

A Corpus-based Study of Inanimate Classifiers in Vietnamese

By

Hai Thi Thanh Tran

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Department of Linguistics

University of Manitoba

Winnipeg

Copyright © 2021 by Hai Thi Thanh Tran

Abstract

Key words: Vietnamese, classifiers, variation, discourse analysis, corpus-based study.

This dissertation investigates inanimate classifiers in Vietnamese to identify classifier use patterns across genres, in spoken and written discourse as well as among different age groups. The study works on three corpora namely the Vietnamese Narrative Corpus consisting of 141 folktales, the Vietnamese Online Newspaper Corpus containing 140 contemporary online newspaper articles, and the Vietnamese Spoken Corpus including 22 talk show episodes with the total duration of 14 hours. As a large-scale discourse analysis study of frequency, distribution and function of attested inanimate classifiers, it attends closely to the use of *cái* (inanimate), double classifiers, and other frequent classifiers in the Vietnamese corpora.

The study found that the classifier frequency in spoken Vietnamese is far higher than in written language. In Vietnamese, a classifier is required for classified nouns, but not for non-classified nouns (Emeneau 1951; Nguyen 1957). However, *cái* (inanimate) frequently appears with non-classified nouns functioning as emphatics in the spoken corpus, but not in the written corpora. I argue that this may lead to the higher classifier frequency in spoken Vietnamese than in written language. Interestingly, there is a decline in classifier frequency among younger speakers compared to older speakers. The data reveals that language change may be in progress in Vietnamese in apparent time, which merits further investigation. Double classifiers, in which *cái* (inanimate) is constructed with a classifier combining with classified nouns, are used often in spoken Vietnamese, but not in written language. I claim that in this construction *cái* (inanimate) functions as an emphatic, while the other classifier performs its own function.

This research found new evidence that *cái* (inanimate) functions as cataphoric reference in Vietnamese. A number of classifiers which can combine with verbs and adjectival verbs function as nominalizers to indicate different semantic types in Vietnamese. The frequent classifiers in the concurrent corpora differ from those in the narrative corpus, suggesting language change over time. The study also found that the choice of classifiers may be dependent on the formality of the genre. The findings reveal that the Vietnamese classifier system is highly complex, and what was observed in naturalistic data of Vietnamese corpora, especially spoken discourse, is different from the traditional description of classifier usage in Vietnamese.

Acknowledgements

It has been a long journey to complete this dissertation, and I have learned a lot during the journey. This dissertation would have not been completed without the support of a number of individuals to whom I would like to express my gratitude. First, I am deeply grateful to my supervisor, Dr. Terry Janzen, for his unwavering support and constructive mentorship throughout this study. I am thankful for his great help and being with me to the successful completion for my PhD. Dr. Janzen has always offered motivation and encouragement since the day he taught me some courses and supervised me on my second Generals Paper. Without his guidance and insightful comments, this study would have never been completed. I also thank him for his faith in me and for his kindness. He is a great linguist and an incredible person. He is always passionate about new things, which motivates me to do this research. It was such a great pleasure to conduct this study under his supervision.

I would like to sincerely thank the committee members for their specialized and erudite inputs. Special thanks are due to Dr. Nicole Rosen for her meticulous and informative comments that inevitably helped me do this research better. Since the day I came to ask for her help to code the variables and put them in an Excel spreadsheet, she showed me how to keep the data in an organized way, which helps me easily keep track on the data and use them for analysis. She has been very helpful and has become a source of inspiration for me. I am grateful to her for her willingness to advise me on my first Generals Paper, which is the first step for the continuance in my program. She is a remarkable professor and an amazing woman. Many thanks also go to Dr. Erin Wilkinson for being my committee member and for her helpful suggestions and comments when I defended my thesis proposal. I am grateful to her for reading the thesis and making interesting, useful comments and suggestions for me to improve my dissertation. I am also thankful to Dr. Doris Payne for her kindly accepting to be my external examiner. Her rigorous examination together with the valued comments and questions have greatly enriched my study and experience.

I would like to express my thanks to all the professors in the Department of Linguistics at the University of Manitoba for enriching my view of linguistics. Many thanks go to Dr. Jila Ghomeshi, the Head of the Department, who taught me the course of Syntactic Theory. The assignments and the final paper of her course for which I worked on the linguistic issues of Vietnamese inspired me a lot. During my study at the department, I also received a lot of help from

others, especially Heather Cherpako, who has provided invaluable help with administrative matters, and I am grateful for her supports. I would like to say thanks to all my colleagues and friends who have supported me along the way. My sincere thanks go to Thu Trang Nguyen, a Vietnamese PhD student in the Department of Linguistics at the University of Manitoba, for her providing some transcription for the data and support. My special thanks go to my friend, Brenda Noble, an ESL teacher, for her help with proofreading the thesis for me.

My doctoral degree was funded by Vietnam International Education and Development (VIED), the Ministry of Education and Training of Vietnam, and I am greatly indebted to them for their supports. I would like to express my sincere thanks to the leaders of Hanoi Foreign Trade University for their facilitation and support in my PhD journey. My colleagues at the Faculty of English Linguistics and my friends at Hanoi Foreign Trade University are also a source of inspiration for me, and I thank them for their supports and encouragement. My special thanks are due to my MA supervisor, Dr. Duc Hoat Nguyen, who has always encouraged me in my study. Last but not least, I would love to express my heartfelt gratitude to my parents, my siblings, and my two beloved children for their love, care, understanding, encouragement and support during the journey of my PhD program completion.

Table of contents

Abstract	ii
Acknowledgements	iii
Table of contents	v
List of abbreviations	ix
List of tables	x
List of figures	xi
Chapter 1 Introduction	1
1.1 Background of the study	2
1.2 Types of classifiers in Vietnamese	5
1.2.1 Human classifiers	5
1.2.2 Animate non-human classifiers	6
1.2.3 Inanimate classifiers	7
1.3 Rationale and scope of the study	9
1.4 Objectives of the study	10
1.5 Organization of the thesis	12
Chapter 2 Literature review	14
2.1 Introduction	14
2.2 Classifier systems	15
2.2.1 Classification of classifier systems	15
2.2.2 Functions of classifiers	16
2.2.3 Numeral classifier construction	17
2.3 The Vietnamese noun phrase	18
2.4 The Vietnamese classifier system	19
2.4.1 Overview of the Vietnamese classifier system	19
2.4.2 Number of Vietnamese classifiers	22
2.4.3 Vietnamese classifier constructions	25
2.4.4 Functions of Vietnamese classifiers	28
2.4.5 Characteristics of Vietnamese classifiers	31
2.5 Categorization of Vietnamese classifiers	34

2.5.1 General classifiers	35
2.5.2 Specific classifiers	36
2.5.3 Type classifiers	37
2.5.4 Classifiers showing contents	37
2.5.5 Event classifiers	38
2.5.6 Double classifiers	38
2.6 Theoretical framework and key terms	41
2.6.1 Variationist framework	41
2.6.2 Terms used in the study	43
2.7 Summary	43
Chapter 3 Corpus and Methodology	45
3.1 Corpora	45
3.1.1 The Vietnamese Narrative Corpus	45
3.1.2 The Vietnamese Online Newspaper Corpus	47
3.1.3 The Vietnamese Spoken Corpus	48
3.2 Criteria for identifying a classifier in Vietnamese	49
3.2.1 Criteria for identifying inanimate classifiers	49
3.2.2 Difficulties in identifying inanimate classifiers in Vietnamese	53
3.3 Methodology	59
3.3.1 Organizing the data	59
3.3.2 Analyzing the data	60
3.3.2.1 Identifying inanimate classifiers and head nouns	60
3.3.2.2 Analyzing nouns and classifiers under some factors/variables	61
3.3.3 Data aggregation for analysis	63
3.4 Summary.....	64
Chapter 4 Frequency and distribution of inanimate classifiers	66
4.1 Major findings of classifier use in the study	66
4.1.1 Frequency of classifier use in the three corpora	66
4.1.2 Overall distribution of the most frequent CLs in the three corpora	67
4.2 Findings from the Vietnamese Narrative Corpus	68
4.2.1 Frequency of classifiers in the Narrative Corpus	68

4.2.2 Overall distribution of frequent classifiers in the Narrative Corpus	71
4.2.3 Overall distribution of infrequent classifiers in the Narrative Corpus	77
4.2.4 Summary	82
4.3 Findings from the Vietnamese Online Newspaper Corpus	83
4.3.1 Frequency of classifiers in the Online Newspaper Corpus	83
4.3.2 Overall distribution of frequent CLs in the Online Newspaper Corpus	84
4.3.3 Overall distribution of infrequent CLs in the Online Newspaper Corpus	90
4.4 Findings from the Vietnamese Spoken Corpus	93
4.4.1 Frequency of classifiers in the Spoken Corpus	94
4.4.2 Overall distribution of frequent classifiers in the Spoken Corpus	96
4.4.3 Overall distribution of infrequent classifiers in the Spoken Corpus	104
4.4.4 Exceptional cases in the Spoken Corpus	105
4.5 Summary	109
Chapter 5 Discussion	111
5.1 Overview of inanimate classifier use in the three corpora	111
5.1.1 Frequency of inanimate classifiers	111
5.1.1.1 Frequency of classifier use in the three corpora	111
5.1.1.2 Frequency of CL use among different age groups	115
5.1.2 Overall distribution of inanimate classifiers in the three corpora	120
5.1.3 Number of inanimate classifiers in the corpora	127
5.1.4 Summary	132
5.2 Classifier use in written discourse and spoken discourse	133
5.2.1 Differences in classifier use in written and spoken discourse	133
5.2.1.1 Frequency of classifier use	133
5.2.1.2 Overall distribution of classifiers	134
5.2.2 Double classifiers and their lexical semantic functions	136
5.2.3 Summary	145
5.3 The use of <i>cái</i> (inanimate) in the three corpora	146
5.3.1 The use of <i>cái</i> (inanimate) in the Narrative Corpus	146
5.3.2 The use of <i>cái</i> (inanimate) in the Online Newspaper Corpus	150
5.3.3 The use of <i>cái</i> (inanimate) in the Spoken Corpus	153

5.3.4 The use of <i>cái</i> (inanimate) with non-classified nouns	157
5.3.5 Summary	162
5.4 Analysis of classifiers with the nominalization function	164
5.5 Vietnamese classifier constructions and other issues	170
5.5.1 Vietnamese classifier constructions	170
5.5.2 Definiteness of the noun	172
5.5.3 Summary	173
Chapter 6 Summary and Conclusions	175
6.1 Summary of the goals and methodology of the study	175
6.1.1 Summary of the goals of the dissertation	175
6.1.2 Summary of the methodology of the study	175
6.2 Major findings and conclusions of the dissertation	176
6.2.1 Frequency of classifier use across genres	176
6.2.2 Distribution of classifiers in the three corpora	177
6.2.3 Functions of classifiers in Vietnamese	179
6.2.4 Other findings	180
6.2.5 Concluding summary	182
6.3 Implications and recommendations for future research	183
6.3.1 Implications for teaching Vietnamese classifiers	183
6.3.2 Recommendations for future research	185
6.3.3 Conclusion	185
References	187
Appendix A	193
Appendix B	200
Appendix C	206
Appendix D	211
Appendix E	218
Appendix F	220
Appendix G	225
Appendix H	230

List of abbreviations

A	absent	O	Online
Abs	abstract	P	present
AdjP	adjective phrase	poss.	possessive
ani.	animate	QUES	question
CL(s)	classifier(s)	S	Spoken
class.	classification	s. s.	social status
Con	concrete	TV	television
DEF	definite	VP	verb phrase
F/f	female	VTV	Vietnam Television
i. e.	it means	VNC	Vietnam Narrative Corpus
inani.	inanimate	VONC	Vietnam Online Newspaper Corpus
Ind	individualization	VSC	Vietnam Spoken Corpus
ling	linguistic	w.	with
M/m	male	w/o	without
Nar	Narrative	y	young
Nom	nominalization	YOB	Year of birth
N	noun		
NP(s)	noun phrase(s)		
No.	number		
Num/NUM	numeral		

List of tables

Table 1: Criteria for identifying a classifier in Vietnamese	53
Table 2a: Examples of data analysis	61
Table 2b: Examples of data analysis	62
Table 3: Overall distribution of frequent CLs in the three corpora	67
Table 4: Frequency of CLs in the Narrative Corpus	69
Table 5: Distribution of frequent CLs in the Narrative Corpus	71
Table 6: Frequency of CLs in the Online Newspaper Corpus	83
Table 7: Distribution of frequent CLs in the Online Newspaper Corpus	85
Table 8: Frequency of CLs in the Spoken Corpus	94
Table 9: Overall distribution of CLs in the Spoken Corpus	97
Table 10: Distribution of <i>cái</i> (inanimate), <i>sự</i> (event), and <i>cuộc</i> (life, strike)	123
Table 11: Distribution of double CLs in the Spoken Corpus	137
Table 12: Comparison of non-classified nouns with or without <i>cái</i> (inanimate)	161
Table 13: Constructions of <i>cái</i> (inanimate) in the three corpora	163
Table 14: Major classifier constructions in the study	170
Table 15: Classifier constructions without the head noun in the three corpora	171

List of figures

Figure 1: Frequency of classifier use in the three corpora	112
Figure 2: Number of cái (inani.) tokens and its frequency in the three corpora	114
Figure 3: Frequency of CL use among different age groups in VSC	116
Figure 4: The use of cái (inani.) among different age groups in VSC	117
Figure 5: Number of actual inanimate CLs in the three corpora	127
Figure 6: Distribution of frequent CLs in the written corpora	135
Figure 7: Distribution of frequent CLs in the spoken corpus	135

Chapter 1

Introduction

The complex system of language that we speak and write is evolving (Tagliamonte 2012), and variation is an “inherent part of language” (Labov 1969:728). It is thus important to investigate variation in the evolving system of language. Vietnamese has a highly complex classifier system with a large inventory of classifiers, which attracts attention from researchers within the country and worldwide (Emeneau 1951; D. H. Nguyen 1957; Thompson 1965; Cao 1988, 1998; Daley 1998; Löbel 2000; P. P. Nguyen 2002; Pham and Kohnert 2008; J. Tran 2011; H. T. Nguyen 2004, 2013; Simpson and Ngo 2018). However, there usually exists a “significant discrepancy between a rich language inventory of classifiers found in dictionaries and prescriptive grammars and the much more restricted set attested in actual speech” (Craig 1986:8). In fact, the existing analyses on Vietnamese classifiers are mainly on written texts (newspapers and/or books) and none of them looks at actual speech (spoken data). I argue that the frequency and distribution of classifiers in Vietnamese spoken discourse is different from their uses in written discourse.

Furthermore, investigating classifiers across different genres will identify variation and classifier use patterns in the Vietnamese classifier system. With this aim, the current research examines inanimate classifiers in three different corpora with the focus on *cái* (inanimate) and double classifiers. I argue that *cái* (inanimate) and double classifiers are used at higher frequency rates in Vietnamese spoken than in written language. From my observations, I hypothesize that there is variation in classifier frequency and distribution across different genres as well as among different age groups. I also assume that the distribution of inanimate classifiers in the concurrent Vietnamese corpora differs from those in the narrative corpus and the traditional studies.

This chapter, which lays the background for the above proposals and presents an overview of the dissertation, is organized as follows. Section 1.1 presents the background of the study and definition of classifiers in Vietnamese. Section 1.2 provides a brief description of the three semantic types of Vietnamese classifiers: human, animate non-human, and inanimate. Section 1.3 introduces the rationale and scope of the study. Section 1.4 describes the objectives of the research. Finally, the organization of the chapters of the thesis is presented in section 1.5.

1.1 Background of the study

As a Mon-Khmer language in the Austroasiatic family, Vietnamese is a classifier, monosyllabic, and non-inflectionary language. As the official language of Vietnam, Vietnamese is spoken by more than 96 million Vietnamese people within the country and overseas, according to the General Statistics Office of Vietnam (2019). Vietnamese has a highly complex classifier system and a rich inventory of classifiers, which has led to arguments among researchers of Vietnamese classifiers in terms of number and functions (Emeneau 1951; Aikhenvald 2000; P. P. Nguyen 2002). However, previous analyses primarily rely on constructed or elicited utterances, so it is worthy to investigate classifiers on a corpus basis to find out the actual uses of classifiers in Vietnamese. This section presents the background of the study and the definition of classifiers in order to clarify the picture of the Vietnamese classifier system and how important to examine classifiers in Vietnamese.

According to researchers of Vietnamese classifiers including Emeneau (1951) and Thompson (1965), a classifier in Vietnamese is a word that categorizes the noun by grouping the thing denoted by the noun it precedes into a generalized classification. D. H. Nguyen (1957:125) claims that a classifier is really “a unit quantity, or number, of that denoted by the noun that it precedes” because Vietnamese nouns do not have number as part of their class meaning, but only serve to identify the kind. P. P. Nguyen (2002) defines a classifier as a part of speech to identify the unit, a semantic unit. As nouns in Vietnamese are non-individuated, they can be “individuated via classifiers or measure phrases” in order to be counted or measured (H. T. Nguyen 2013:59). Thompson (1965) discusses that classifiers are considered as function words or ‘empty words’ that could not be independent due to their inability to stand alone as a noun phrase. However, Luu (2000) argues that Vietnamese classifiers could be considered content words as they contain varying degrees of lexical-semantic meaning. Pham and Kohnert (2008) claim that from a semantic perspective, classifiers are unbound function words that categorize the head noun based on inherent or salient features of the noun’s referent, such as animacy, shape, length, dimension, function, or material. Researchers have differing views on how to define classifiers as they look at classifiers from different perspectives. However, these researchers share the same approach to identifying classifiers in Vietnamese. This will be further reviewed in the literature chapter.

They all agree that Vietnamese nouns themselves do not carry any notion of number, so they are all somewhat like English mass nouns (Thompson 1965). Vietnamese nouns such as *bếp* (kitchen) and *mèo* (cat) are non-individuated and thus can only be counted or measured in the presence of a classifier or a measure phrase (H. T. Nguyen 2004, 2013) as in (1) and (2).

