để ổng leo lên một cái cây để
3.SG climb PREP one CL ladder so that 3.SG climb PREP one CL tree to

ổng hái. Ổng hái xong, thì ổng leo xuống cái thang đó,
3.SG pluck 3.SG pluck RES CONJ 3.SG climb down CL ladder DEM

xuống đó. Rồi ổng, **dưới mặt đất** sẽ có ba cái giỏ ...
down DEM CONJ 3.SG down CL ground FUT have three CL basket

'He climbed a ladder to get on [a] tree to pick [the fruits]. Having picked [them], he went down [the] ladder. Then, he, down on the ground, there were three baskets...'

In contrast to (23), *đất* 'earth, ground' does not have a classifier in the non-contrastive situation of the following example:

(24) (Oral text 28, sentence 20)

Thì có ba sọt trái cây dưới **đất**,
CONJ have three CL fruit under ground

không có ai trông nom hết.
NEG have who take-care at.all

'There were three baskets of fruit on the ground, but nobody was taking care of them.'

Another context that induces classifier use is the context of focus particles, which typically mark the inclusion or exclusion of alternatives (König 1991). The other two examples to be discussed here both belong to this type of focus. The first example (25) is on the [+unique/+relational] noun *mặt* 'face', which occurs with the two focus particles *chỉ còn* 'only' and *mỗi* 'only'. The noun *mặt* 'face' takes the position between these two particles to emphasize the fact that the foam

[19]One of our reviewers suggests that *dưới mặt đất* 'down on the ground' is a frame-setter (e. g., Krifka 2008). This interpretation cannot be fully excluded. However, we would like to point out that the contrast between the position 'up in the tree' and the position 'down on the ground' is clearly given in the way the scenes are presented in the film.

covers almost the whole of the husband's body, leaving only his face unaffected. Thus, the two particles exhaustively single out one part of the body, which is excluded from the disturbing presence of foam:

(25) (Written text 29, sentence 31)

Lúc bấy giờ, người chồng nghe thấy bèn trồi lên khỏi mặt
time that CL husband hear RES CONJ rise out out of surface
nước, toàn thân ông là bọt xà phòng chỉ còn thấy mỗi **khuôn mặt**.
water whole body 3.SG COP foam only see only CL face
'At that time, the husband heard (the bell), then he moved out of the water. His whole body was full of soap foam, except **[the] face** [lit.: one can just only see **[his] face**].'

Our next two examples are not included in the statistics in Table 3 because they contain a possessive construction, and thus go beyond the distinction of bare noun vs. [CL+N]. In spite of this, they are relevant because classifiers very rarely occur with non-sortal head nouns of possessor constructions. In (26), the [+unique, +relational] possessee head noun *chồng* 'husband' in *chồng của mình* [husband CL self] 'husband of her' takes the classifier *ông*. Since non-sortal nouns of this type do not have a classifier in our data (e. g., *chồng (của) mình* [husband (possessive marker) self-$_{\text{reflexive pronoun}}$] '[her] husband', *vợ (của) mình* [wife (possessive marker) self-$_{\text{reflexive pronoun}}$] '[his] wife', *con trai họ* [son (possessive marker) selves-$_{\text{reflexive pronoun}}$] '[their] son', etc.), it is reasonable to assume that the presence of the classifier is due to the focus particle *ngoài* 'except':

(26) (Written text 30, sentence 8)

Khi giật mình vì bị tạt nước, bà vợ liền
when startle because PASS throw water, CL wife immediately
thức giấc nhìn xung quanh xem ai làm và chả có ai **ngoài ông**
awake look around see who do CONJ NEG have who except CL
chồng của mình.
husband POSS self
'Being startled by the water, the wife awoke immediately, looked around to see who did it. But there was nobody, except [her] husband.'