- (1) *cái* *bếp*
 CL(inanimate) kitchen
 ‘the kitchen’ (N1.39)¹
- (2) *một con* *mèo*
 one CL(animate) cat
 ‘a/one cat’ (N1.138)
- (3) *người* *mẹ*
 CL(human) mother
 ‘the mother’ (N2.120)

The three general classifiers in Vietnamese which are widely recognized by all linguists are *cái* (inanimate), *con* (animate non-human), and *người* (human) (Cao 1998; Emeneau 1951; D. H. Nguyen 1957; P. P. Nguyen 2002) as in the examples in (1)-(3). These three classifiers belong to three semantic classifier types of Vietnamese: inanimate, animate non-human, and human. Each of the classifiers represents each type respectively. These are the only classifiers recognized by Cao (1998). However, many other researchers believe that Vietnamese has a large number of classifiers, which may go up to 195 as P. P. Nguyen (2002) claims. In my corpus-based study, 150 actual classifiers including 34 human, 12 animate non-human, and 113 inanimate classifiers were identified (Tran 2018). The discrepancy in the number of classifiers claimed by different researchers is due to the differing views on classifiers and head nouns in Vietnamese, which will be explained in detail in the literature in section 2.4.2.

According to researchers including Emeneau (1951), D. H. Nguyen (1957), Thompson (1965), P. P. Nguyen (2002), and H. T. Nguyen (2004, 2013), there are many other specific classifiers in Vietnamese in addition to the three general classifiers mentioned above. These

¹ Unless otherwise indicated, the examples in this chapter are from the 2018 corpus of my previous research (Tran 2018), which is a subset of the Vietnamese Narrative Corpus in the current study. It will be fully described in chapter 3. (N1.39) in (1) means the example is from the Narrative Corpus, Book 1, and page number 39.

specific classifiers include *chiếc* (individual), *cây* (tree, long) and *quả* (fruit, round) as in (4)-(6).

(4) bốn chiếc bánh

four CL(individual) cake

‘four cakes’ (N1.43)

(5) một cây tre trăm đốt dài lắm

one CL(tree, long) bamboo hundred knots long very

‘a very long bamboo tree of hundred knots’ (N1.25)

(6) hai quả bầu khô

two CL(fruit, round) gourd dry

‘two dried gourds’ (N2.69)

Despite the fact that researchers of Vietnamese claim different numbers of classifiers, there are a great variety of inanimate classifiers in Vietnamese since at least 113 inanimate classifiers were found in the corpus in my prior research (Tran 2018). However, in this study, I do not aim to focus on exploring how many classifiers exist in Vietnamese. I attempt to investigate what inanimate classifiers are used in naturalistic data of Vietnamese corpora with the focus on how *cái* (inanimate) and frequent classifiers are used in the corpora. Furthermore, I will identify classifier use patterns across different genres in Vietnamese with an attempt to find out whether there is variation in the choice of classifiers in these genres.

Different types of numeral classifiers may co-occur, and they carry different properties (Aikhenvald 2000). The data in the corpus-based study shows evidence that two classifiers can co-occur, and they display different properties (Tran 2018) as in the example in (7).

(7) cái đám ma kia

CL(inani.) CL(procession) ghost that

‘that funeral procession’ (N1.142)

In this construction, *cái* (inanimate) occurs in the first position and the specific classifier *đám* (procession) in the second position. This noun phrase is definite due to the presence of the demonstrative *kia* (that). In this case, *đám* (procession) classifies and individuates the noun, while *cái* (inanimate) is supposed to be used for emphasizing the noun. In (7), *cái* (inanimate) can be omitted without changing the meaning of the noun phrase, but the specific classifier *đám*

(procession) cannot. In fact, *cái* (inanimate) in this case is of optional use and receives a lot of argument from researchers (Emeneau 1951; D. H. Nguyen 1957; H. T. Nguyen 2013; Simpson and Ngo 2018). In this study, I will investigate the use and function of classifiers in this doubling construction. It is hypothesized that double classifiers occur more often in spoken discourse than in written discourse. The findings in my prior research (Tran 2018) show that a limited number of double classifiers are used, and the function of each classifier differs from one another. I hypothesize that the omission of one classifier in this double classifier construction is possible. This issue will be examined in the current study.

1.2 Types of Vietnamese classifiers

According to previous researchers including Emeneau (1951), D. H. Nguyen (1957), Thompson (1965), and P. P. Nguyen (2002), Vietnamese classifiers belong to three semantic types: human, animate non-human, and inanimate. In this research, I focus on inanimate classifiers only. However, human and animate non-human classifier types are also briefly introduced in this section for clarifying the overall picture of the Vietnamese classifier system.

1.2.1 Human classifiers

Classifiers for nouns denoting human beings belong to two categories: general and special. General classifiers do not indicate any status, but special ones do (D. H. Nguyen 1957). That means the use of special classifiers is governed by the “status factor”, the social distance between the speaker and the person referred to (D. H. Nguyen 1957:132).

Người (human) is the general human classifier used before kinship terms as in (3) given in section 1.1 and other nouns indicating the occupation of the person as in (8).

- (8) một *người* thợ
 one CL(human) worker
 ‘a worker’ (Nguyen 1957:133).

In Tran (2018)’s study, thirty-four actual human classifiers were identified in which the most frequent human classifier is *người* (human) with 321 occurrences, accounting for 31% out of 1012 human classifier tokens in the data. Other common classifiers include *ông* (human, male, old), *cô* (human, female, young), *thằng* (human, male), and *đứa* (human) as in (9)-(10).

- (9) a. *ông* *ăn mày*
 CL(human, male, old) beg
 ‘an old male beggar’ (N1.155)
- b. *một cô* *gái đẹp*
 one CL(human, female, young) girl beautiful
 ‘a beautiful girl’ (N2.179)
- c. *một đứa* *con* *trai* *khôi ngô*
 one CL(human, young) offspring male smart
 ‘a smart son’ (N2.100)
- (10) a. *hai thằng* *kẻ* *trộm*
 two CL(human, male, low s. s.) CL(human, low s. s.) steal
 ‘two male thieves’ (N1.155).
- b. *hai thằng* *trộm*
 two CL(human, male) steal
 ‘two male thieves’ (N1.155) (Tran 2018:63).

The co-occurrence of two classifiers as in (10a) is an interesting phenomenon in Vietnamese. In this example, either one of them can be omitted without changing the meaning of the noun. The classifier *kẻ* (human, low social status) is omitted as in (10b). However, if the classifier *thằng* (human, male, low social status) is taken out, the noun does not specify the gender of the persons it refers to (Tran 2018).

1.2.2 Animate non-human classifiers

In Tran (2018)’s study, twelve animate non-human classifiers were found with 374 classifier tokens in the corpus. The general animate classifier *con* (animate non-human) as in (11) in Vietnamese is very common. It occurs 316 times, accounting for 84% of all the animate non-human classifier tokens in the corpus (Tran 2018).

- (11) *một con* *nai*
 one CL(ani.) deer
 ‘a/one deer’ (N2.112)

Tran (2018)’s study shows that eight human classifiers are used with nouns denoting animals anthropomorphically. One of the two kinship terms which goes with nouns denoting animals is *chú* (human, male) as in (12). In this case, it appears that personification is used with the appearance of the human classifier *chú* (human, male) (Tran 2018).

- (12) một *chú* gấu
 one CL(human, male) bear
 ‘a bear’ (N2.102)

The general classifier *con* (animate non-human) is widely recognized as an animate classifier by all researchers. However, this classifier can be used for nouns denoting non-living things as in (13) and for nouns denoting human beings as in (14) (D. H. Nguyen 1957).

- (13) a. *con* dao
 CL(ani.) knife
 ‘knife’ (N2.101)
- b. một *con* sông rất sâu
 one CL(ani.) river very deep
 ‘a very deep river’ (N2.114)
- (14) a. *con* bạc
 CL(ani.) gambler
 ‘gambler’ (Nguyen 1957:127)
- b. *con* người
 CL(ani.) human
 ‘human being’ (N2.08) (Tran 2018).

1.2.3 Inanimate classifiers

The classifier *cái* (inanimate) is widely recognized as a general inanimate classifier in Vietnamese. It was found to be the most frequent among 113 actual inanimate classifiers identified in Tran (2018)’s corpus with 235 occurrences, accounting for 25% out of 930 inanimate classifier tokens in the corpus. Other common classifiers including *cây* (tree, long), *quả* (fruit, round), *chiếc* (individual) and *hòn* (stone, round) occur 29 to 140 times each, accounting for 3% to 15% of all

the inanimate classifier tokens in the corpus (Tran 2018). In many cases, a specific classifier can be used instead of *cái* (inanimate) without a functional difference as in (15a-d) (Tran 2018).

- (15) a. *cái* *nhà* *này*
 CL(inani.) house this
 ‘this house’ (N1.161)
- b. *một ngôi* *nhà* *kia*
 one CL(house) house that
 ‘that house’ (N1.156)
- c. *cả ba* *tòa* *nhà*
 all three CL(building) house/building
 ‘all three buildings’ (N1.127)
- d. *một dãy* *nhà*
 one CL(row) house
 ‘a row of houses’ (N2.99) (Tran 2018).

The general classifier *cái* (inanimate) appears with the noun ‘*nhà*’ (house) to make it individuated as in (15a). The classifier *ngôi* (house) combines with this noun as in (15b) to denote a normal house, while *tòa* (building) goes with the same noun as in (15c) to indicate a big house or a building. Additionally, *dãy* (row) used with this noun refers to a row of houses as in (15d). Thus, these four classifiers can be used interchangeably with the noun ‘house’ or ‘building’ without a functional difference although each of the classifiers appears to add different properties to the noun. Specifically, combining with the noun ‘*nhà*’ (house), *cái* (inanimate) just individuates and classifies the noun as an inanimate thing, while *ngôi* (house) individuates and makes the noun sound more literary. Going with *tòa* (building), the noun *tòa nhà* (CL house) indicates a big house or a building, while appearing with *dãy* (row), the noun *dãy nhà* (CL house) means a row of houses. This evidence shows that one noun can be used with several different classifiers, but each classifier may be selected to describe different properties of the noun phrase (Tran 2018).

There are several exceptions in the use of *cái* (inanimate) in Vietnamese. Some nouns denoting small living things can go with *cái* (inanimate) such as *cái kiến* (CL ant) or *cái ong* (CL bee) (D. H. Nguyen 1957:127). However, I will not look at these exceptional cases in which inanimate classifiers are used extendedly for animate nouns, limiting my study to inanimate

classifiers occurring with nouns indicating non-living things in the corpora only. Inanimate classifiers with animate nouns are left for future research. In sum, classifiers in Vietnamese fall into three semantic types: human, animate non-human, and inanimate. This study limits its focus to a corpus-based investigation on how inanimate classifiers are used in Vietnamese.

1.3 Rationale and scope of the study

The fact that the Vietnamese classifier system is highly complex with a large number of classifiers captured my mind, and I became more interested in examining Vietnamese classifiers after doing the General Paper study (Tran 2018) on this topic. There have been prior studies on Vietnamese classifiers in written language which primarily work on constructed or elicited utterances and the basic vocabulary (Emeneau 1951; D. H. Nguyen 1957; Thompson 1965; P. P. Nguyen 2002). No previous research has extensively investigated Vietnamese classifiers in spoken discourse or on an extensive corpus-based study. Daley (1998) worked on classifiers in a corpus-based study, but it is a small-size corpus of four folktale stories. Pham and Kohnert (2008) is a corpus study but they only analysed *con* (animate) and *cái* (inanimate). According to McEnery and Wilson (1996), studies on language corpora may reveal much information on frequency and distribution of linguistic patterns in actual language use. Therefore, with an aim to explore how inanimate classifiers are actually used in naturalistic data, it is necessary to carry out an investigation on Vietnamese classifiers in a corpus study.

Furthermore, variation is an inherent part of language (Labov 1969) and variation in language can be observed in daily life from a piece of news in the newspaper to a conversation (Tagliamonte 2012). In fact, we use language in a variety of ways and interact with many different text types as well as via various means of media. To capture how a complex classifier system is used in a language such as Vietnamese, investigating classifier use in various genres in different modes of discourse as well as among different age groups is expected to identify interesting patterns of classifier use. Since the use of classifiers in actual oral speech is hypothesized to be different from their use in written texts, it is important to examine classifier use in spoken Vietnamese with an attempt to explore variation in classifier frequency and distribution across different genres as well as in spoken and written discourse. However, until now research studies have focused on analyses on the functions of Vietnamese classifiers and what nouns each classifier is used with. No single study has examined the variety and variation of classifier use across

different genres. Thus, in my dissertation, with the aim to capture a broad and realistic picture of Vietnamese inanimate classifiers, I will investigate the use of inanimate classifiers in three genres: narratives, current online newspapers, and conversations in talk shows in Vietnamese.

The reason for my choice of examination of inanimate classifiers in Vietnamese in three different genres is as follows. For written language, I have chosen to work on two genres: folktale fiction genre and online newspaper nonfiction genre. As a popular conventional genre of daily language in the past, the folktale stories, which can be real or imaginary, were written in an informal and informational way. With the choice of this genre, the language use is expected to be of a traditional style, which may be conservative and even archaic. Very different from the folktales, the articles in the online newspaper genre provides factual information and views. The newspaper articles reflect contemporary written language used by a variety of writers who may be journalists, scientists, or even readers. Thus, comparing the uses of inanimate classifiers in these two different genres of Vietnamese written language is expected to find out differences in classifier use diachronically. Furthermore, with an attempt to explore synchronic variation in classifier use in written and spoken discourse, the oral speech nonfiction genre is to be selected for investigation and to be compared with the current online newspaper. For this genre, the analysis of inanimate classifiers in the talk show episodes is expected to clarify how they are used in spoken language, which is naturally occurring, conversational, and interactive. In short, these three genres are different regarding their characteristics, purposes of communication, target audience, time frame and settings. With an investigation of inanimate classifiers in these genres, the dissertation is expected to identify their use patterns and differences in spoken and written discourse as well as among different age groups. The findings of the study are expected to contribute to the knowledge of the Vietnamese classifier system specifically and of the world languages generally.

As I have mentioned above, in this study I examine inanimate classifiers only due to the following two reasons. First, I focus on inanimate classifiers because they are of the largest number of classifiers in Vietnamese compared to human and animate non-human classifier types (Tran 2018). This semantic classifier type goes with numerous nouns describing the nature of the world. Second, narrowing the topic of the dissertation helps me do an in-depth investigation into one semantic classifier type instead of attempting all the three types in Vietnamese.

1.4 Objectives of the study

With the aim to explore how inanimate classifiers are used in Vietnamese, the dissertation investigates all inanimate classifiers that appear in three corpora. The main purpose of this study is to identify what inanimate classifiers are used and how they pattern across three different genres, especially their uses in spoken discourse since this has never been examined in prior research to date. Then the use of inanimate classifiers in the spoken discourse corpus is compared to their usage in the written corpora with an attempt to find out similarities and differences of classifier use in written and spoken discourse with respect to frequency and distribution. The frequency of classifiers among different age groups in the spoken corpus is also examined and discussed.

The study focuses on the uses of *cái* (inanimate) and double classifiers. The use of *cái* (inanimate) is the focus of the study because, as mentioned above, it is hypothesized that this classifier is more frequently used in spoken discourse than in written discourse. As double classifiers in Vietnamese are a special construction, their uses and the functions of classifiers in this construction are also the focus of this study. The findings of Tran (2018)'s study show that there are a limited number of double classifiers, and the function of each classifier in the double classifier constructions varies. Although they rarely appear in the narrative corpus of Tran (2018)'s study, it is observed by the researcher as a Vietnamese native speaker that double classifiers are used more often in spoken discourse than in written discourse, especially the construction with *cái* (inanimate) in the first position. This study further explores other frequent classifiers in the corpora with an attempt to identify how they are distributed across the three genres. The findings about the uses of classifiers in the corpora are also compared to the results of previous research. With these objectives of the study, my research questions are:

1. How are inanimate classifiers used in different genres in Vietnamese?
2. Is there variation in classifier use in spoken and written discourse and among different age groups with respect to frequency and distribution?
3. How are *cái* (inanimate) and double classifiers used across genres? What are the lexical semantic functions of each classifier in the classifier doubling construction?

Besides these questions, I look for the answers to the related issues regarding the number of inanimate classifiers in Vietnamese, the typical classifier construction, and the definiteness of

the noun in relation to the classifier because the results of this study will reveal the answers to these questions. With the findings of frequency, distribution and variation in inanimate classifier use in the three corpora, the study is expected to make substantial contributions to the knowledge of how such a complex classifier system of Vietnamese is used in naturalistic data of Vietnamese corpora, especially spoken Vietnamese. Based on the results, implications for teaching Vietnamese language will be considered.

1.5 Organisation of the thesis

The thesis consists of 6 chapters. Following this introductory chapter, Chapter 2 reviews literature for the research. In the chapter, classifier systems and primary functions of classifiers from a typological perspective are presented. The Vietnamese noun phrase is also briefly described. Then the Vietnamese classifier system with its functions, number of actual classifiers, classifier constructions, and characteristics of classifiers in Vietnamese are introduced. Different kinds of classifiers including general classifiers, specific classifiers, type classifiers, mensural classifiers or classifiers showing contents, event classifiers, and double classifiers are reviewed in the categorization of classifiers. I also discuss the theoretical framework of this study and some key terms used in the study in chapter 2. Finally, the approach I employ for this study with the focus on inanimate classifiers in Vietnamese is presented in the summary section of chapter 2.

Chapter 3 presents the data used for this study, criteria for identifying a classifier in Vietnamese, and methodology applied in the research. First, the three corpora that the current study works on are fully described. Then criteria for identifying an inanimate classifier in Vietnamese are defined based on the foundations of the study. Finally, the methodology that is employed in this study is presented. Specifically, this chapter also presents how the data is organized and analyzed.

Chapter 4 reports the major findings of inanimate classifier use in the study with respect to frequency and distribution. First, the findings on the overall uses of classifiers in the three corpora of this study are presented. Then, the frequency and distribution of the most frequent classifiers in each of the three corpora are analyzed with a variety of examples to illustrate how they are used in the three genres and which nouns they appear with. Moreover, the distribution of some

infrequent classifiers in the corpora are presented with examples for illustration. Exceptional cases in the spoken corpus are also analyzed in this chapter.

Chapter 5 discusses an overview of Vietnamese classifier use across the three genres. This shows how differently inanimate classifiers are used in terms of frequency and distribution in the three corpora. Then the chapter compares the use of inanimate classifiers in written and spoken discourse to see how different classifiers are used in these modes of discourse. The chapter discusses the differences in classifier frequency in written and spoken discourse as well as among different age groups in the Spoken Corpus. The use of double classifiers and the lexical semantic functions of each classifier in this doubling construction are also discussed in the chapter. This chapter examines the use of *cái* (inanimate) extensively since it is very frequently used in the three corpora, especially in the spoken corpus. The hypothesis that this classifier can even occur with non-classified nouns or optional-classifier nouns is also discussed in the chapter. Additionally, classifiers functioning as nominalizers are analyzed and discussed. Furthermore, other findings about the typical classifier construction in Vietnamese and the definiteness of the noun regarding the presence of a classifier are also discussed in this chapter.

Finally, Chapter 6 summarizes the major findings and conclusions for the study. It then recommends implications for teaching Vietnamese language for native and second language learners of Vietnamese.

Chapter 2

Literature review

2.1 Introduction

There have been various classifier analyses, some of which work on classifiers from the typological viewpoint such as Greenberg (1972, 1974), Allan (1977), Dixon (1986), Adams (1989), Denny (1976), Bisang (1993, 1999), Löbel (2000), Aikhenvald (2000), Grinevald (2015). Other studies focus on Vietnamese classifiers including Emeneau (1951), D. H. Nguyen (1957), Thompson (1965), Hoang (1996), P. P. Nguyen (2002), H. T. Nguyen (2004, 2013), and Simpson and Ngo (2018). In fact, researchers do not categorize classifiers in the same way. They focus on different characteristics when defining and categorizing classifiers. As Grinevald (2015) claims, a great variety of classification systems and great variability within different types of systems challenge classifier studies. This chapter reviews the primary literature on classifiers from the typological viewpoint generally and on the Vietnamese classifier system specifically.

In this chapter, first I will introduce classifier systems of the world languages from a typological perspective in section 2.2. It begins with Allan (1977)'s two criteria to define classifiers and Aikhenvald (2000)'s definition of classifiers typologically, then Allan (1977)'s four types of classifier languages including numeral classifier languages, concordial classifier languages, predicate classifier languages, and intra-locative classifier languages. Next, four main types of classifier systems in the world languages based on morpho-syntactic features proposed by Grinevald (2015), which are numeral classifier, noun classifier, genitive classifier, and verbal classifier, are presented in 2.2.1. The four primary functions of classifiers proposed by Bisang (1993, 1999) including individualization, classification, referentialization, and relationalization are mentioned in 2.2.2. The numeral classifier construction claimed by Greenberg (1972) is introduced in 2.2.3. Next section 2.3 briefly describes the Vietnamese noun phrase. Then the Vietnamese classifier system is fully reviewed in section 2.4. An overview of the Vietnamese classifier system and definitions of Vietnamese classifiers are presented in 2.4.1, while the number of Vietnamese classifiers and Vietnamese classifier construction are discussed in 2.4.2 and 2.4.3 respectively. The functions of Vietnamese classifiers are reviewed in 2.4.4, and their characteristics in 2.4.5. Section 2.5 categorizes classifiers in Vietnamese. The two primary types: general classifiers and specific

classifiers are described in subsections 2.5.1 and 2.5.2 respectively. Other types of classifiers in Vietnamese are also introduced, in which type classifiers are reviewed in 2.5.3, classifiers showing contents in 2.5.4, event classifiers in 2.5.5, and double classifiers in 2.5.6. Section 2.6 presents the variationist framework and some key terms used in the study. Finally, section 2.7 summarizes the literature of Vietnamese classifiers and presents the position as well as the approach that will be employed in the current research.

2.2 Classifier systems

2.2.1 Classification of classifier systems

Allan (1977) states that there are two criteria to define classifiers. First, classifiers occur as morphemes in surface structures under specifiable conditions. Second, they have meaning, in the sense that a classifier denotes some salient characteristic of the entity to which a noun refers. Classifiers could be words to categorize word classes based on an attribute such as animacy, form, shape, or size (Aikhenvald 2000). According to Allan (1977), there are four types of classifier languages: numeral classifier languages, concordial classifier languages, predicate classifier languages, and intra-locative classifier languages. Numeral classifier languages, as defined by Allan (1977), are the paradigm type, in which in many expressions of quantity a classifier is obligatory. In all numeral classifier languages, the classifiers occur not only in anaphoric or deictic expressions but also in expressions of quantity. Vietnamese, Burmese, Thai, and Khmer languages are of this type, as claimed by Allan (1977). Concordial classifier languages are those in which classifying formatives are affixed, usually prefixed, to nouns, plus their modifiers, predicates, and pro-forms. Many African (Bantu and Semi-Bantu) and Australian languages belong to this type. Predicate classifier languages are the ones in which “verbs of motion/location consist of theme and a stem which varies according to certain discernible characteristics of the objects or objects conceived as participating in an event as actor or goal” as in Navajo (Allan 1977:287). Intra-locative classifier languages are those in which noun classifiers are imbedded in some of the locative expressions obligatorily accompanying nouns in most environments.

Grinevald (2015) finds it apparently necessary to acknowledge the diversity of classifier system and the existence of several subsystems of classifiers. She proposes four main types of classifier systems in the languages of the world based on morpho-syntactic features: numeral

classifier, noun classifier, genitive classifier, and verbal classifier. According to her, numeral classifiers (Num + CL) are used in “quantifying expressions”, which is the best known and is considered as the “prototype of classifiers” (Grinevald 2015:812). Noun classifiers (CL + Noun) are classifiers which just occur with a bare noun, not linked to the quantifying expressions or possessions. Genitive or possessive classifiers (poss + CL) are morphemes attached to constituents of possessive constructions. Verb classifiers (verb + CL) are morphemes referring to nominal arguments inside the verb form, often linked to processes of incorporation (Grinevald 2015).

Greenberg (1972, 1974) claims that one major difference between classifiers and non-classifiers is that classifiers add no information or have no meaning other than ‘unit’ in a numeral phrase. However, Adams (1989) states that classifiers in Vietnamese can also function as nouns. From a semantic, cognitive, and cultural viewpoint, the function of classifiers is “to communicate a few especially important classes that objects fall into by virtue of the way we interact with them” while nouns have the function of establishing references to things in the world (Denny 1976:125).

In sum, there exist a variety of classifications of classifier systems and subsystems in the world’s languages. Each of them is based on certain criteria. It is thus important to acknowledge the diversity of classifier systems and the existence of several subsystems of classifiers as Grinevald (2015) recognizes. In the next section, the main functions of classifiers put forward by Bisang (1999) will be presented.

2.2.2 Functions of classifiers

With the focus on the functions of classifiers, Bisang (1993, 1999) proposes four operations of nominal concretization used in classifier systems: individualization, classification, referentialization, and relationalization. According to him, the operation of classification is “subsuming a given entity under a certain class according to its properties” (Bisang 1999:115). Classification can be employed to differentiate one “particular sensory perception and its properties” to the other sensory perceptions’ properties to “identify that particular perception by subsuming it under a certain concept”, which is called identification (Bisang 1999:115). Classification can also be used to “establish a sensory perception as an individuum by actualizing the inherent properties which constitute its conceptual unity”, which is called individualization (Bisang 1999:115). Identification can take place without referring to individualization. However,

it seems difficult to “individualize a sensory perception” without identifying it, so the identification function of the classifier “forms the point of departure for classifiers to take on the function of referentialization” (Bisang 1999:116). He also claims that realization is connected to the previous identification of the head noun before it can be modified by a possessor or a relative clause. Classification and individualization, which have often been the only functions described in connection with numeral classifiers, are the primary ones.

According to Bisang (1999), the two primary functions: classification and individualization (including identification) are present in all the classifier languages of East and Southeast Asia, including Vietnamese. These two main functions can explain for the fact that classifiers are used in the context of counting, but individualization “does not necessarily imply that classifiers must be used obligatorily with counting” as Bisang (1999:116) claims. This is true for Vietnamese classifiers. Seiler (1986:95) points out that “classification is a mental operation that causes an object or a multitude of objects to fall under a concept X”. Based upon classification, a set of properties forming a certain class are found. Although these two operations are necessary conditions for a language to be called a classifier language, identification of reference is always described as a “subsidiary or secondary function” of classifiers (Seiler 1986:107).

What I have just presented are the primary functions of classifiers generally and two main functions of classifier languages of East and Southeast Asia, including Vietnamese. Next comes the numeral classifier constructions.

2.2.3 Numeral classifier constructions

Greenberg (1972) established four possible constituent orders in numeral classifier constructions as follows.

- (i) [NUM – CL] – N
- (ii) N - [NUM – CL]
- (iii) [CL – NUM] – N
- (iv) N - [CL – NUM]

Aikhenvald (2000) claims that the Vietnamese numeral classifier construction is the first order. This means Numeral - Classifier - Noun is the Vietnamese numeral classifier pattern. However, in the Vietnamese classifier construction, the classifier combines with the noun before with the

numeral as Simpson and Ngo (2018) argue. This will be further reviewed in 2.4.1 and 2.4.3.

In sum, Vietnamese is one of the numeral classifier languages with the constituent order of Numeral - Classifier - Noun (Aikhenvald 2000). What has been presented in section 2.2 are the classifier systems of the world languages from a typological perspective. In the next section, the Vietnamese noun phrase will be briefly reviewed.

2.3 The Vietnamese noun phrase

Researchers of Vietnamese all recognize the construction of Vietnamese noun phrases: premodifiers + (classifier) + head noun + postmodifiers (Diep 2005:410). The classifier may be present in the Vietnamese noun phrase. The classifier in the Vietnamese noun phrase is the “classifying word” by grouping the objects referred to by the following noun in a general type (Diep 2005:411). Premodifiers are the ones to modify the quantity, while postmodifiers are the items that modify the quality of the noun, demonstratives, and possessive morphemes (Diep 2005). The numeral in Vietnamese can be a cardinal number or a quantifier such as ‘mấy’ or ‘vài’ (several, some), ‘nhiều’ (much, many, a lot of), ‘không ít’ (not little), ‘mỗi’ (every, each). Plural markers such as ‘các’, ‘những’ can also be in the position of the numeral.

He analyses an example of a noun phrase in Vietnamese as in (16), the head noun ‘mèo’ (cat) is unmarked for number and pre-modified by the quantifier ‘tất cả’ (all) and the plural marker ‘những’. It is postmodified by the adjectives ‘đen’ (black), ‘xinh đẹp’ (beautiful), ‘dễ thương’ (adorable), the demonstrative ‘ấy’ (that), and the possessive ‘của nhà Giáp’ (of the Giap family).

(16) Tất cả những *cái* *con* mèo đen xinh đẹp dễ thương ấy của nhà Giáp

All PL CL(inani.) CL(ani.) cat black beautiful adorable that of house Giap

‘All those adorable beautiful black cats of the Giap family’ (Diep 2005:412).

Since the head noun unmarked cannot be individuated or counted, it requires the presence of a classifier. In this case, the two classifiers *cái* (inanimate) and *con* (animate) precede the head noun ‘mèo’ (cat). This is a special phenomenon in the Vietnamese classifier system because a general inanimate classifier co-occurs with a general animate classifier. This phenomenon is unusual, but in fact, it is not a rare case. This receives a lot of argument from various researchers, especially about the function of the extra *cái* (inanimate). However, these two classifiers co-occur

before an animate noun, which will not be investigated in the present study. I will further review the co-occurrence of two classifiers in section 2.5.6.

Nouns in Vietnamese are morphologically unmarked for number. A bare noun can, therefore, refer to one or more than one entity as in (17a), or to mass substances as in (17b).

- | | |
|------------------|--------------------------|
| (17) a. mèo | b. đường |
| cat | sugar |
| ‘(a/the) cat(s)’ | ‘(the) sugar’ |
| c. *hai mèo | d. *hai đường |
| two cat | two sugar |
| ‘two cats’ | ‘two kilograms of sugar’ |
| e. hai con mèo | f. hai kí đường |
| two CL(ani.) cat | two kilogram sugar |
| ‘two cats’ | ‘two kilograms of sugar’ |
- (H. T. Nguyen 2013:59).

In Vietnamese, there is no lexical distinction between count and mass nouns (H. T. Nguyen 2013). Although Vietnamese nouns may refer to discrete entities, they are typically like English mass nouns (Thompson 1965). They cannot be directly counted without the presence of a classifier as in (17c) or measure phrase (17d). In other words, nouns in Vietnamese are non-individuated and thus they need to be individuated via classifiers or measure phrases before they can be counted or measured as in (17e-f) (H. T. Nguyen 2013:59). In sum, the Vietnamese noun phrase has been reviewed with its possible constituents. The Vietnamese classifier system and its properties will be introduced in subsection 2.4.

2.4 The Vietnamese classifier system

2.4.1 Overview of Vietnamese classifier system

Numeral classifier is one of the four classifier systems discussed by Grinevald (2015). Vietnamese is claimed to be a numeral classifier system with the construction of Numeral - Classifier - Noun by researchers including Allan (1977), Aikhenvald (2000), Bisang (1999), and P. P. Nguyen (2002). I agree with their claims since the Vietnamese classifier system appears to be the numeral classifier language, one of the four classifier languages categorized by Allan

(1977). That means, the classifier in Vietnamese is needed to precede nouns so that nouns can be counted or individuated. However, as D. H. Nguyen (1957) points out, based on Grinevald (2015)'s morpho-syntactical categorization, Vietnamese also has noun classifiers and verbal classifiers in addition to numeral classifiers. According to D. H. Nguyen (1957), there are a fair number of noun classifiers with over 100 being common in Vietnamese. This is understandable since several types of classifiers may co-occur in a single language (Craig 1992). When several types of classifiers exist in a language, the labelling of the classifier system is usually based upon the majority of classifiers and/or the primary functions of classifiers in that language, but not all classifier types that occur in the language. Therefore, Vietnamese is the numeral classifier system based on the primary functions of the majority of classifiers.

Prior research on classifiers in Vietnamese including Emeneau (1951), D. H. Nguyen (1957, 1997), T. C. Nguyen (1975), Thompson (1965), and P. P. Nguyen (2002) are primarily on descriptive grammar. However, T. T. Hoang (1996) describes differences in classifier use in language styles or subgenres. He discusses that classifiers are used differently in terms of frequency, distribution, structural and semantic features in the language of literature, science, politics and administration. Investigating 45 classifiers in 14 literary works, he found that the majority of classifiers had low frequency, while only 11 classifiers which can combine with words in surprising and rich expressions have higher rates. They are *cái* (inanimate), *chiếc* (individual), *con* (animate), *mảnh* (thin piece), *miếng* (small piece), *tấm* (large thin piece), *nỗi* (worry, sad, scare), *niềm* (sentiment), *sự* (event), *việc* (activity), and *cuộc* (life, strike, match) (T. T. Hoang 1996:5). He found that the frequency rate of classifiers in prose is higher than in poetry due to the characteristics of literature works, expressive capability of semantic traits of classifiers, and artistic inspiration of writers. He claims that *cái* (inanimate) does not occur often in prose because of the expressive rhetoric purpose, but it neutralizes nuances and generalizes the meanings of the following nouns. He also analyses that in poetry, *con* (animate) was “put on a new coat” in terms of meaningful and expressive nuances (T. T. Hoang 1996:5). Having many commonalities regarding semantic and grammatical features with *cái* (inanimate), *chiếc* (individual) occurs with plentiful semantic nuances such as lonely, single, fragile, unstable, small, and little. He concludes that studies on classifiers in the perspective of pragmatics are needed to satisfy the practical requirements of communication.

It has been claimed by Simpson and Ngo (2018:213) that in Chinese, Japanese and Korean, the presence of a classifier is usually required when numerals are combined with “entity-denoting count nouns” whereas in Vietnamese, not all nouns require classifiers when combined with numerals. For them, nouns in Vietnamese can be divided into three basic types: ‘obligatory-classifier nouns’, ‘optional-classifier nouns’, and ‘non-classified nouns’ as in (18)-(20).

Obligatory-classifier nouns

- | | |
|---|--|
| <p>(18) a. hai *(con) chó
two CL dog
‘two dogs’</p> <p>c. hai *(bông) hoa hồng
two CL flower rose
‘two roses’</p> | <p>b. bốn *(cuốn) sách
four CL book
‘four books’</p> <p>d. hai *(chiếc) xe đạp
two CL vehicle cycle
‘two bicycles’</p> |
|---|--|

Optional-classifier nouns

- | | |
|---|--|
| <p>(19) a. bốn (căn) phòng
four CL room
‘four rooms’</p> <p>c. hai (khối) thiên thạch
two CL sky stone
‘two meteorites’</p> | <p>b. tám (cái) làng
eight CL village
‘eight villages’</p> <p>d. hai (cái) rạp chiếu phim
two CL house to.show movie
‘two cinemas’</p> |
|---|--|

Non-classified nouns

- | | |
|---|---|
| <p>(20) a. hai màu
two color
‘two colors’</p> <p>c. hai vương quốc
two king nation
‘two kingdoms’</p> | <p>b. hai nước
two country
‘two countries’</p> <p>d. hai chính phủ
two government
‘two governments’</p> |
|---|---|

(Simpson and Ngo 2018:214-215).