Finally, the classifier even occurs with non-sortal head nouns of possessive constructions, if the relevant focus situation can only be derived from context without the explicit presence of a focus marker. This is illustrated by (27), in which we find the two non-sortal nouns *chân* 'foot' and *mông* 'buttocks', the

former without a classifier, the latter with a classifier. The interpretation of this sentence crucially depends on the function of the adverbial subordinator *nên* 'therefore, to the extent that' which creates a context in which the situation becomes worse and worse until it culminates in a rather unexpected situation, in which the husband even burns his buttocks. This situation can be compared to the situation created by a focus particle like *even*:

(27) (Written text 13, sentence 35)

Ông ấy bị bỏng và đau quá nên ôm **chân** lên và không giữ
3.SG PASS burn CONJ hurt very CONJ hold foot RES and NEG keep

được thăng bằng nên té vào chiếc chảo đang cháy đó, **cái mông**
RES balance CONJ fall into CL pan PROG burn DEM CL buttock

của ông ấy đã bị phỏng.
POSS 3.SG PERF PASS burn.

'He got burnt and he got hurt, therefore, he lifted [his] leg up to hold it, then he was no longer able to keep his balance to the extent that he fell down on [the] burning pan and [as a consequence] even [his] buttocks got burnt.'

4 Classifiers and indefiniteness

Classifiers with indefinite interpretation are limited to particular contexts: the indefinite function of classifiers in the subject position of thetic statements is presented in §4.1. §4.2 discusses the [CL+N] construction in existential clauses, while §4.3 describes [CL+N] constructions in combination with verbs of appearance.

4.1 Thetic statements

As can be seen from Table 4, indefinite subjects are rather rare: 97.8% of the preverbal [CL+N] constructions of the written and the spoken corpus together are definite (cf. §3.3). The vast majority of the remaining 2.2% of indefinite preverbal [CL+N] constructions are subjects of thetic constructions (Kuroda 1972; Sasse 1987; 1995). Thetic utterances are seen in contrast to categorical utterances. Sasse (1995) defines both types as follows:

> Categorical utterances are said to be bipartite predications, involving a **predication base**, the entity about which the predication is made, and a **predicate**, which says something about the predication base. In other words, one of the arguments of the predicate is picked out as a "topic" in the literal sense, namely, an object about which something is asserted. Thetic utterances, on the other hand, are **monomial** predications (called "simple assertions" in Sasse 1987); no argument is picked out as a predication base; the entire situation, including all of its participants, is asserted as a unitary whole. (Sasse 1995: 4-5)

In utterances of this type, the entire clause is an 'all-new' utterance that is seen as inactivated information (often backgrounded) that is assumed by the speaker not to be present in the hearer's mind. Thus, nominal participants of thetic utterances are generally indefinite. The following two examples constitute the beginning of the story as told by two different informants. They provide a description of the initial scene as it was presented in the film. In the first sentence of both examples, the subject *đồng hồ báo thức* 'alarm clock' is marked by the classifier *chiếc*. Similarly, the subject *đàn ông* 'man' has the default classifier for humans, *người*, in the second sentence of both examples:

(28) Indefinite Subject (Written text 12, sentence 1, 2)

Chiếc đồng hồ báo thức reo lên lúc 8 giờ đúng.
CL clock alarm ring up at eight o'clock exactly

Người đàn ông đang ngủ thì bị nước văng tung tóe vào mặt.
CL man PROG sleep CONJ PASS water splatter PREP face

'The alarm clock rang at exactly eight o'clock. There was a man, who was sleeping and then [his] face was splattered with water.'

(29) Indefinite Subject (Written text 1, sentence 1, 2)

Chiếc đồng hồ báo thức reo lên báo hiệu đã tám giờ sáng.
CL clock alarm ring RES signaling PERF eight o'clock morning

Người đàn ông mở mắt liếc nhìn sang vợ-mình.
CL man open eye glance toward wife-self

'[The] alarm clock rang to signal that it was already 8 o'clock in the morning. [A] man opened his eye and glanced at [his] wife.'

4.2 Existential expressions

Existential sentences of the type 'there is an X' are typically used to introduce previously unidentified referents. Thus, [CL+N] constructions occurring in this type of construction are typically indefinite. Since they are positioned after the verb, they form a considerable part of the indefinite object classifiers in our data (but cf. inanimate nouns below). A good example is (30) from our oral corpus, in which the [CL+N] construction is preceded by the verb *có* 'have, there is':

(30) (Written text 20, sentence 28)

Lúc này, có **viên cảnh sát** vào hỏi xem tình hình vì hai
time DEM have CL policeman enter ask see situation because two

vợ chồng cãi nhau.
wife husband argue RECIP

'This time, [a] policeman entered and asked why this couple was arguing with each other.'

Another verb that implies indefiniteness is the copula verb *là* 'be', which is used in identificational contexts ('this is an X') as well as in locative contexts ('Y is [placed] in/at/on an X'). The following example starts out with a locative expression in the topic position (*bên cạnh đó* 'at the side of it, beside'). The three subsequent objects following the copula *là* are introduced as previously unmentioned elements into the scene by being situated within that locative topic:

(31) (Written text 14, sentence 2)

Bên cạnh đó là **cái kệ nhỏ**, **cái bình** và **ly nước** được đặt
beside DEM COP CL shelf small CL bottle and CL water PASS place

lên trên nó.
move.up top 3.SG

'Beside [him] was a small shelf with a bottle and a glass of water placed on it.'

Previous Context: The man who wore glasses awoke, opened his eyes for a moment, had a look around himself, ignored the alarm clock and went on sleeping.

In contrast to the thetic utterances of the preceding subsection, existential constructions can also be combined with constructions other than [CL+N]. For that reason, their impact on postverbal indefinite classifiers in our data is less strict than the impact of thetic utterances on indefinite classifiers in the subject

position. As is shown by the following example, existential expressions can also occur with the [*một* 'one'+CL+N] construction:

(32) (Oral text 20, sentence 1, 2)

Câu chuyện được bắt đầu vào một buổi sang tại một cánh đồng,
CL story PASS begin in one morning at one CL field

có **một người nông dân** leo lên trên một cái thang, đang hái
have one CL farmer climb up on one CL ladder PROG pluck

một loại trái cây nào đó giống trái lê.
one kind fruit some certain like CL pear

'The story began in a morning in a field. There was a farmer, who was climbing up a ladder to pick a kind of fruit like a pear.'

Finally, there are also some instances of inanimate nouns which occur without a classifier in existential constructions. This is illustrated by the following example with the noun *xe cứu hoả* 'fire truck' in its bare form:

(33) (Written text 20, sentence 25)

Gần đó, có **xe cứu hỏa** và lập tức đến xịt nước vào
nearby have car save fire CONJ immediately arrive spray water into

chữa cháy nhưng làm cho mọi thứ hỏng hết.
extinguish fire CONJ cause everything ruin

'Nearby, there was [a] fire truck, it arrived immediately to extinguish the fire. However, it also ruined everything.'

The extent to which the use of the classifier ultimately depends on the animacy of the noun cannot be determined from our data because we do not have enough examples.[20]

[20] In an alternative analysis, readers may be tempted to argue that the absence of the classifier is related to the complexity of the head noun (compounds vs. simple nouns) or to its status as a lexical item borrowed from Chinese. Since Emeneau (1951), it has often been claimed that nouns of this type take no classifiers. In spite of this, the noun *cảnh sát* 'policeman', which is borrowed from Chinese 警察 *jǐngchá* 'police(man)', does occur with the classifier *viên* in (30). Thus, we can at least exclude borrowing from Chinese as a strong factor for determining classifier use in existential constructions. In (31) it seems that animacy is more important. Ultimately, more data would be needed to enable more precise conclusions to be reached.