According to Simpson and Ngo (2018), Vietnamese can be considered a language which presents the empirical evidence Bale and Coon (2014) suggest would clearly identify a Chierchia-type classifier-noun system. They claim that in Vietnamese, whether classifiers are overtly present

or absent in counting constructions is determined by the type of noun, and not by the numeral, regardless of the type of numeral they occur with (Simpson and Ngo 2018:217). As illustrated in (18)-(20), whether the nouns require the overt presence of classifiers, optionally permit the overt use of classifiers, or never allow the overt occurrence of a classifier depends on the type of noun, not the type of numeral. They conclude that Vietnamese is a language in which the overt presence or absence of classifiers mainly appears to be an “idiosyncratic non-predictable property of nouns”, not numerals (Simpson and Ngo 2018:243). They add this provides evidence for a [numeral + [classifier + noun]] syntactic alignment in which classifiers are first combined with nouns before being built together with numerals, rather than the [[numeral + classifier] + noun] alignment that Bale and Coon (2014) argue for, based on the analysis of Chol and Mi’gmaq. According to Simpson and Ngo (2018:243), the Vietnamese patterns support the position that “classifiers are for nouns, not numerals” as assumed in Chierchia (1998), and the hypothesis that “classifiers are for numerals, not nouns” posed in Bale and Coon (2014), cannot be universally correct.

Since Vietnamese nouns do not in themselves “contain any notion of number or amount, they are all somewhat like English mass nouns such as milk, water, flour” (Thompson 1965:193). Most Vietnamese nouns require classifiers to be individuated and counted. On the contrary, in classifier languages, some nouns cannot take a classifier, for instance, the name of time units and/or uncountable nouns (Dixon 1986). Vietnamese has a large number of nouns which do not occur with a classifier (Allan 1977). Apart from the nouns which require a classifier, a number of nouns in Vietnamese do not. These include nouns denoting substance matter, colour, smell, taste, noise; nouns denoting time units such as minutes, days, weeks, months, years, century; nouns denoting geographical areas, regions such as place, district, village, area, province; collective nouns such as nhà cửa (dwelling, house), giày dép (footwear), chăn chiếu (bed clothing), sách vở (books and supplies); or abstract compound nouns such as độc lập (independence), tự do (freedom), dân chủ (democracy), thống nhất (unity), ảnh hưởng (influence), hạnh phúc (happiness), kết quả (result) (D. H. Nguyen 1957:131-132).

2.4.2 Number of Vietnamese classifiers

As one of the isolating languages, Vietnamese tends to have a large number of numeral classifiers (Aikhenvald 2000). Most researchers claim that Vietnamese has a great variety of classifiers although they report different numbers. Emeneau (1951) states that different numbers

of classifiers in Vietnamese are declared in previous studies, and many researchers claim about 140 or 150 classifiers. He affirms that Vietnamese has 121 classifiers, while Adams (1989) estimates that Vietnamese has about 140 classifiers. The difference in the number of classifiers in Vietnamese reported by various researchers can be attributed to differing viewpoints on the definition of classifiers and on the classification of nouns including ‘classified’ and non-classified’, and the identification of the head in noun phrases.

Although researchers have different viewpoints towards classification of nouns and classifiers in Vietnamese, many linguists who claim Vietnamese is of over 100 classifiers have a similar way to identify classifiers in Vietnamese, following the approach of Emeneau (1951). On the contrary, Cao (1988, 1998) has a different approach to Vietnamese classifiers, and recognizes three general classifiers only. He analyses that the structure of two noun phrases in (21) is exactly one and the same. In both cases, there are noun phrases (NPs) with a count noun as the head followed by a mass noun as its qualifier (Cao 1988:41). He refers to not only mass-denoting nouns such as ‘oil’ as in (21b) but also nouns denoting discrete objects such as ‘knife’ as in (21a) as ‘mass nouns’. However, he treated *dao* (knife) in (21a) as the ‘classified’ head noun and *cái* as its classifier, while he analysed *giọt* (drops) in (21b) as the ‘unclassified’ head noun, which is qualified by ‘unclassified’ *dầu* (oil) (Cao 1988:41). This means that only nouns denoting discrete objects such as *dao* (knife) in Vietnamese can be ‘classified’ nouns which are preceded by a classifier such as *cái* (inanimate). However, mass-denoting nouns such as *dầu* (oil) are just ‘unclassified’ nouns which are preceded by another ‘unclassified’ head noun such as *giọt* (drops). In contrast, other researchers including Emeneau (1951), D. H. Nguyen (1957), and P. P. Nguyen (2002) treat such nouns as *giọt* (drops) in (21b) as a mensural classifier preceding the head noun *dầu* (oil). Thus, Cao’s different viewpoint on defining the head in Vietnamese NPs and classifying nouns makes his recognition on classifiers differ from other researchers’.

(21) a. *mấy cái dao*

some CL knife

(Lit. ‘some thing knife’)

b. *mấy giọt dầu*

some drop oil

‘some drops of oil’ (Cao 1988:41)

In sum, different numbers of classifiers in Vietnamese have been reported by researchers. The discrepancy between Cao (1998)'s and other linguists' claims is attributed to his different viewpoint towards the definition of classifiers and classification of 'classified' or 'unclassified' nouns. I am following the approach of the other researchers including Emeneau (1951), D. H. Nguyen (1957), and P. P. Nguyen (2002). In my view, classifiers are words which precede the head and perform the function of classification, individualization, and/or nominalization. Syntactically, classifiers may appear in Vietnamese noun phrases containing a numeral, with or without an overt head noun followed by several potential attributes, and with or without a demonstrative. Semantically, classifiers are unbound function words that categorize the head noun based on inherent or salient features of the noun's referent, such as animacy, shape, length, dimension, function, or material. I will discuss this in more detail in the section of criteria for identifying a classifier in Vietnamese (section 3.2). Furthermore, the number of classifiers which varies among different researchers may also be due to the scale of the study and the norms of constructed and elicited utterances. Studies on a large-scale corpus may reveal more about this issue. I assume that the number of Vietnamese classifiers may be over 200. Despite differences in the number of classifiers claimed, it is undeniable that Vietnamese has a great variety of classifiers.

According to Emeneau (1951:84), the two major subclasses of Vietnamese nouns are "classified nouns" and "non-classified nouns". He claims that a clear distinction can be made between these two subclasses of nouns. For him, the basic vocabulary number of nouns and classifiers is 770, which includes 121 classifiers, 471 classified nouns, and 178 non-classified nouns. He adds each noun may be used in one or two of these subclasses (Emeneau 1951:93). In my viewpoint, categorizing nouns and classifiers in the basic vocabulary seems not objective or appropriate because in this way it does not reflect the nature of actual language use. In brief, differing viewpoints on how to define a classifier and how to classify nouns as well as the data of the study can result in different claims on the number of classifiers in Vietnamese. However, the number of classifiers in Vietnamese is not the focus of this study.

2.4.3 Vietnamese classifier constructions

As mentioned in 2.2.3, Aikhenvald (2000) claims that the Vietnamese numeral classifier construction has the pattern of Numeral - Classifier - Noun, the first in the four possible constituent orders established by Greenberg (1972). The most typical structure of all substantial Vietnamese

classifier constructions is Numeral - Classifier - Noun (D. H. Nguyen 1957). Thompson (1965) also argues that the construction consisting of a numeral preceding a classifier and a noun as in (22) may be the most typical of all Vietnamese classifier patterns.

- (22) *ba cái ghế*
 three CL(inani.) chair
 ‘three chairs’ (Thompson 1965:193).

However, in my prior corpus-based study, the data shows that Classifier - Noun would be potentially the prototypical classifier construction since over 60% of the classifier tokens have this pattern (Tran 2018). The typical classifier pattern might be Classifier - Noun, without the numeral. This suggestion is also made in Daley (1998)’s research. She says that the prototypical classifier construction of Vietnamese might be Classifier - Noun since the majority of classifiers found in her corpus study follow this pattern (Daley 1998). She also argues that the data in her study would challenge the prototypical Vietnamese classifier construction claimed by previous researchers.

According to Emeneau (1951), a numeral classifier construction contains a numeral preceding a classifier and a noun as in (23a), or a demonstrative following the classifier and noun as in (23c), or both a numeral and a demonstrative as in (23d) if it is a classified noun. This construction may consist of all the constituents including a numeral preceding a classifier and a classified noun, followed by an attributive and a demonstrative as in (23f). For non-classified nouns, the construction would contain a numeral and a non-classified noun as in (23b), or a numeral and a non-classified noun preceding a demonstrative as in (23e).

- (23) a. *hai cái cổng*
 two CL(inani.) gate
 ‘two gates’
 b. *hai chuyện*
 two story
 ‘two stories’
 c. *cái cổng ấy*
 CL(inani.) gate that
 ‘that gate’

- d. hai *cái* cổng ấy
 two CL(inani.) gate those
 ‘those two gates’
- e. hai chuyện ấy
 two story those
 ‘those two stories’
- f. hai *cái* cổng gỗ lớn kia
 two CL(inani.) gate wood big those
 ‘those two big wooden gates’ (Emeneau 1951:84).

The numeral classifier construction can be illustrated as in the following schema in (24).

(24) A schema of the numeral classifier phrase constructions (Emeneau 1951:85):

Numeral	Classifier	Classified noun	+/- Attribute(s)	Demonstrative
	Non-classified noun			

Emeneau claims that in a “numerated substantive phrase” in Vietnamese, a noun of the classified type may be omitted but the classifier must remain when the head noun is identified in the preceding context as in (25) (Emeneau 1951:84). In the question in which the classifier precedes the noun *cuốn* sách (CL book), the noun is identified. Thus, in the answer in (25), the classifier *cuốn* (volume) occurs with the demonstrative *kia* (that) in the absence of the noun as the noun is identified in the context. In this case, the entity is definite due to the presence of the demonstrative.

(25) Anh muốn *cuốn* sách nào? - *Cuốn* kia.

You want CL(volume) book which CL(volume) that

‘Which book do you want? - That one.’

(Emeneau 1951:84).

When the head noun is omitted as it has been previously identified, numeral plus classifier is the common construction (Emeneau 1951:91). In this construction, the entity is individuated due to the appearance of the classifier, but the noun is indefinite because of the presence of the numeral. In (26a), with the omission of the noun when it is identified in the preceding context, in the presence of the classifier *quyển* (volume), the entity is individuated and it is indefinite due to the

occurrence of the numeral *một* (one), but no demonstratives. According to Thompson (1965), in many cases the head noun may be omitted when it is identified in the context, so only the classifier occurs with numerals as in (26a) or with demonstratives as in (26b). The example in (26b) is similar as the one in (25) because the classifier occurs with the demonstrative ‘*này*’ (this) with the omission of the noun when it is identified in the context. Clearly, the individuation and the definiteness of the entity is explicitly identified in this case due to the occurrence of the classifier and the demonstrative.

(26) a. Tôi lấy *một quyển*.

I get one CL(volume)

‘I’ll buy one (volume/book).’

b. Tôi lấy *quyển này*.

I get CL(volume) this

‘I’ll buy this (volume/book).’

(Thompson 1965:192)

In contrast, without any “explicit indication of number”, a noun is entirely free from reference to the number category (Emeneau 1951:85). That means, a “non-numerated substantive phrase” or a non-numeral classifier phrase which has no “indication of number or of individuation” has neither numerals nor demonstratives as in (27) (Emeneau 1951:85). In this case, the individuation and definiteness of the entity is ambiguous since the noun *sách* (book) does not occur with either numerals or demonstratives. In this example, the speaker does not indicate how many books he/she intends to buy. Clearly, in this case, the meaning of the numeral plus classifier combination as in (26a) is different from the case when the noun is unclassified as in (27).

(27) Tôi muốn mua *sách*.

I want buy book

‘I want to buy book(s).’

(Emeneau 1951:85).

In summary, the schema of Vietnamese classifier/noun phrase construction in (24) clearly shows all possible constituents. A classifier phrase in Vietnamese consists of three main constituents in the order: Numeral - Classifier - Classified Noun. However, this construction might

not be the most typical in Vietnamese since the number of classifiers following the pattern of Classifier - Noun may exceed those with the former pattern as in Daley (1998)'s and Tran (2018)'s corpus-based studies. Vietnamese classifiers can occur with numerals, and/or demonstratives in the absence of the noun when it is identified in the context, which shows variation in Vietnamese classifier constructions. The classifier construction is also to be examined and discussed in the current research with an attempt to find out which pattern would be the prototypical in Vietnamese.

2.4.4 Functions of Vietnamese classifiers

Like all the classifier languages of East and Southeast Asia, Vietnamese classifiers have two primary functions: classification and individualization (including identification), according to Bisang (1999:116). Ly (1998) also claims that Vietnamese classifiers perform two functions. The main function of classifiers in Vietnamese is to individuate the object denoted by the noun. The secondary function of classifiers is to classify, characterize or describe objects through definite features (Ly 1998). However, apart from these two primary functions, another function of a classifier, as Löbel (2000:296) claims, is “syntactic referentialization”, which is strongly connected with particularization. By using the term “syntactic referentialization”, she means a classifier only appears with a noun when the noun has a referent. In other cases when the noun does not have a referent, the classifier is not used in the noun phrase. What she means is that Vietnamese classifiers are used with a classified noun when the noun refers to a particular reference. As in (28a), the noun phrase (in square brackets) is syntactically non-referential, whereas in (28b), it is syntactically referential and, therefore, can be modified by a relative clause or other attributive constructions.

(28) a. Trong nhà hát kia có [17 ghế].

In cinema that have 17 chair

‘There are seventeen chairs in this cinema.’

Lit: ‘That cinema is seventeen-chaired.’

b. Trong nhà hát kia có [17 cái ghế làm bằng cây tốt].

in cinema that have 17 CL(inani.) chair made out of wood good

‘There are seventeen wooden chairs in that cinema.’

Lit: ‘There are seventeen chairs in that cinema which are made of precious wood.’

(Löbel 2000:296).

She also argues that in Vietnamese, the occurrence or non-occurrence of a classifier is restricted neither to any special numerals, nor to any lexical selection or “core semantic properties of nouns” (Löbel 2000:296). She adds that the referential function of the classifier can be illustrated by comparing the generic sentence with the non-generic one (Löbel 2000:297). There is another type of construction without the presence of the verb *có* (have, there is/are) as in (29). These are nominal attributes from Truong (1970:246).

- (29) a. máy bay bốn động cơ
 airplane four engine
 ‘four-engined airplane’
 b. máy bay với bốn chiếc động cơ lớn
 airplane with four CL(individual) engine big
 ‘airplane with four big engines’ (Truong 1970:246).

These above examples illustrate contexts where the presence or absence of a classifier in combination with numerals correlates with a difference in meaning, according to Löbel (2000). What she analyses means that Vietnamese classifiers might be required only for the noun which is syntactically referential and modified by attributes as in (28b) and (29b), not for the noun without particular references as in (28a) and (29a). To my understanding, these two nouns do not take classifiers when being used to modify another noun as in (28a) and (29a). They function as modifiers without classifiers in this case. That means, the use of classifiers before a noun may depend on the discourse context and/or pragmatics.

Furthermore, Bisang (1999) discusses that the noun in classifier languages can be omitted from the classifier construction if it is previously identified. In this case, the classifier anaphorically refers to the noun (Bisang 1999). He claims that in almost all classifier languages, especially in most East and Southeast Asian languages, the classifier can combine with numerals and/or demonstratives without the presence of the noun when the noun is identified in the preceding context, and does not occur alone in its anaphoric function in this case, as shown in the example in (30). Thus, apart from the two primary functions, classification and individualization (including identification), Vietnamese classifiers also have anaphoric function.

(30) Đây là sách. Lan mua một *quyển*, tôi mua hai *quyển* này.

This be book Lan buy one CL I buy two CL this

‘These are books. Lan bought one (of them), I bought these two (of them).’

(Bisang 1999:148).

Additionally, according to H. T. Nguyen (2004:168), the subset called ‘event classifiers’ characterizes abstract concepts. An event classifier usually precedes a verb, an adjective or a verbal adjective, or a noun referring to an abstract notion. In most cases, these event classifiers participate in the process of nominalization as in the example in (31) (H. T. Nguyen 2004).

(31) *sự* ném bom

CL throw bomb

‘bomb raid’ (H. T. Nguyen 2004:168).

In summary, besides the two primary functions claimed by Bisang (1999), classification and individualization including identification, the Vietnamese classifiers have anaphoric function. Furthermore, they perform the function of syntactic referentialization (Löbel 2000; Truong 1970), and nominalization (H. T. Nguyen 2004). The characteristics of Vietnamese classifiers will be reviewed in the next section.

2.4.5 Characteristics of Vietnamese classifiers

A noun can combine with more than one independent classifier, and the choice of classifiers depends on a “particular, shape-related, property of the referent” which speakers want to focus on (Aikhenvald 2000:114). As Bisang (1999) analyses, the selection of classifiers depends on many different important interacting factors such as semantics and discourse. He states that style and age are also factors which might influence the choice of classifiers. In Chinese and other languages, some factors determining classifier use include “level of formality, discourse type, presence of the referent, familiarity of the referent, and age of the hearer” (Erbaugh 1986: 413). For many classifier languages, different classifiers may be used with the same noun, depending on the “context and indicating different characteristics of the referents” (Behrens 2003:65).

In many cases, a specific classifier can be used instead of *cái* (inanimate) without a functional difference (D. H. Nguyen 1957). However, the choice of classifiers can vary as in (32)-

(33). The meanings of the noun phrases in (32a) and (32b) are the same. That means *cái* (inanimate) and *chiếc* (individual) can be used interchangeably in this case when going with the noun ‘cake’. In contrast, with the classifiers *dòng* (flow) and *con* (animate), the noun phrases *dòng suối* (CL stream) and *con suối* (CL stream) as in (33b-c) respectively sound more formal and literary than *cái suối* (CL stream) with the classifier *cái* (inanimate) as in (33a). Clearly, the use of specific classifiers can communicate the formality of the language in this case. This means that the use of different classifiers can produce differences in genres or stylistics although it does not make a difference in functions as D. H. Nguyen (1957) claims.

- (32) a. bốn *cái* bánh
 four CL(inani.) cake
 ‘four cakes’ (N1.43)
- b. bốn *chiếc* bánh
 four CL(individual) cake
 ‘four cakes’ (N1.43)
- (33) a. một *cái* suối nước chảy rì rì
 one CL(inani.) stream water flow slowly
 ‘a stream flowing slowly’ (N1.126)
- b. những *dòng* suối nhỏ
 PL CL(current) stream little
 ‘the little streams’ (N2.115)
- c. những *con* suối
 PL CL(ani.) stream
 ‘the streams’ (N2.114).