4.3 Verbs and situations of appearance

Vietnamese has quite a few verbs with the meaning of 'appear, come up', 'turn out to be' or 'reveal', whose subsequent nouns introduce previously unidentified elements into the discourse. In such cases, the postverbal noun is indefinite. In the following example with the verb *lòi ra* 'come to light, appear', the noun *tẩu thuốc* 'smoking pipe' takes the general classifier *cái*. Since the pipe was hidden in the husband's pocket, it is unknown to the audience/reader of the text and is interpreted as indefinite.

(34) (Written text 15, sentence 61)

Nhưng sau đó, cái túi áo của ông chồng bị lủng, lòi ra **cái**
CONJ after that CL pocket POSS CL husband PASS burst, show out CL
tẩu thuốc, chứ không phải vật gì có thể gây nguy hiểm.
smoking-pipe $\text{CONJ}_{\text{emph}}$ NEG something can cause danger
'However, after that, his pocket burst and what came to light was a smoking pipe, definitely nothing that may cause any danger.'

Sometimes, the meanings of verbs implying the emergence of unidentifiable concepts are highly specific. This can be shown by the verb *vấp* 'trip, walk into, stumble over', which creates a situation in which the object is unpredictable and has the status of being unidentifiable as in the following example:

(35) (Oral text 25, sentence 13)

Mải mê nhìn gái nên nó vấp phải
passionately look girl CONJ 3.SG trip PASS
cục đá và té xuống đường.
CL stone CONJ fall down road
'[He] looked at the girl passionately and thus stumbled over [a] stone and fell down on the road.'

Thus, the object *đá* 'stone' is marked by the classifier *cục* in (35). The boy, who is one of the two protagonists in the story, as well as the audience, cannot know what will happen when the boy is looking at the girl rather than at the road while riding his bike. The stone is clearly not activated and is interpreted as indefinite.

5 Conclusion

The aim of this study was to reach a better understanding of the referential functions of Vietnamese classifiers based on the systematic analysis of data from a corpus of written and oral texts which was designed to generate a broad variety of contexts which may trigger classifier use. The main results on the use and the functions of the Vietnamese classifier in [CL+N] can be summarized as follows:

(i) Classifiers can be interpreted as definite as well as indefinite but there is a clear preference for using the classifier in definite contexts (cf. §3.1).

(ii) There is a clear clustering of animacy and definiteness: definite animate nouns occur much more frequently with a classifier than definite inanimate nouns (§3.1).

(iii) There is a clear clustering of [CL+N] with [+definite, +animate, subject] and of bare nouns [N] with [−definite, −animate, object] (§3.3).

(iv) The overwhelming majority of nouns occurring in the [CL+N] construction are sortal nouns [−unique, −relational] (§3.2).

(v) Discourse and information structure play an important role in the function as well as in the presence/absence of a classifier:

   a. The definiteness with which classifiers are associated in [CL+N] is based on identifiability in discourse (§3.4.1);

   b. Information structure is an important factor for determining the use of a classifier in [CL+N] (§3.4.2 and §3.4.3) and its interpretation in terms of definiteness vs. indefiniteness (particularly cf. §4.1 on indefiniteness and theticity).

(vi) There are certain semantic environments which support the indefinite interpretation of the classifier (existential clauses and verbs of appearance; §4.2 and §4.3).

The results in (i) to (iii) on animacy, definiteness and subject/preverbal position tie in with general findings on prominence at the level of the morphosyntax-semantics interface as they manifest themselves in hierarchies like the animacy hierarchy (Silverstein 1976; Dixon 1979) or the accessibility hierarchy (Keenan &

Comrie 1977) (for a survey, cf. Bornkessel-Schlesewsky & Schlesewsky 2009).[21] The clustering observed in (ii) and (iii) additionally reflects a universal tendency to associate animate subjects in clause-initial positions of SVO languages with definiteness (Keenan & Comrie 1977; Givón 1979; Du Bois 1987; and many others). This tendency is also well known for word order in Sinitic languages (Li & Thompson 1976; Sun & Givón 1985; LaPolla 1995). Chen (2004: 1166) talks about definiteness-inclined preverbal positions and indefiniteness-inclined postverbal positions in Mandarin Chinese. As can be seen from (i), word order does not determine the (in)definiteness interpretation of the classifier in Vietnamese as rigidly as it does in Cantonese or in the Wu dialect of Fuyang (cf. the discussion of (11) and (12); for the discourse-based reasons for this, cf. below).[22]

[21]Based on the relevance of (in)definiteness and animacy, one may think of analyzing the use of the classifier in [CL+N] in the light of Differential Object Marking (DOM) as suggested by one of our reviewers. In our view, such an account would be problematic for at least the following reasons: (i) The use of the classifier in the [CL+N] construction is strongly associated with sortal nouns ([−relational]/[−unique]), while DOM marking is not limited to this type of nouns. (ii) As pointed out by Aissen (2003: 439), "it is those direct objects which most resemble typical subjects that get overtly case-marked". If one takes the use of the classifier as a DOM marker, one would expect the highest frequency of classifier use with [+definite] and [+animate] objects. This is clearly not borne out in the case of definiteness. As can be seen from Table 4, the ratio of definite subjects with CL is much higher than the ratio of definite objects with CL. There are 1,329 [= 1012 + 317] definite subjects with CL vs. 98 [= 86 + 12] definite subjects with no CL, i. e., 93.3% of the definite subjects in our two corpora have a classifier. In contrast, only 34.1% of the definite objects have a classifier (795 [= 432 + 363] definite objects with CL contrast with 1,533 [= 1206 + 327] without CL). In the case of animacy, the difference between the two ratios is smaller but it is still higher with animate subjects. As can be seen from Table 6, there are 1,259 [= 997 + 262] animate subjects with CL and only 10 [= 9 + 1] animate subjects with no CL, i. e., 99.2% of the animate subjects have a classifier. In the case of animate objects, the ratio is 72.7% (312 [= 179 + 133] animate objects with CL contrast with 117 [= 114 + 3] animate objects with no CL). (iii) The results discussed in (ii) are remarkable from the perspective of split vs. fluid DOM languages in terms of De Hoop & Malchukov (2007). In split languages, DOM marking is obligatory for a particular feature, while it is optional in fluid systems. In most DOM languages, DOM is split for at least one category. As can be seen in (ii), this is not the case with the use of the classifier. Vietnamese classifiers are not obligatory with definite objects nor are they obligatory with animate objects.

[22]In the case of Sinitic, Li & Bisang (2012) argue that the definiteness interpretation of subjects is due to a process of grammaticalization in which the definiteness properties of the topic position were passed on to the subject position (cf. the classical grammaticalization pathway from information structure to syntax in Givón 1979). In a similar way, the observation that postverbal [CL+N] constructions are preferably indefinite but do not exclude definiteness in Sinitic can be derived from the association of informational focus with the postverbal position (Xu 2004). As Lambrecht (1994: 262) points out, focus differs from topic inasmuch as it is not necessarily identifiable or pragmatically salient in discourse. For that reason, it is open to indefinite and definite interpretation even though the default interpretation is indefinite. If this

The observation in (iv) that the vast majority of nouns occurring in the [CL+N] construction are sortal nouns in the terms of Löbner (1985) confirms and further specifies the findings of Simpson (2017: 324) on the Wu variety of Jinyun, that nouns denoting "specifically unique individuals/elements" predominantly appear as bare nouns [N] (cf. the three instances of [+unique] nouns taking a classifier in Table 3). These results show the potential relevance of Löbner's (1985; 2011) four basic types of nouns for understanding definiteness/indefiniteness as associated with the [CL+N] construction in East and mainland Southeast Asian languages.