It seems probable that if the set of classifiers in a language is larger, it will allow for a greater number of choices (Adams 1986:244). Since Vietnamese has a large variety of classifiers, many classified nouns in Vietnamese occur with two or more different classifiers without a difference in meaning (Emeneau 1951:96) as in (34).

- (34) a. *cây* gươm
 CL(long) sword
 ‘(the) sword’

- b. *cái* *guôm*
 CL(inani.) sword
 ‘(the) sword’
- c. *thanh* *guôm*
 CL(long) sword
 ‘(the) sword’ (Emeneau 1951:96).

As shown in the examples in (34), the noun *guôm* (sword) is used with three different classifiers without much difference from the semantic perspective. However, from a pragmatic perspective, each of the classifiers may be used with different intentions of the speaker. It could be that the general classifier *cái* (inanimate) is used to just indicate that the sword is a non-living thing while the specific classifiers *cây* (long object) and *thanh* (long object, thin) are used to depict the physical shape of the sword, which is long with *cây* (long object) or long and thin with *thanh* (long object, thin). Moreover, these specific classifiers are also more formal, especially the later one which appears with the noun and adjective when referring to a ‘precious’ sword. In contrast, the general classifier *cái* (inanimate) is not used when a ‘precious’ sword is the intended meaning. This exemplifies why researchers claim that different classifiers can be used with the same noun in different contexts and/or with different focus of pragmatic semantic indications (D. H. Nguyen 1957; Thompson 1965; P. P. Nguyen 2002).

According to Adams (1986), in some Austroasiatic languages, certain genres require special application of the set of classifiers. “Register such as formal and informal may also indicate a different choice of a classifier” (Adams 1986:244). In Vietnamese, some of Sino-Vietnamese morphemes are classifiers in the literary register (more formal register). This type of register change of classifiers is meaningful in terms of pragmatics, but not semantically, she adds. However, for many other Vietnamese nouns, the use of different classifiers makes the meaning of the noun different. The use of the subclass of nouns indicating names of plant parts as classifiers is the most obvious case. Emeneau (1951:95) discusses that all names of plant species are classified by *cây* (tree) to denote one specimen of the plant as in (35a). The morpheme *cây* (tree) is also used as a classifier for many nouns denoting long, sticklike-shaped objects. The classifier *trái* (fruit, round) as in (35b), which is the synonym of *quả* (fruit, round), is used with names of many plant species to denote the fruits of those and with other nouns to denote round or globe-shaped objects.

Similarly, *lá* (leaf) as in (35c) is used as a classifier for many plant species for denoting the leaf of those plants or with other nouns denoting leaf-shaped objects.

- (35) a. một *cây* cam
one CL(tree, long) orange
'an orange tree'
- b. một *trái* cam
one CL(fruit, round) orange
'an orange'
- c. một *lá* cam
one CL(leaf)
'an orange leaf'² (Emeneau 1951:97).

In Vietnamese, a number of nouns may be classified nouns in some cases and nonclassified nouns in other cases, usually with a different meaning as Emeneau (1951:95) claims. For instance, the noun *cửa* (door) is a classified noun denoting a physical object when appearing with a classifier as in (36a), but it is a non-classified noun denoting the way in or out like 'entrance' or 'exit' without a classifier as in (36b).

- (36) a. cái cửa
CL(inani.) door
'the door'
- b. cửa vào/ra
door into/out
'entrance/exit' (Emeneau 1951:95).

In sum, Vietnamese classifiers categorize the head noun based on the inherent feature or characteristic of the noun's referent such as animacy, shape, size, length, depth, dimension, function, or material. They are mainly function words, but a number of classifiers may belong to the category of content words since they are 'real nouns' in other cases, especially classifiers showing contents. A number of nouns can go with several different classifiers, and the choice of

² In Emeneau (1951), 'orange' appears as the adjective in (35c), but in Vietnamese 'orange' remains as a noun and the classifier categorizes the noun. In (35c), it means *a leaf of a kind of tree 'orange'*.

classifiers depends on the speaker's intention. These are the characteristics of Vietnamese classifiers. Categorization of classifiers in Vietnamese will be reviewed in the next section.

2.5 Categorization of classifiers in Vietnamese

As mentioned earlier, Vietnamese is claimed to be a numeral classifier language by researchers including Allan (1977), Aikhenvald (2000), and P. P. Nguyen (2002) due to the fact that a classifier is obligatorily required in many expressions of quantity as well as in anaphoric or deictic expressions in Vietnamese. According to the researchers whose studies focus on Vietnamese classifiers, in the inanimate classifier type, there are two major kinds of classifiers: general classifiers and specific classifiers (Emeneau 1951; D. H. Nguyen 1957; Thompson 1965; P. P. Nguyen 2002). There are subtypes of classifiers including “type classifiers”, “classifiers showing contents” (D. H. Nguyen 1957:128), and “event classifiers” (H. T Nguyen 2004:168). In addition, two classifiers which co-occur are called double classifiers (Tran 2018). These types of classifiers will be reviewed in more detail in the next subsections.

D. H. Nguyen (1957:128) distinguishes “classifiers showing contents” from general and other specific classifiers which are called “proper classifiers”. Grinevald (2000:64) makes a distinction between two semantic subtypes of numeral classifiers: “sortal or true classifiers” and “mensural or quantitative classifiers”. The terms “proper classifier” or “true classifier”, indicating the same type of classifiers, are used throughout this study. The terms “mensural classifiers” or “classifier showing contents” indicating another type of classifiers, which are akin to measure terms Grinevald (2000:64), are also used in the study. Furthermore, she claims that in numeral classifier systems which are said to have a large number of classifiers, the majority of classifiers are in fact mensural classifiers, while the number of true classifiers is very limited.

2.5.1 General classifiers

The general classifier *cái* (inanimate) is the most common of the classifiers that classify nouns denoting nonliving things (Emeneau 1951; Tran 2018). Emeneau (1951:103) claims that out of the 471 classified nouns in the basic vocabulary, 173 are classified with *cái* (inanimate). According to Löbel (2000), the general classifier *cái* (inanimate) denotes exactly the property of being a nonliving thing. This classifier accounts for a quarter of all the inanimate classifier tokens in the corpus in Tran's (2018) study with 235 out of 930 inanimate classifier tokens as in (37).

- (37) *cái* chum vàng
 CL(inani.) jar gold
 ‘a jar of gold’ (N1.16) (Tran 2018:26).

Although Emeneau (1951) claims that the general classifier *cái* (inanimate) classifies no nouns that denote living things, D. H. Nguyen (1957) points out that nouns denoting small insects can be classified with *cái* (inanimate) as in (38).

- (38) a. *cái* kiến
 CL(inani.) ant
 ‘the ant’
 b. *cái* ong
 CL(inani.) bee
 ‘the bee’
 (Nguyen 1957:127; 144).

2.5.2 *Specific classifiers*

Beyond the three general classifiers, many researchers agree that there are many other individual/specific classifiers in Vietnamese such as *chiếc* (individual), *cây* (tree, long) and *quả* (fruit, round) as in (4)-(6). The classifier *cây* (tree, long) is used to count trees, mushrooms, posts, pillars, fans, guns, axes, writing tools, fuel, candles, lamps, firecrackers (Adams 1986:250). Among 113 inanimate classifiers found in Tran (2018)’s corpus, a number of frequent specific classifiers include *cây* (tree, long object) (14%), *quả* (fruit, round) (5%), *chiếc* (individual) (4%), and *hòn* (a stone, round) 3%. They are among the ten core classifiers of Vietnamese claimed by Löbel including *cái* (inanimate), *cây* (tree, long object), *chiếc* (individual), *con* (animate), *hòn* (a stone, round), *quả* (fruit, round), *quyển* (a volume), *sợi* (a hair, thread, cord), *tấm* (a flat piece of material) and *tờ* (a sheet of paper, document) (Löbel 2000:299). According to Bisang (1999:139), these ten classifiers particularly refer to inherent properties of the noun. He claims that they actualize the semantic boundaries which already belong to the concept of a given noun, and this kind of actualization also takes place with more specific classifiers. Löbel (2000:272) states that the classifier *cây* (tree, long object) denotes the class of the subclass *rau* (vegetable plant), and accordingly, can be used as “designating the property of being a plant”. Both *cái* (inanimate) and

cây (tree, long object) denote a property which is inherent to the meaning of the classified noun as in (39a-b), she adds. It is similar for *chiếc* (individual) and *quyển* (a volume) as in (39c-d).

- (39) a. hai cái cây
two CL(inani.) tree/plant
'two trees/plants'
- b. hai cây rau
two CL(tree, plant) vegetable
'two vegetables'
- c. hai chiếc xe
two CL(individual) car
'two cars'
- d. hai quyển sách
two CL(volume) book
'two books' (Löbel 2000:272-273).

2.5.3 Type classifiers

Type classifiers such as *thứ* (kind, sort) classify most nouns when one sample of the species is to be distinguished in quality from another sample (Emeneau 1951: 109). For instance, a piece of silk is compared to another piece of silk in quality as in (40).

- (40) *Thứ lụa này đắt hơn thứ ấy*
CL(kind) silk this expensive than CL(kind) that
'This kind of silk is more expensive than that kind' (Emeneau 1951:109).

2.5.4 Classifiers showing contents

According to D. H. Nguyen (1957:127), when the material thing designated is not discrete, the classifier indicates *quantity* rather than *number*. For instance, *ấm* (kettle, teapot) and *bát* (eating bowl) denote units of measurement or 'contents', in which case these words require the general classifier *cái* (inanimate) as in (41a). However, according to researchers including Emeneau (1951) and D. H. Nguyen (1957), *ấm* (teapot) and *bát* (eating bowl) are also classifiers showing contents or mensural classifiers as in (41b-c).

- (41) a. một *cái* *ấm*
 one CL(inani.) teapot
 ‘a teapot’
 b. một *ấm* trà
 one CL(pot) tea
 ‘a potful of tea’.
 c. hai *bát* cơm
 two CL(bowl) rice
 ‘two bowlfuls of rice’ (D. H. Nguyen 1957:128).

Li and Thompson (1981:106) state that “any measure word can be a classifier”. However, D. H. Nguyen (1957) argues that classifiers showing contents are distinguished from measure words in Vietnamese as in (17f) repeated below for ease of comparison. He proposes that unit names such as kilogram, meter are not classifiers. In contrast, H. T. Nguyen (2013) makes a distinction between classifiers and measure words. In his analysis, he does not treat measure words as in (41b-c) as classifiers while D. H. Nguyen (1957) treats them as classifiers showing contents, but not unit names. I am following D. H. Nguyen (1957) on this point as these NPs actually denote the contents, not their containers.

- (17) f. hai *kí* đường
 two kilogram sugar
 ‘two kilograms of sugar’ (H. T. Nguyen 2013:59).

2.5.5 Event classifiers

According to H. T. Nguyen (2004:168), all classifiers designating actions, states, processes, or activities are categorized into a subset called “event classifiers”. This subset which characterizes abstract concepts differs from the other subgroups denoting concrete entities. In most cases, these event classifiers take part in the process of nominalization as in (42) (H. T. Nguyen 2004).

- | | |
|---|---|
| <p>(42) a. <i>sự</i> thật
 CL true
 ‘truth’
 (H. T. Nguyen 2004:168).</p> | <p>b. <i>nền</i> văn hóa
 CL culture
 ‘culture’</p> |
|---|---|

An event classifier usually precedes a verb, an adjective, or a noun referring to an abstract notion. The verbs or adjectives following the event classifier may be considered as abstract nouns representing action, with the classifier having the function of nominalizing the verbs (Nguyen 2004).

2.5.6 Double classifiers and an extra *cái* (inanimate)

In Vietnamese, two classifiers can co-occur before a noun. This is a special phenomenon which receives a lot of argument from various researchers of Vietnamese. Below are the examples of an extra *cái* (inanimate), in which it appears in the doubling construction with an animate classifier *con* (animate) as in (43a), with the noun indicating human in (43b), or with ‘mass’ noun as in (43c). It is called ‘extra’ by the researchers as they claim its appearance is optional.

- (43) a. *cái con voi (mà) anh thấy lúc nãy*
 CL(inani.) CL(ani.) elephant which you see while ago
 ‘the elephant (which) you saw a while ago’
- b. *Cái người (mà) tôi vừa chào*
 CL(inani.) human whom I just greet
 ‘the person (whom) I just said “hello” to’
- c. *Cái độc lập (mà) chúng tôi muốn*
 CL(inani.) independence which we want
 ‘the independence (that) we want’
- (D. H. Nguyen 1957:130).

There has been much discussion about whether or not the particle *cái* (inanimate) as in (43a) is a classifier and what function it performs. This particle *cái* (inanimate) is considered to be equivalent as a definite article with the function of a determiner by T. K. Tran et al. (1960). P. P. Nguyen (2002) and Diep (2005) call it a special demonstrative, and claim that it can produce the definiteness for the preceding noun. This particle clearly exists, but it is not easy to identify what it actually is (Diep 2005). Emeneau (1951) and D. H. Nguyen (1957) claim that the use of an extra *cái* (inanimate) preceding *con* (animate) and *người* (human) as in (43a-b) or non-classified nouns as in (43c) is considered to be old-fashioned. They both claim that the noun is identified by an attribute consisting of a relative clause introduced optionally by ‘*mà*’ (which/whom), used as a coordinating conjunction and as a final particle emphasizing the content of a clause. D. H. Nguyen

(1957:131) states that the classifier *cái* (inanimate) preceding such abstract terms as *độc lập* (independence) as in (43c) is likely to “strike” young people as being “incorrect”.

H. T. Nguyen (2013:67) also pays special attention to *cái* in this case and this particle is proposed to be a “focus marker” in the Vietnamese noun phrase, which serves as a formal device to signal that there is a focus in the noun phrase. He analyses this construction with the extra *cái* syntactically. He argues that this particle is homonymous with the classifier *cái* (inanimate) but has a function different from that (or any other) classifier (H. T. Nguyen 2013:65). He adds that it cannot be used before the homonymous classifier. Unlike classifiers, this particle can occur with any kind of nouns, whether classified nouns, non-classified nouns as in (44a), mass-denoting nouns as in (44b), or measure nouns as in (44c), as he claims.

(44) a. *cái* ngày ấy

cái day that

‘that very day’

b. *cái* thịt ít mỡ

cái meat little fat

‘the lean meat’

c. Uống thử *cái* ấm trà này coi có ngon không.

drink try *cái* pot tea this see QUES delicious QUES

‘Try this very potful of tea to see if it’s delicious.’

(H. T. Nguyen 2013:66)

However, in this special construction, the extra *cái* (inanimate) significantly strengthens the interpretation of definiteness (Simpson and Ngo 2018). According to Simpson et al. (2011), a combination of Classifier - Noun may be interpreted as either definite or indefinite, depending on the context. However, the potential ambiguity in (in)definiteness disappears completely when an extra *cái* (inanimate) is added, and this particle thus forces a definite interpretation (Ngo 2012; Simpson 2008; H. T. Nguyen 2004; D. H. Nguyen 1957). It is noted that the special use of the extra *cái* (inanimate) in this construction is always phonologically stressed although generally classifiers in numeral constructions do not receive phonological stress (H. T. Nguyen 2013; Simpson and Ngo 2018). This prosodic property of the extra *cái* (inanimate) along with its necessary definiteness are useful for identifying which function any occurrence of *cái* (inanimate)

has in certain situations when only one classifier occurs. In order to distinguish the two functions of *cái* (inanimate), they “gloss the regular use of the classifier *cái* (inanimate) simply as ‘CL’ and the special, definite use of *cái* (inanimate) as ‘CL.DEF’” (Simpson and Ngo 2018:224). According to them, it is possible for an extra *cái* (inanimate) to occur with all three kinds of nouns in Vietnamese: obligatory-classifier nouns, optional-classifier nouns, and non-classified nouns. They state that the extra *cái* (inanimate) must precede the regular classifier for the noun when occurring with obligatory-classifier nouns, and it is ungrammatical to omit the regular classifier, as shown in (45). The extra *cái* (inanimate), consequently, does not replace the regular classifier, but performs another function, relating to definiteness.

- (45) hai *cái* *(cuốn) sách
 two CL.DEF CL book
 ‘the two books’
 (Simpson and Ngo 2018:225)

They claim that it is similar for the case of optional-classifier nouns. That means, the regular classifier for the noun must be present when the extra *cái* (inanimate) occurs despite the optionality of the regular classifier to appear with such nouns in other cases as illustrated in (46a-b). The addition of an extra *cái* (inanimate) thus has a clear effect on optional-classifier nouns and constrains the optionality of classifiers to such nouns in the absence of the extra *cái* (inanimate) (Simpson and Ngo 2018).

- | | |
|---|--|
| <p>(46) a. hai (<i>người</i>) nhân viên
 two CL(human) employee
 ‘(the) two employees’
 (Simpson and Ngo 2018:225)</p> | <p>b. hai <i>cái</i> *(<i>người</i>) nhân viên
 two CL.DEF CL(human) employee
 ‘the two employees’</p> |
|---|--|

When non-classified nouns combine with the extra *cái* (inanimate), no additional classifier occurs in the structure, and the extra *cái* (inanimate) appears to be able to go directly with the noun, as illustrated in (47).

- (47) a. hai câu/ màu/ góc/ tiếng
 two sentence/color/corner/sound
 ‘(the) two sentences/colors/corners/sounds’

- b. hai *cái* câu/ màu/ góc/ tiếng
 two CL.DEF sentence/color/corner/sound
 ‘the two sentences/colors/corners/sounds’ (Simpson and Ngo 2018:226).

In brief, *cái* (inani.) which appears in the doubling construction with classified nouns, optional-classifier nouns, and non-classified nouns is called an extra *cái* (inani.) by prior researchers. This extra *cái* (inani.) is labelled differently by researchers. A number of researchers do not recognize it as a classifier while others argue that it is a classifier. Despite their disagreement on naming it, the extra *cái* (inani.) has a special function, which is different from other classifiers and even different from the general classifier *cái* (inani.) in other cases. Following Simpson and Ngo (2018), I call it a classifier with a special function. The functions of each classifier in the doubling construction are to be examined in the current study. The next section is my theoretical framework for the study.

2.6 Theoretical framework and key terms

2.6.1 Variationist framework

In this section I will present what I will specifically look at in this research in terms of linguistic variation employing corpora with an aim to study differences in classifier use in Vietnamese across genres. According to McEnery and Hardie (2012), variation can be interpreted in a number of ways: diachronic variation and synchronic variation. Diachronic variation is language change over time while synchronic variation describes differences in a language at a specific point of time, usually the present. In exploring corpus-based approaches to synchronic variation, there are two distinct approaches: a multi-dimensional (MD) approach and a variationist approach. The MD approach looks at “variation across genre (or register), with the individual *text* as the unit of variation” while variationist sociolinguistics considers variation across class, gender or other social category, with the individual speaker as the unit of variation (McEnery and Hardie 2012:94). The MD approach was first introduced in a study by Biber (1986) which aimed to explain certain findings on variation between speech and writing, and between different genres or registers. Biber suggests that investigating the use of a wide range of features of language in different genres and using statistical techniques to integrate them into a complicated and subtle picture of how language is used differently across genres. Also, to measure the frequency of each

of the features within a corpus from a heterogeneous set of genres is important. Variation in language is variation across or within genres or registers where “the unit of variation is the individual text” (McEnery and Hardie 2012:115). In research for the production data contained in a speech, variationists seek for discovering patterns of usage, which is related to the relative frequency of occurrence or co-occurrence of structures, rather than to their existence or grammaticality (Polack 1993:252). This study is thus interested in synchronic variation at the level of the text and discourse. It attempts to uncover the systematic differences in classifier use in Vietnamese among different genres by investigating inanimate classifiers in three corpora.

It is important to do study language variation on corpus because “corpus-based studies typically use corpus data in order to explore a theory or hypothesis, typically one established in the current literature, to validate it, refute it or refine it” (McEnery and Hardie 2012:6). The development of corpus linguistics, as they discuss, has facilitated the exploration of theories which draw their inspiration from attested language use and the findings drawn from it. Thus, working on corpus data allows the production of frequency data of a classifier list, which contains all classifiers appearing in a corpus and specifies how many times each classifier occurs in the corpus. As McEnery and Hardie (2012:2) analyze, “concordances and frequency data exemplify respectively the two forms of analysis”, qualitative and quantitative, that are of equal importance to corpus linguistics. Therefore, investigating variation in classifier use across genres in Vietnamese in a corpus-based study is necessary because the corpus dataset will reveal similarities and differences in their use among different genres in actual writing and speech.

2.6.2 Terms used in the study

In this section, I will introduce some fundamental terms that I will use in this study. As mentioned earlier, this study will investigate all inanimate classifiers that appear in three corpora of different genres. Therefore, the first concept I would use in this study is “token”. Token is an individual occurrence of a linguistic unit in speech or writing (Bybee 2006). In this study, a token or a classifier token indicates an occurrence of an actual classifier or classifier type in the dataset. It is differentiated from a type by Bybee (2006). In the current research, a classifier type means an actual classifier (type) that may occur once or many times in the corpus.

The term “language variation” or “linguistic variation” or simply “variation” used in this study refers to differences in language use across genres or genre differences. According to Tagliamonte (2012), the rate of occurrence of a linguistic form is a very important feature in understanding variation and the frequency of a feature depends on the contexts. Frequency is the count of the occurrence of a particular word in texts (Bybee 2006). In this study, the term “frequency” is used to indicate the frequency rate of occurrence of a classifier (type) per 10,000 words. On the other hand, the term “distribution” refers to the overall distribution of a classifier (type) in the dataset. This indicates how often a classifier (type) is generally distributed in the corpus compared to other classifier (types) in terms of percentages.

2.7 Summary

In summary, there are several approaches to classifiers in Vietnamese. Firstly, Emeneau (1951) and many other researchers such as D. H. Nguyen (1957), Thompson (1965), and P. P. Nguyen (2002) share the same viewpoint in the perception of Vietnamese classifiers. They primarily work on descriptive grammar and their analyses are based on the basic vocabulary, constructed or elicited utterances. Emeneau (1951:85) divides Vietnamese nouns into two major subclasses: “classified nouns” and “non-classified nouns”, and illustrates the structure of the Vietnamese noun phrase in (24) in section 2.4.3. This schema will be employed for identifying classifiers in the corpora this study will be working on. Following their approach of perceiving and identifying classifiers, this study investigates inanimate classifiers in the Vietnamese corpora.

More recently, H. T. Nguyen (2013) analyses the Vietnamese noun phrase syntactically and focuses on discussing the structure with the occurrence of an extra *cái* (inani.), which is considered as a focus marker. He perceives classifiers in the same way as do many researchers including Emeneau (1951) and D. H. Nguyen (1957), but H. T. Nguyen (2013) does not recognize classifiers showing contents as the other researchers do. He argues that they are measure words. Simpson and Ngo (2018) syntactically analyse Vietnamese classifiers within the noun classification. In their analysis of the functions of three constituents in the combination of numeral + classifier + noun, they conclude that in Vietnamese, the overt presence or absence of classifiers mainly appears to be a property of nouns, and classifiers are first combined with nouns before being built together with numerals. They discuss and argue that the special extra classifier *cái* (inani.) occurs with optional-classifier nouns with the same function as when it precedes another

classifier plus an obligatory-classifier noun. In contrast, the approach of Cao (1988) is very different from the other researchers' because he recognizes the general classifiers only as discussed in section 1.1. He has an interesting analysis on Vietnamese classifiers and nouns in comparison with English and other European languages. However, I am not following his approach.

Bisang (1999) and Löbel (2000) focus on analysing the functions of Vietnamese classifiers. Bisang claims that Vietnamese has two primary functions, classification and individualization including identification, and also anaphoric function. Löbel argue that Vietnamese classifiers also have syntactic referentialization. They discuss that classified nouns do not require classifiers in certain environments, where a noun is modifying another noun and/or the noun has no referent. H. T. Nguyen (2004) claims that Vietnamese classifiers perform the function of nominalization.

Although researchers have different views on certain points, many of them agree with Emeneau's perceptions of Vietnamese classifiers such as D. H. Nguyen (1957), Thompson (1965), P. P. Nguyen (2002), Q. B. Diep (2005), and Simpson and Ngo (2018). According to them, a classifier in Vietnamese is a word that categorizes the noun it precedes into a generalized classification and individuates the noun so that the noun can be specified and counted. Following their approach and employing the schema in (24), I will investigate inanimate classifiers in the three Vietnamese corpora, which will be described in the next chapter.

Chapter 3

Data and Methodology

In this chapter, I will describe three corpora that I will be working on for investigating inanimate classifiers in Vietnamese in the current research in section 3.1. Then criteria for identifying a classifier in Vietnamese will be identified in 3.2, with the foundations in 3.2.1 and criteria in 3.2.2, and problems 3.2.3. Section 3.3 will present methodology for the study. Data organization, data analysis, and data aggregation will be described in subsections 3.3.1, 3.3.2, and in 3.3.3 respectively. Finally, section 3.4 summarizes the main points in the chapter.

3.1 Corpora

This study investigates inanimate classifiers in Vietnamese in three corpora: Vietnamese folktales (hereafter referred to as the Vietnamese Narrative Corpus or the Narrative Corpus for short), recent online newspapers (the Vietnamese Online Newspaper Corpus or the Online Newspaper Corpus) and spoken discourse from TV talk shows (the Vietnamese Spoken Corpus or the Spoken Corpus).

3.1.1 The Vietnamese Narrative Corpus

Folk narratives are an integral part of cultural heritage which can be a valuable resource for folk narrative studies since moral values and beliefs, and identities of groups and individuals over time are reflected in folktales (Meder 2010). In addition, studying folk narratives, a treasure of literature, can illustrate how folk people use the language in traditional ways. These folktales were told by Vietnamese native speakers at least about sixty years ago, and are still read and loved by many young readers nowadays. The language in these folktales is understood to be natural for the time but traditional and archaic in comparison with current language. In choosing this corpus, I attempt to find out whether classifier use in ‘real life stories’ is different from their usage in the other two genres, recent online newspapers and oral conversations, which will be described in detail in the following sections.

The Narrative Corpus consists of one hundred forty-one Vietnamese folktales randomly selected from two books: “Truyện cổ nước Nam” (Vietnamese folktales) Volume 1 and “100

truyện cổ tích Việt Nam” (100 Vietnamese folktales). Each story in the corpus is about three to fifteen pages long. The corpus contains a total of about 115,000 words. The first book was written by Ôn Như Văn Ngọc Nguyễn in 1932, and first published in 1957. The author was born in 1890 in a village in Hai Duong province in the North of Vietnam. He was one of the first students at the College for Teachers’ Training in the early 20th century and became a teacher at a primary school in Hanoi, the capital city of Vietnam. He then changed jobs several times and travelled extensively to collect stories from many different villagers in various regions throughout the country. He was the Director of the Education Department of Ha Dong province. He, in fact, wrote many different books and had influence on many generations since the early 20th century. The stories in this book are typical and carry the significant value of the Vietnamese folktales since they are “completely folk”, which reflect the farmers’ way of speaking, as the author’s grandson, Chien Tran, wrote in the Acknowledgement of the book (Nguyen 2016:6). This book was republished in 2016 in Vietnam, and covers a wide variety of topics including animals, country, family, talented people, and festivals. For short, in this study, we call it Book 1.

The second book, ‘*100 truyện cổ tích Việt Nam*’ (100 Vietnamese folktales), was written by many different authors, among whom Huy Nguyên Lữ and Văn Lung Đặng are also the editors. This book was republished in 2013, and it is called Book 2 in this study. Particularly, apart from the very familiar folktales with Vietnamese people, this book contains special folktales of different minority ethnic peoples from every region of the country including minority ethnic groups in the Northern mountainous areas, Highlands, Mekong Delta, and the South Central region. The folktales, which are supposed to have originated many years ago, before 1954, were orally transmitted from generation to generation. They were then collected, written and edited by the authors. They describe the Vietnamese farmers’ spiritual and material life as well as their culture, work, thinking and habits, and are closely related to fields, forests, mountains, rivers, and oceans because 90% of the Vietnamese population used to be in agriculture.

In brief, the folktales reflect the use of natural narrative language of Vietnamese native speakers in the past which still has an important position in the treasure of Vietnamese literature. The Narrative Corpus provides a valuable source for traditional linguistic research, so it can demonstrate differences in classifier use in the narrative language versus online contemporary discourse, and actual conversation in the other two corpora described in sections 3.1.2 and 3.1.3.

3.1.2 The Vietnamese Online Newspaper Corpus

The second corpus is named the Vietnamese Online Newspaper Corpus or the Online Newspaper Corpus for short. This corpus consists of one hundred forty online articles. They are collected from four popular Vietnamese websites, which are <https://vnexpress.net>, <https://dantri.com.vn>, <https://vietnamnet.vn>, and <https://tuoitre.vn>. The e-articles are written by Vietnamese native speakers. The articles in this corpus cover a wide variety of topics including news, events, world, science, health, life, laws, education, culture, business, sports, tourism, and entertainment. They describe all aspects of people's life in the current society as well as news and events worldwide. These articles are written by many different people and 'e-published' mainly in 2019 and 2020, so the language is more current compared to the language in the Narrative Corpus.

The online articles are randomly selected and collected by the researcher of this study. A number of articles under every topic have been selected in each of the websites earlier mentioned. I copied all the e-articles that I collected and put them in a word document file with notes of the necessary information such as the year of publication and the source. I labelled the articles by numbering them from 1 to 140. I can count the total number of words of the corpus and have the data for analysis. The Online Newspaper Corpus has a word count of about 135,900 words.

In investigating classifiers in the Online Newspaper Corpus, I aim to explore whether classifier use in this genre is different from their use in the folktale narrative, in terms of distribution and frequency. The online newspaper articles are contemporary, and their target audience are adults. They cover a wide range of topics for communicating current news and information within the country and worldwide. However, traditionally folktales in the Narrative Corpus mainly talk about animals and are geared towards children. I attempt to find out whether there is variation in classifier use in these genres because they are so different in terms of their audiences and contexts of use. For the Narrative Corpus, the folktales, which were written years ago and used to be orally transmitted, were later published in paper-printed form. They are thus more familiar for the older generations when the majority of the Vietnamese population were farmers. However, the e-articles have been written recently within the last one or two years by younger writers. With new advanced technology, they are published electronically. These two corpora appear to belong to two different times and two generations. They are in two different forms of publication, paper-printed and e-published. I hypothesize that there is variation in

classifier use among these genres. However, these two corpora are in written discourse. With an aim to compare classifier use in written texts and spoken discourse, I will investigate classifier use in the Vietnamese Spoken Corpus, which is described in the next section.

3.1.3 The Vietnamese Spoken Corpus

The third corpus is the Vietnamese Spoken Corpus or the Spoken Corpus for short. It consists of twenty-two talk show episodes broadcasted on Vietnam Television (VTV), vtv.vn and/or reposted on youtube.com. The talk shows took place in recent years from 2011 to 2019. Each episode video clip is between thirty minutes to sixty minutes in length. The total duration of all the talk shows is approximately 14 hours. I watched and listened to the episodes, and then transcribed them myself since there are no available transcriptions of the dialogs for research. A number of talk show episodes have been previously transcribed by other native speakers. They were transcribed in 2018 for another research project by a scholar named Thu Trang Nguyen, from whom I got them through personal contact. For these, I listened to the talk show episodes again and edited the transcriptions to make sure that they are accurately transcribed. I also checked and noted all the necessary information such as the speaker's age and gender, and the year that they were published and/or broadcasted. After I had the transcriptions for all the talk show episodes, I put them in a word document file. In this way, I can count the total number of words and have the data for analysis. This corpus has a word count of approximately 151,000 words.

The Spoken Corpus is actual spontaneous speech, so the language used is likely to be natural and conversational. All the speakers in the talk shows are Vietnamese native speakers. They are all quite famous and recognized by their names and positions in Vietnam, so their age can be identified by looking them up on Vietnamese websites. There are 46 speakers altogether in the corpus. They are divided into three groups based on their age. The first group called 'older speakers' consists of 14 speakers. They are over 50 years old (those who were born in 1968 or earlier). The second group called 'middle-aged speakers' consists of 18 speakers. They are between 30 and 50 years old (those who were born in 1969 to 1987). There are 14 'younger speakers' in the third group. They are under thirty years old (those who were born in 1988 or later). The age of speakers in this corpus is contingent, but it is good to have three groups of different age. In this way, I will compare classifier use among these groups to identify similarities and differences in classifier use among them in terms of frequency. The differences in classifier use

among these groups in the Spoken Corpus, as a part of the whole study, are expected to reveal some tendency of classifier use by Vietnamese native speakers of different ages. In sum, the three corpora that the current study is working on have been described. Setting out criteria for identifying a classifier in Vietnamese in this study is important, which is specified in the next section.

3.2 Criteria for identifying a classifier in Vietnamese

There appears to be a large number of classifiers in Vietnamese although there are disagreements in the literature about what constitutes a classifier in Vietnamese and in the number of classifiers. It is sometimes difficult to determine whether a morpheme functions as a classifier, so it is important to set out criteria for identifying a classifier in the study. Before describing the method to distinguish it from other constituents, I briefly present the definition and functions of classifiers as the foundations again here. As discussed in section 1.1, Vietnamese classifiers are words that are used to individuate nouns and categorize nouns into a different classification (D. H. Nguyen 1957; Thompson 1965; P. P. Nguyen 2002, Diep 2005). Classification, individualization, and nominalization are the major functions of Vietnamese classifiers (Bisang 1999; H. T. Nguyen 2004, 2013). The structure of the Vietnamese noun phrase put forward by Emeneau (1951) in (24) is employed in this study for identifying a classifier as well.

3.2.1 Criteria for identifying inanimate classifiers

Based on the numeral classifier structure of Vietnamese in (24) claimed by Emeneau (1951), in the presence of a numeral, the morpheme or word between the numeral and head noun can be a classifier. In the event that the numeral does not occur, the morpheme preceding the head noun can be considered as a classifier. However, based on the functions of classifiers in Vietnamese, the morpheme can be identified as a classifier only if it carries the lexical semantic function of classifying and individuating and/or nominalizing the head noun. That is, a morpheme which can be identified as a classifier must satisfy the criteria regarding structure and lexical semantic function of a classifier in Vietnamese. For instance, different constituents in (48a) are analysed as follows: the numeral ‘một’ (one), the head noun ‘phim’ (movie), and *bộ* (set) positioned between them. The morpheme *bộ* (set) can be identified as a classifier because it individuates the noun ‘phim’ (movie). Similarly, in (48b), with the numeral ‘3500’ and the head noun ‘sông’ (river), the morpheme ‘con’ between them is identified as a classifier, individuating

the noun. In (48c), positioned between the numeral ‘mấy’ (several) and the head noun ‘thuốc’ (medicine), the morpheme *viên* (pill) is identified as a classifier, which individuates the head noun. In (48d), the morpheme ‘sự’ positioned between the numeral ‘nhiều’ (much) and the head ‘hỗ trợ’ (support) is identified as a classifier because it not only nominalizes but also individuates the head ‘hỗ trợ’ (support). Following the same process of analysing constituents in the NPs, the morpheme positioned between the numeral and the head noun can then be identified as in (48e-g).