Even though the factors of semantics (animacy, uniqueness, relationality) and syntax (subject, object) clearly have an impact on the presence or absence of the classifier in contexts of definiteness and indefiniteness, we have evidence that discourse and information structure are stronger than these factors. The dominance of discourse is reflected in the very function of the classifier itself. As discussed in §3.4.1, classifiers mark identifiability rather than uniqueness (cf. point (v.a), also cf. Li & Bisang 2012 on Sinitic). Thus, they express pragmatic definiteness rather than semantic definiteness in terms of Löbner (1985; 2011) or anaphoric ("strong") definiteness rather than unique ("weak") definiteness in terms of Schwarz (2009; 2013). In addition to the discourse-based definiteness expressed by the classifier, contrastive topics (§3.4.2), as well as contrastive focus and focus particles (§3.4.3), enhance the use of the [CL+N] construction. Thetic statements, as another instantiation of information structure, play an important role in the indefinite interpretation of [CL+N] in the subject position (§4.1; also cf. (v.b)). Moreover, there are more specific discourse-based environments as mentioned in point (vi) which support the use of a classifier in contexts of indefinite interpretation (§4.2 and §4.3). Finally, evidence of the dominance of discourse comes from data outside of our corpus. In order to disentangle the semantic effects of animacy vs. discourse effects associated with protagonists, we looked for narrative texts with inanimate protagonists. In the three texts we found, the inanimate protagonists generally occur in the [CL+N] construction (Quang forthcoming). One of the stories is about a flying carpet, which is already mentioned in the title, *Tấm thảm bay* [CL carpet fly] 'The Flying Carpet'.[23] After the protagonist is introduced by an indefinite construction of the type [one CL N], the noun *thảm* 'carpet' consistently occurs with a classifier. It is important to add in this

analysis is true, one may argue that in Sinitic the classifier in [CL+N] is like a variable that takes on the [±definite] function that corresponds to its syntactic position if it is not overwritten by stronger factors. In Vietnamese, such a syntactic scenario turns out to be problematic because the classifier generally favours definite interpretation (cf. point (i)).

[23]The story was published by Viet Nam Education Publisher in 2003.

context that the carpet has no anthropomorphic properties in the story, i. e., it does not act in any way. It is just the element that keeps the story going through many different events and episodes. Needless to say, such examples are hard to find in a corpus, no matter how large it is, because they are rare overall. The fact that even inanimate protagonists generally can take a classifier together with the findings summarized in (v) are good evidence for the dominance of discourse and information structure over semantics and syntax.

Taking these findings together, the classifier in [CL+N] is used as a variable whose use and interpretation depend on prominence in discourse and interact with factors from the morphosyntax-semantics interface. The details of that interaction will undoubtedly need more research. What is remarkable and makes the data on Vietnamese and other East and mainland Southeast Asian languages particularly relevant from a typological perspective is the observation that the different factors associated with (in)definiteness are well known, while cross-linguistic variation in how they interact is still under-researched. In Vietnamese, factors of discourse are particularly prominent. In order to further corroborate these observations and compare them with the situation in other mainland Southeast Asian languages, it is necessary to look at how classifiers are used in actual discourse in text corpora. We understand the corpus discussed here as a starting point for Vietnamese.

Acknowledgements

We would like to thank the editors for their time and support in the editorial processes. Comments from two anonymous reviewers have greatly improved the content. We owe special thanks to our 46 informants and friends in Ho Chi Minh city, Vietnam, for their help in participating in our experiments.

References

Aissen, Judith. 2003. Differential object marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21. 435–483.

Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: Counting and beyond. In Jadranka Gvozdanovic (ed.), *Numeral types and changes worldwide* (Trends in Linguistics, Studies and Monographs (TiLSM), Vol. 118), 113–185. Berlin: Mouton De Gruyter.

Bisang, Walter & Yicheng Wu. 2017. Numeral classifiers in East Asia. *Linguistics* 55. 257–264.

Bornkessel-Schlesewsky, Ina & Matthias Schlesewsky. 2009. The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass* 3(1). 19–58.