- (48) a. Trong một bộ phim tài liệu
 In one CL(set) movie documentary
 ‘in a documentary movie’ (O5.214)
- b. với hơn 3500 con sông có chiều dài lớn hơn 10 cây số
 with over 3500 CL(ani.) river have length big than 10 km
 ‘with over 3500 rivers of more than 10 km long’ (O53.5099)
- c. mới mua được mấy viên thuốc.
 just buy get several CL(pill) medicine
 ‘just got several pills.’ (O31.2325)
- d. nhận được nhiều sự hỗ trợ của mọi người đến thế.
 receive get much CL(action) support of every human such that
 ‘received a lot of support from everyone like that.’ (O69.6743)
- e. khiến không ít vụ án tham nhũng bị kéo dài, bế tắc.
 cause not little CL case corruption PASS last long stuck
 ‘caused many corruption cases last long and get stuck.’ (O60.6054)
- f. trong mỗi chuyến du lịch
 in every CL(trip) travel
 ‘in every tour’ (O676562)
- g. Các cuộc điều tra đang tiếp tục
 PL CL investigation PROG continue
 ‘The investigations are continuing’ (O17.992)³

³ Unless otherwise indicated, the examples given starting from (48) and later are from the corpus of the current study. They will be coded by the abbreviations of the corpus name, article number for the VONC or talk show episode for the VSC, and token number. For instance, (48g) is coded as (O17.992), meaning that it comes from the VONC, article number 17, and token number 992.

For cases in which the numeral does not appear in the NP, once the head noun is identified, any morpheme or word preceding the noun might be considered as a classifier if it serves the function of a classifier. For instance, in (49a) there are two NPs ‘*chiếc* nón’ (bamboo hat) and ‘*món* quà của bạn’ (your present). For the first NP, in the absence of a numeral, the morpheme *chiếc* (individual) preceding the head noun ‘nón’ (hat) is identified as a classifier because it individuates the noun. For the second NP, in the absence of a numeral, the morpheme ‘*món*’ preceding the noun ‘quà’ (present) is identified as a classifier as it classifies and individuates the noun. The noun is also followed and modified by the possessive ‘của bạn’ (your). Similarly, in (49b), in the NP ‘*căn* nhà này’, without a numeral, the morpheme ‘*căn*’ precedes the noun ‘nhà’ (house), followed by the demonstrative này (this). This morpheme is then identified as a classifier which classifies and individuates the noun. In (49c), in the absence of a numeral, the morpheme ‘*sự*’ preceding the head ‘độc lập’ is identified as a classifier because it not only nominalizes but also individuates the head.

- (49) a. Đây là *chiếc* nón, là *món* quà của bạn.
 Here be CL(individual) hat be CL present of you
 ‘This is a hat, your present.’ (S17.663)
- b. *Căn* nhà này ban đầu mang phong cách tân cổ điển,
 CL house this initially have style new classic
 ‘This house initially had the ‘new classic’ style,’ (O73.7103)
- c. khuyến khích *sự* độc lập của con,
 encourage CL(state) independent of him
 ‘encourage his independence,’ (O6.278).

For cases when the head noun is identified in the context, the classifier can occur with numerals and/or demonstratives with the omission of the head noun as in (50). In this case, the numeral is first identified as in (50a) and (50c), then in order to see what noun the morpheme following the numeral refers to, we have to look back to the preceding context. The morpheme *chiếc* (individual) in (50a) refers to the noun ‘car’ and ‘*bông*’ in (50c) refers to the noun ‘flower’ identified in the context. These two morphemes are then identified as classifiers for the nouns ‘car’ and ‘flower’ because they carry their functions of classifying and individualizing the nouns. In (50c), the classifier *bông* (flower) precedes the demonstrative ‘này’ (this) which makes the NP become definite. In (50b), without the numeral and head noun, *chiếc* (individual) precedes a

modification ‘còn lại tại hiện trường’ (remaining at the scene) which makes the noun definite and helps readers identify it. In fact, the head noun is omitted in this case, so we have to refer to the preceding context and find that *chiếc* (individual) is used as a classifier for the noun ‘shoe’ previously mentioned in the context. When the head noun is identified in the context, the classifier can occur with numerals and/or demonstratives or a defining modifier of the noun.

- (50) a. suy giảm gần 400 *chiếc*
 decrease nearly 400 CL(individual)
 ‘decreased by nearly 400 cars’ (O71.6828)
- b. giống *chiếc* còn lại tại hiện trường
 alike CL(individual) remaining at scene
 ‘the same as the remaining one at the scene’ (O10.531)
- c. nếu chỉ một *bông* này mà nó héo
 if only one CL(flower) this which it dry
 ‘if only this one dries up’ (S2.5245).

Many non-classified NPs in Vietnamese do not occur with a classifier. For instance, in the NP ‘một công trường xây dựng’ (a construction site) as in (51a), after the numeral ‘một’ (one), the word ‘công trường’ (site) precedes another word ‘xây dựng’ (construction, building). In this case, in the slot of a classifier, ‘công trường’ (site) is the head noun, and the word following it is just its modifier. It means this NP which consists of a head noun followed by another modifying noun is non-classified. In other cases, NPs consists of only a numeral with a non-classified noun as in (51b), the numeral ‘rất nhiều’ (a lot of) precedes the noun ‘khó khăn’ (difficulty).

- (51) a. Hà Nội như một công trường xây dựng,
 Hanoi like one site construction
 ‘Hanoi like a construction site,’ (O44.3963)
- b. Trải qua rất nhiều khó khăn
 experience very a lot of difficulty
 ‘experienced a lot of difficulties’ (S3.6597).

In most cases, when analysing constituents in the NPs, considering their position and function is an effective way to identify an inanimate classifier in Vietnamese. In sum, for

identifying a classifier in Vietnamese, it is the position and the function that need to be considered. These criteria are put in Table 1. Only the constituents that satisfy both criterion 1 and 2 can be considered a classifier in Vietnamese.

Table 1: *Criteria for identifying a classifier in Vietnamese*

Criteria		Classifiers	Non-classifiers
1. Position	Following a numeral (if present)	+	+
	Preceding a head N	+	-
	Preceding a demonstrative or a defining modifier	+	+
2. Lexical semantic function	Classifying the head N	+	-
	Individualizing the head N	+	-
	Nominalizing the head	+	-

Table 1 shows clearly the criteria for identifying a classifier in Vietnamese. However, in some cases, it is more difficult to determine whether the morpheme preceding the head noun is a classifier or not because that morpheme/word carries its lexical meaning and can stand independently as a content word in other cases. The next section will discuss some difficulties when analysing constituents in the NPs in identifying an inanimate classifier in Vietnamese.

3.2.2 *Difficulties in identifying classifiers in Vietnamese*

In Vietnamese, we cannot depend on the word class to identify the head noun or any other constituents because there are no markers or form of words to help identify the part of speech as in English (H. T. Nguyen 2013). Thus, we have to rely on the context and semantics of the NP and/or of the clause or sentence to identify the head noun. One difficulty when analysing constituents in NPs and considering whether a morpheme/word is a classifier or not is with *việc* (job, activity). This is a free morpheme and can be a content word. It means it can be a classifier in some cases, but a noun in other cases. For instance, in the NP *việc tăng giá cả* (increase of prices) as in (52a), the morpheme *việc* (job, activity) precedes the head *tăng* (increase) without a numeral while the word *giá cả* (prices) follows and modifies the head. In this case, *việc* (job, activity), which nominalizes and classifies the head ‘tăng’ (increase), is identified as a classifier. Similarly,

in the NP *việc cách ly* (isolation) as in (52b), preceding the head *cách ly* (isolate) without a numeral, *việc* (job, activity), which classifies the head *cách ly* (isolate), is identified as a classifier. In contrast, in other cases, *việc* (job, activity) can be a noun with the lexical meaning of ‘job’ or ‘work’, and it may combine with another morpheme ‘*làm*’ (do/work) to become a compound noun *việc làm* (job). For example, structurally the NP *việc ổn định* (stable job) as in (52c) looks exactly like the pattern CL - N. However, *việc* (job) is the head noun while *ổn định* (stable) is an adjective, functioning as a modifier for the noun and adding the quality ‘stable’ to the noun *việc* (job). If *việc* (job, activity) is treated as a classifier in this case, the NP does not make sense as in (52d).

- (52) a. sẽ kiềm chế trong *việc* tăng giá cả
 will restrain in CL(activity) increase price
 ‘will restrain the increase of prices’ (O16.877)
- b. thì *việc* cách ly đặc biệt quan trọng.
 then CL(activity) isolate special important
 ‘then isolation is especially important.’ (O31.2377)
- c. khi đã có *việc* ổn định
 when PAST have job stable
 when (they) have stable jobs’ (O39.3387)
- d. *khi đã có *việc* ổn định
 when PAST have CL stable/stabilize
 ‘when (they) have stabilizing’
- e. có *việc* làm ổn định,
 have job stable
 ‘have a stable job,’ (O124.9587)
- f. *có *việc* làm ổn định,
 have CL do stable
 ‘have stabilizing,’ ⁴

Similarly, in the NP *việc làm ổn định* (stable job) as in (52e), the noun *việc làm* (job) has the same meaning as the noun *việc* (job) while *ổn định* (stable) is an adjective, modifying the head

⁴ The examples (52d) and (52f), which are given for illustration purposes only, are not from the corpus.

noun and adding the quality ‘stable’ to the head noun *việc* (job) as in (52c). It does not make sense if *việc làm* (job) is treated as a classifier in this case as in (52f).

For the words that are considered “proper classifiers” by Nguyen (1957) at all times such as *cái* (inanimate), *chiếc* (individual), *con* (animate), *bộ* (set), *cây* (tree, long object), *bông* (flower), *bức* (picture, mail), *cuộc* (match, strike), *cuốn* (book, volume), *quả* (fruit, round object), *quyển* (book), *chuyến* (trip), *tấm* (picture, thin), *tờ* (sheet), and *viên* (pill), it is quite easy to identify them. However, for words or morphemes that can be classifiers in some cases and lexical words in other cases, it is difficult to identify them. The word *việc* (job, activity) that is analysed in (52) is an example. It is worth noting that another morpheme that can be a classifier in some cases but can be part of a compound noun in other cases as shown in (53). The morpheme *sự* (event) preceding the head *khác biệt* (different) in the absence of a numeral as in (53a) is identified as a classifier. This classifier not only classifies but also nominalizes the head adjective and turns it into the noun *sự khác biệt* (difference) in this case. However, in other cases, this morpheme combines with another morpheme to become a noun such as *sự cố* (incident) as in (53b)⁵. This can be a non-classified noun as in (53b) or a classified noun preceded by *cái* (inanimate) as in (53c). In other words, we can call this noun an optional-classifier noun as Simpson and Ngo (2018) do. Clearly, in this case, the morpheme ‘*sự*’ is not a classifier, but it is simply a part of a bi-syllable noun. It is similar for other cases in which this morpheme is not a classifier, but just a part of a noun such as *sự kiện* (event) as in (53d-e) and *sự nghiệp* (career) as in (53f-g). These nouns can be non-classified as in (52d) and (52f) or classified nouns which occur with *cái* (inanimate) as in (53e) and (53g). Similarly, these can be called optional-classifier nouns as Simpson and Ngo (2018) do.

- (53) a. *Sự khác biệt của Việt Nam*
 CL(event) different of Vietnam
 ‘The difference of Vietnam’ (O26.1765)
- b. *sau sự cố Formosa*
 after incident Formosa
 ‘after the incident of Formosa’ (O114.9098)

⁵ It should be noted that we cannot separate these two morphemes because it does not make sense if these two morphemes of the noun *sự cố* (incident) are taken apart. It is similar for the case of the noun *sự kiện* (event) and *sự nghiệp* (career).

- c. có những *cái* sự cố, nhiều lắm
 have PL CL(inani.) incident many so
 ‘(we) have incidents, so many’ (S11.2055)
- d. tham gia rất nhiều sự kiện khác nhau.
 participate very many event different
 ‘participated in a lot of different events.’ (S13.3296)
- e. mỗi một *cái* sự kiện xảy ra trong *cuộc* sống của mình
 each one CL(inani.) event happen in CL live of self
 ‘each of the events happened in our life’ (S13.3295)
- f. trong sự nghiệp của thầy giáo Ngô Mạnh Cường,
 in career of teacher Ngo Manh Cuong
 ‘in the career of teacher Ngo Manh Cuong,’ (S7.8051)
- g. *cái người* phụ nữ sẽ hy sinh *cái* sự nghiệp của mình
 CL(inani.) CL(human) woman will sacrifice CL(inani.) career of self
 ‘the woman will sacrifice her own career’ (S22.999).

Another morpheme that may easily cause difficulty when determining if it is a classifier or not is *điều* (thing). This is a free morpheme and has its own lexical meaning, so in many cases it is identified as the head noun as in (54a-b). In the NP *rất nhiều điều bất ngờ nữa* (a lot more surprising things) in (54a), preceded by the quantifier *rất nhiều* (very many, a lot of), *điều* (thing) is identified as the head noun, with the attribute *bất ngờ* (surprising) modifying the head noun and the particle *nữa* (more). Similar for the NP in (54b), as a Vietnamese noun can be post-modified by adjectives and/or a demonstrative and/or a possessive, *điều* (thing) is the noun post-modified by the adjective *khó khăn* (difficult) in (54b). However, if we just separate different constituents in the NP *rất nhiều - điều - bất ngờ* (a lot of - thing - surprising), it looks like the NP following the pattern Numeral - CL - N. In this assumptive case, *điều* (thing) would be a classifier and *bất ngờ* (surprising) would be a noun. It does not make sense because *điều* (thing) occurs independently as a noun without any post-modifiers as in (54c-d). In contrast, this word can be a classifier in other cases as in (54e-f). It classifies the head *ước* (wish) in the NP *một điều ước* (a wish) as in (54e) and the head noun *luật* (law) in the NP *điều luật* (law article) as in (54f). In these cases, *điều* (wish, law) is a classifier because it classifies and individuates the nouns. This classifier is also recognized in such cases by prior researchers including Emeneau (1951), D. H. Nguyen (1957), and Thompson (1965).

- (54) a. sẽ còn có rất nhiều điều bất ngờ nữa
will still have very many thing surprising more
'there will be more surprising things' (S5.7211)
- b. phải trải qua những điều khó khăn như vậy.
have to experience PL thing difficult such
'had to experience such difficult things.' (O63.6231)
- c. để đạt được những điều anh mong muốn.
for obtain get PL thing you desire
'to obtain the things/what you desire for.' (S11.1784)
- d. Nhiều người sẽ không tin điều này
many human will not believe thing this
'Many people do not believe this thing' (O58.5819)
- e. chỉ có một *điều* ước thật giản dị
just have one CL wish really simple
'(I) just have a really simple wish' (S7.8113)
- f. Không có *điều* luật nào
not have CL(article) law any
'Not any law articles' (O31.2333).

Furthermore, it is difficult to identify a mensural classifier or a classifier showing contents. Following the approach that D. H. Nguyen (1957) and other researchers distinguish mensural classifiers from nouns presented in the literature, I am very cautious when identifying this kind of classifiers. For instance, *vườn* (garden) is a non-classified noun as in (55a) in the absence of a numeral and a classifier, whereas it is a classified noun preceded by *cái* (inanimate) as in (55b). Similarly, this noun is classified by the classifier *mảnh* (piece) as in (55c). On the contrary, *vườn* (garden) becomes a mensural classifier for the noun *cây cảnh bonsai* (bonsai plant) in the NP *một vườn cây cảnh bonsai* (a garden of bonsai plants) as in (55d). When analysing constituents in this NP, it appears that the NP has two nominal components: *vườn* (garden) and *cây cảnh bonsai* (bonsai plant). However, the noun *cây cảnh bonsai* (bonsai plant) is the actual direct object of the verb *trồng* (grow) while *vườn* (garden) in this case is used as a mensural classifier for the noun 'bonsai plant'. It is similar for *thảm* (carpet) and *hoa* (flower) as in (55e). Both of them are originally nouns, but in this case, *hoa* (flower) is the head noun with its own reference functioning

as the direct object of the verb *chiêm ngưỡng* (gaze) while *thảm* (carpet) is used to describe yellow flowers as a ‘carpet of flowers’. Clearly, *thảm* (carpet) does not have its own reference. This means, *thảm* (carpet) is a mensural classifier for the noun *hoa* (flowers).

- (55) a. Vườn nhà tôi có cả nghìn cây các loại
garden house me have all thousand tree PL type
‘My garden has thousands of trees of all types’ (O89.7726)
- b. Quét thế cho cái vườn nó ấm.
paint so for CL(inani.) garden it warm
‘Painting them like that makes the garden warm.’ (O34.2765)
- c. từ mảnh vườn tổ tiên xứ Bắc
from CL(piece) garden ancestor region North
‘from the piece of garden of the ancestor in the North’ (O38.3267)
- d. không chỉ đầu tư trồng một vườn cây cảnh bon sai tiền tỉ
not only invest grow one CL(garden) plant bonsai money billion
‘not only invested in growing a garden of bonsai plants worth billions of VND’ (O89.7715)
- e. lượng du khách đến chiêm ngưỡng thảm hoa vàng sụt giảm đáng kể.
quantity visitor come gaze CL(carpet) flower yellow decrease considerably
‘the number of visitors coming to gaze the carpet of yellow flowers decreases considerably.’ (O116.9206).

I have analysed and made some comments on a number of cases which can easily cause problems when determining whether a morpheme is a classifier or not. The issues that may cause difficulties in identifying a classifier in these cases can be listed below.

- Classifiers that can be words with lexical meaning (content words) in other cases.
- Classifiers that can be a part of the multi-syllabled words in other cases.
- Classifiers that can be nouns modified by attributes in other cases.
- Mensural classifiers or classifiers showing content which can be nouns in other cases.

In these cases, it is important to carefully employ the criteria in analysing constituents in the NPs and identifying classifiers to avoid mistakes. The next section will present the methodology that I apply for this study.

3.3 Methodology

As mentioned above, this is a corpus-based study and the three corpora that have been described in section 3.1 are used as the data for this research. In this section, the organization of the data is described in section 3.3.1. The methodology for analysing the data is discussed in section 3.3.2. How the data in the three corpora are aggregated will be presented in section 3.3.3. The summary of the corpora and methodology used in this study is in section 3.3.4.