Chafe, Wallace L. 1980. *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chen, Ping. 2004. Identifiability and definiteness in Chinese. *Linguistics* 42. 1129–1184.

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

Croft, William. 2003. *Typology and universals*. 2nd edition. Cambridge: Cambridge University Press.

De Hoop, Helen A. & Andrej Malchukov. 2007. On fluid differential case marking: A bidirectional OT approach. *Lingua* 117. 1636–1656.

Dixon, Robert M. W. 1979. Ergativity. *Language* 55. 59–138.

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63. 805–855.

Emeneau, Murray B. 1951. *Studies in Vietnamese (Annamese) Grammar* (University of California Publications in Linguistics 8). Berkeley/Los Angeles: University of California Press.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Greenberg, Joseph. 1972. Numerical classifiers and substantival number: Problems in the genesis of a linguistic type. *Working Papers on Language Universals* 9. Department of Linguistics, Stanford University, 1–39.

Jiang, L. Julie. 2015. Marking (in)definiteness in classifier languages. *Bulletin of Chinese Linguistics* 8. 319–343.

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.

König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London/New York: Routledge.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4). 243–276.

Kuroda, Shichiro. 1972. The categorical and the thetic judgment: Evidence from Japanese syntax. *Foundations of Language* 9(2). 153–185.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press.

LaPolla, Randy J. 1995. Pragmatic relations and word order in Chinese. In Pamela A. Downing & Michael Noonan (eds.), *Word order in discourse*, 297–329. Amsterdam/Philadelphia: John Benjamins.

Li, Charles N. & Sandra A. Thompson. 1976. Subject and topic: A new typology of language. In Charles N. Li (ed.), *Subject and topic*, 457–489. New York: Academic Press.

Li, XuPing & Walter Bisang. 2012. Classifiers in Sinitic languages: From individuation to definiteness-marking. *Lingua* 122. 335–355.

Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4. 279–326.

Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.

Nguyen, Kim Than, Hai Thuy Ho & Duc Duong Nguyen. 2005. *Từ điển tiếng Việt [Vietnamese Dictionary]*. Ho Chi Minh City: Saigon Cultural Publisher.

Nguyen, Tuong H. 2004. *The structure of the Vietnamese noun phrase*. Boston University. (Doctoral dissertation).

Quang, Kim Ngoc. Forthcoming. *Vietnamese classifiers and (in)definiteness: A text-based analysis*. University of Mainz. (Doctoral dissertation).

Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25. 511–580.

Sasse, Hans-Jürgen. 1995. 'Theticity' and VS order: A case study. *Sprachtypologie und Universalienforschung* 48. 3–31.

Schwarz, Florian. 2009. *Two types of definites in natural language*. Amherst, MA: University of Massachusetts. (Doctoral dissertation).

Schwarz, Florian. 2013. Two types of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–558.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. New Jersey: Humanities Press.

Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asian languages. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 806–838. Oxford: Oxford University Press.

Simpson, Andrew. 2017. Bare classifier/noun alternations in the Jinyun (Wu) variety of Chinese and the encoding of definiteness. *Linguistics* 55(2). 305–331.

Sun, Chao-Fen & Talmy Givón. 1985. On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language* 61. 329–351.

Tran, Jennie. 2011. *The acquisition of Vietnamese classifiers*. Honolulu: University of Hawaii. (Doctoral dissertation).

Trinh, Tue. 2011. Nominal reference in two classifier languages. In Ingo Reich, Eva Horch & Dennis Pauly (eds.), *Sinn und Bedeutung 15: Proceedings of the*

2010 annual conference of the Gesellschaft für Semantik, 629–644. Saarbrücken: Saarland University Press.

Wang, Jian. 2015. Bare classifier phrases in Sinitic languages: A typological perspective. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 110–133. Oxford: Oxford University Press.

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

Xu, Liejiong. 2004. Manifestations of informational focus. *Lingua* 114. 277–299.