3.3.1 Organizing the data

As described in section 3.1, I collected all the texts for the Narrative Corpus as well as the Online Newspaper Corpus and did the transcriptions for the Spoken Corpus. I keep the texts for the Narrative Corpus in printed copies while the texts and transcriptions for the Online Newspaper and Spoken Corpus are kept in word document files. After identifying all inanimate NPs in the three corpora by highlighting them in printed copies or in word files, I extracted and typed or copied all the phrases or clauses containing these NPs and put them into a column in an Excel spreadsheet (Microsoft Excel 2016) under the heading ‘context’. The ‘context’ shows the linguistic environment where the classifier and/or noun appears. The data of the three corpora are kept separately in three spreadsheets for easy tracking and reference.

For every NP in the spreadsheet, I noted down all the relevant information I need such as the source of texts for the corresponding corpus. For the data in the Narrative Corpus, I coded all the NPs by labelling them as N (standing for Narrative Corpus), followed by 1 or 2 as it appears in the Book 1 or 2 accordingly, then the page number where the NP appears. For example, N1.39 means the NP is from the Narrative Corpus, Book 1, page 39. For the data in the Online Newspaper Corpus, the NP is coded as O (standing for the Online Newspaper Corpus), followed by the article number, then the token number. For the data in the Spoken Corpus, I noted down the talk show episode by numbering them as TS1 to TS22 in a different column in the spreadsheet. I also put in the speakers’ age and gender for the Spoken Corpus. I coded all the tokens in the Spoken Corpus by labelling them with S (standing for the Spoken Corpus), followed by the talk show episode number and then the token number. In this way, I have all the data based on the three corpora in the spreadsheets with distinctive and necessary information I need for analysis.

3.3.2 *Analysing the data*

Based on the criteria to identify a classifier in Vietnamese as presented in section 3.2, I follow a number of steps below to analyse the data.

3.3.2.1 **Identifying inanimate classifiers and head nouns**

I examined all inanimate NPs in the extracted phrases or clauses in the spreadsheet in chronological order within each corpus. I analysed each of the NPs into different constituents including numeral, classifier, noun if they are present. After that, I place classifiers and inanimate nouns into two other columns in the spreadsheet under their own headings of ‘classifier’ and ‘inanimate noun’. In the NP if the numeral occurs, P (standing for present) is put in another column under the heading ‘presence or absence of numeral’. If the numeral does not occur, A (standing for absent) is put in this column accordingly. The presence of the numeral in the NP is counted for the purpose of identifying the classifier construction. In fact, it would reveal whether the typical classifier construction in Vietnamese would potentially be Numeral - CL - N or CL - N.

For NPs in which a classifier is present, I then determined whether the classifier is single or double, and I noted it down as single or double accordingly in a separate column under the heading ‘classifier type’. For NPs in which a classifier is absent, I put ‘null’ in the column in the spreadsheet. The next step is to identify classifiers, I found out what lexical semantic function the classifier has and put it in a column under the heading ‘lexical semantic function of CL’. For this, I attempt to find out what function classifiers in Vietnamese have.

Identifying all inanimate noun phrases and classifiers that appear in the three corpora of this study is a very important and time-consuming step in the process of analysing the data. After this stage, I analysed every inanimate classifier and noun under a set of factors, which are described in section 3.2.2.2.

The examples to illustrate how classifiers and head nouns were analysed and organized in the study are given in Table 2a.⁶

⁶ It is noted that due to the paper size, Table 2a is half of the spreadsheet that is used in Microsoft Excel for the study. The other half for other factors in the spreadsheet is given in 3.3.2.2.

Table 2a: Examples of data analysis

Extracted phrases/ clauses	CL	Inani. NP	ling variable 1	ling variable 2	ling variable 3	ling variable 4
			Noun referent: generic/ specific	Classifier type	Lexical semantic functions of CLs	presence/ absence of numerals (P or A)
Tao lại thấy cái chum ở bờ ruộng	cái	chum	specific	single	Ind	A
T đánh rơi một chiếc giày.	chiếc	giày	specific	single	Ind	P
Khi lên trên chỏm một quả núi	quả	núi	specific	single	Ind	P
để chống sự lan rộng của dịch bệnh này	sự	lan rộng	specific	single	Nom	A
mà đây là cái sự quan tâm đích thực, đúng k?	cái sự	quan tâm	specific	double	Emphasis- Nom	A
yêu thương cái con rầy mỡ màu bắp, lúa	cái con	rầy	specific	double	Emphasis- Ind	A
Để giải quyết được vấn đề này		vấn đề	specific	null		A

3.3.2.2 Analysing nouns and classifiers under some factors/variables

I analysed every noun with or without classifiers under a number of factors concerning the noun and classifier. Specifically, I first determined whether the noun has a generic or specific referent. I attempt to find out whether the noun's referent is related to the use of classifiers or not because from my own observations, nouns with generic reference usually do not require classifiers. I noted down this information in a separate column in the spreadsheet under the heading 'noun referent' by putting 'generic' or 'specific' correspondingly.

Second, the definiteness of the noun is examined in order to see what decides the definiteness of the noun and whether the use of classifiers determines it. This information is noted

down in a separate column under the heading ‘definiteness’ in the spreadsheet by putting Y (yes) for a definite noun or N (no) for an indefinite noun accordingly. Because the previous mention of the noun in the context is one factor that influences the definiteness of the noun, it is also examined in the study. This is to find out whether ‘previous mention’ determines the definiteness of the noun or other factors do. Thus, I made a note of whether the noun is previously mentioned in the context in another column under the heading ‘previous mention’ in the spreadsheet by writing down Y (yes) for a noun previously mentioned or N (no) for a noun not previously mentioned appropriately. The examples are illustrated in Table 2b.⁷

Table 2b: Examples of data analysis

ling variable 5	ling variable 6	ling variable 7	Genre	Speaker gender	Speaker age (for VSC)	Year of publicat ion	Source
Definiteness Y/N	Previous mention in discourse Y/N	Kind of nouns: Concrete or Abstract	Spoken (S)/ Narrative (Nar)/ Online (O)	Male (M) / Female (F)	YOB		Coding of tokens
Y	Y	Con	Nar	M		2013	N2.206
N	N	Con	Nar	M		2013	N2.161
N	N	Con	Nar	M		2016	N1.148
N	N	Abs	O	M		2020	O31.2291
Y	N	Abs	S	M	1966	2014	S16.233
Y	N	Con	Nar	M		2013	N2.282
Y	N	Abs	S	F	1968	2012	S10.1645

Third, because the kind of nouns might influence the choice of classifier use, it is also investigated in the study. This is specifically to see whether a certain classifier can be used with a concrete or abstract noun only or with both kinds of nouns. I, therefore, noted down the kind of

⁷ Table 2b is the second half of the spreadsheet attached to the first half given in Table 2a in 3.3.2.1.

noun by entering ‘concrete’ or ‘abstract’ into a distinctive column under the heading ‘kind of noun’ in the spreadsheet accordingly.

Apart from these linguistic variables, the study also considers some demographic variables, such as genres, and speakers’ age for the Spoken Corpus. This is to examine how classifiers are used across genres and among younger, middle-aged, and older speakers. This attempts to identify whether there is any variation in classifier use in the three genres and among speaker groups of different ages in the spoken corpus. The information of the genre and speakers’ age was thus noted in distinctive columns in the spreadsheet. In this study, I only analyze data in the spoken corpus within the variationist framework since the age of the speakers in the talk shows can be identified via social media. This is because of the fact that they are all well-known people in Vietnam. Meanwhile, the age of all the writers of the folktales and newspapers cannot be identified. That is the rationale for not analyzing data within the variationist framework for both narrative and online newspaper corpora in the study.

In sum, after analysing all the NPs in the three corpora, I have all the data I need for the study. Then I continued with data aggregation as described in the next section.

3.3.3 Data aggregation for analysis

After analysing all the inanimate classifier tokens and NPs in the three corpora, I have all the data I need for the study. I aggregated the data in each of the three corpora into tables and calculated the frequency as well as overall distribution of each classifier. I then generalized and put them together and compared similarities and differences in the use of classifiers in the three corpora. The use of classifiers in spoken and written corpora was also compared and discussed. The purpose is to find out if there is any variation in classifier use across three genres as well as in written versus spoken discourse in Vietnamese.

Next, I sorted out classifiers used by groups of younger, middle-aged, and older speakers in the Spoken Corpus as mentioned in section 3.1. This is to compare classifier use among three age groups of speakers with an aim to identify differences in classifier use in terms of frequency. The group of younger speakers are under 30 years old (those who were born in 1988 or later). The middle-aged group are between 30 and 50 years old (those who were born in 1969 to 1987). Older

speakers are more than 50 years old (those who were born in 1968 or earlier). The result may be used to predict the tendency of using classifiers among groups of speakers of different ages.

Furthermore, the co-occurrence of two classifiers are treated as double classifiers, which is also the focus of this study. This is for comparison and further analysis on their lexical semantic functions since they are a special phenomenon and a significant part of the Vietnamese classifier system. Also, I separated and calculated the number of classifiers in the presence or absence of numerals. This is for comparison and discussion on the typical Vietnamese classifier construction.

Finally, in this research, special cases in which classifiers occur with unclassified nouns and optional-classifier nouns will be discussed. This is an attempt to examine whether there are any cases in which classifiers can appear with unclassified nouns in Vietnamese. It is hypothesized that a number of unclassified nouns can appear with *cái* (inanimate), and the choice of using this classifier depends on speakers' intention for a certain purpose.

3.4 Summary

In sum, the three corpora which I collected and used for the data in the current study, criteria for identifying an inanimate classifier in Vietnamese, and the methodology employed for this research have been described. With three corpora of different genres and means of media, the study attempts to capture variation in inanimate classifier use in Vietnamese across genres as well as in written and spoken discourse.

Based on the findings in previous research, two sets of primary criteria for identifying a classifier in Vietnamese have been discussed. These criteria are based on the position or structure and lexical semantic functions of classifiers in Vietnamese as listed in Table 1. Employing these criteria, it is not very difficult to determine a classifier in Vietnamese generally, except mensural classifiers because they can be nouns in other cases. Moreover, we cannot rely on word class to identify the head noun because there is no markers or form of words to indicate parts of speech in Vietnamese (H. T. Nguyen 2013). We have to depend on the context and semantics of discourse to identify the head noun or the classifier if it is present in the NP.

Employing these criteria, I examined all the inanimate NPs in the three corpora that I collected. Then inanimate classifiers that appear in the three corpora were analysed. Thus, I had all the data I need for the study. I did data aggregation and tabulated statistics for analysis and

comparison in the next chapters. Frequency and overall distribution of inanimate classifiers in the three corpora will be reported and analysed in chapter 4.

Chapter 4

Frequency and distribution of inanimate classifiers

In this chapter, the major findings of classifier use in the three corpora of this study in terms of frequency and distribution are reported. The first section begins with the findings regarding variation across the three genres. Then the following sections provide the findings and analysis of classifier use in each of the corpora separately for a clear distinctive picture, which makes up an overall picture of how inanimate classifiers are used in the three different genres in Vietnamese. Section 4.2 reports the major findings from the Narrative Corpus. The findings from the Online Newspaper Corpus are presented in section 4.3, and the results from the Spoken Corpus in section 4.4. Section 4.5 is the summary of the chapter.

4.1 Major findings of classifier use in the study

4.1.1 Frequency of classifier use in the three corpora

After having analysed 24351 inanimate noun phrases that appear in the three corpora of the study, a total of 8626 classifier tokens have been identified. To be specific, different numbers of tokens have been found in the three corpora, with 1828 in the Narrative Corpus, 2472 in the Online Newspaper Corpus, and 4326 in the Spoken Corpus. Because of the different word counts of these corpora, the comparison of the totals of tokens among them is not reliable. Thus, they will not be compared. However, frequency of classifier use in the three corpora is more important for comparison. In this study, the frequency of classifiers used in each of the corpora means the rate of classifier tokens per 10,000 words. The results of the study show variations in the frequencies of classifier use across the three genres. The frequency of classifier use in the Spoken Corpus is highest at the rate of 286 per 10,000 words while it is lowest at 160 per 10,000 words in the Narrative Corpus. The frequency of classifier use in the Online Newspaper Corpus is 182 per 10,000 words, a little higher than that of the Narrative Corpus. Furthermore, it is interesting that the differences in frequencies of classifier use among three age groups of speakers in the Spoken Corpus have been found. Specifically, the frequency rates of classifier use among older, middle-aged, and younger speakers are 304, 291, 269 tokens per 10,000 words respectively. These findings will be discussed in chapter 5.

Additionally, the differences in the overall distribution of classifiers in the three corpora have been identified. In this study, the distribution of classifiers means the percentage of a classifier type in comparison with all the other classifier types in the corpus. A noticeable difference is that a greater variety of double classifiers occur in the Spoken Corpus while a limited number of them appear in the Narrative and Online Newspaper Corpora. It is important to note that the overall distribution of classifiers greatly differs across the three genres. The next section briefly presents an overview of the most frequent inanimate classifiers found in the three corpora.

4.1.2 Overall distribution of the most frequent classifiers in the three corpora

The overall distribution of nine most frequent inanimate classifiers in the three corpora with the number of tokens is shown in Table 3. Each of them accounts for over 2% of all the tokens that appear in each of the corpora. The other remaining classifiers which occur much less often in the corpora are put into the ‘others’ category in this table.

Table 3: Overall distribution of frequent CLs in the three corpora

Narrative Corpus (115,000 words)		Online Newspaper Corpus (135,900 words)		Spoken Corpus (151,000 words)	
CLs	No. of tokens	CLs	No. of tokens	CLs	No. of tokens
<i>cái</i> (inanimate)	404	<i>sự</i> (event)	277	<i>cái</i> (inanimate)	2658
<i>cây</i> (tree, long)	180	<i>cuộc</i> (life, strike)	187	<i>bài</i> (song, text, lesson)	204
<i>quả</i> (fruit, round)	66	<i>cái</i> (inanimate)	180	<i>cuộc</i> (strike, life)	201
<i>chiếc</i> (individual)	59	<i>chiếc</i> (individual)	144	<i>cái sự</i> (inani., event)	144
<i>bờ</i> (bank, shore)	46	<i>việc</i> (activity)	116	<i>sự</i> (event)	129
<i>gốc</i> (root)	45	<i>bộ</i> (set)	68	<i>chiếc</i> (individual)	65
<i>thứ</i> (type)	40	<i>vụ</i> (catastrophe)	61	<i>tình</i> (relationship)	61
<i>hòn</i> (round)	38	<i>con</i> (animate)	58	<i>con</i> (animate)	58
<i>con</i> (animate)	36	<i>dòng</i> (river, line)	57	<i>đám</i> (procession, patch)	47
‘others’ (<2% each)	913	‘others’ (<2% each)	1324	‘others’ (<2% each)	759
Total	1828	Total	2472	Total	4326

As shown in Table 3, the three corpora have a number of most frequent classifiers in common and some different classifiers. The use of these classifiers in each of the corpora will be reported and analysed in more detail in the coming sections. The findings reveal variation in the number of classifier types used in these genres, with the largest number of 192 in the narrative, the lowest number of 134 in the spoken, and 153 in the online newspaper. To make it clear, we should note that double classifiers found in this study are considered and counted as different from single classifiers that use the same morphemes. This is because double classifiers are treated as different classifiers as they naturally occur in a ‘fused’ construction in speech or writing. The double classifiers found in the corpora are composed of *cái* (inanimate) and a specific classifier. Most of the specific classifiers in these constructions appear as single classifiers in the corpora. As this is a special phenomenon, they will be investigated and discussed in section 5.2.3 in chapter 5.

The list of all the inanimate classifier types identified in the three corpora is in Appendix D, in which overlapped classifiers are also distinguished and contrasted. As mentioned earlier, the findings regarding frequency and distribution of classifiers in the three corpora will be reported and analysed in the following sections. The next section presents with the major findings from the Narrative Corpus.

4.2 Findings from the Vietnamese Narrative Corpus

4.2.1 Frequency of classifiers in the Narrative Corpus

After analyzing 5377 inanimate NPs in the Narrative Corpus of 115,000 words, 1828 tokens have been identified. The overall frequency of inanimate classifiers in the corpus is 160 per 10,000 words although the frequency of different classifiers greatly varies. Altogether, 192 actual inanimate classifier types found in the corpus including 189 single and three double classifiers. The frequency rates of the nine most frequent classifiers in the corpus are shown in Table 4. All the other 183 infrequent classifiers, which are in the ‘others’ category in the table, are not discussed in detail. Although the ‘others’ category includes 913 tokens altogether, each of the classifiers in this category appears only one to 22 times. As shown in Table 4, the most frequent classifier *cái* (inanimate) occurs at a rate of over 35 per 10,000 words in the corpus. The second most frequent classifier *cây* (tree, long object) appears at a rate of approximately 16 per 10,000 words, followed by *quả* (fruit, round object) and *chiếc* (individual) at about 5 per 10,000 words. Next, *bờ* (bank,

shore, fence) and *gốc* (root) occur at a rate of 4 per 10,000 words. The other three classifiers including *thứ* (type, kind), *hòn* (round object, stone), and *con* (animate)⁸ have the same frequency rate of 3 per 10,000 words in the corpus.

Table 4: Frequency of classifiers in the Narrative Corpus

CLs	No. of occurrences	Frequency (No. of CLs per 10,000 words)
<i>cái</i> (inanimate)	404	35.46
<i>cây</i> (tree, long object)	180	15.80
<i>quả</i> (fruit, round object)	67	5.88
<i>chiếc</i> (individual)	59	5.18
<i>bờ</i> (bank, shore, fence)	46	4.04
<i>gốc</i> (root, foot)	45	3.95
<i>thứ</i> (kind, sort)	40	3.51
<i>hòn</i> (round object, stone)	38	3.34
<i>con</i> (animate)	36	3.16
‘others’	913	80.14
Overall	1828	160.46

These nine most frequent inanimate classifiers are illustrated in the examples in (1)-(9). As a general classifier, *cái* (inanimate) classifies a great variety of nouns. In (1), when combining with the noun *cầu* (bridge), *cái* (inanimate) simply classifies and individuates it. The classifier *cây* (tree, long) appears with the noun *kiếm* (sword) as in (2). In addition to individuating the noun, it also indicates that the ‘sword’ is a long object. The noun *bầu* (gourd) is classified as a type of fruit by *quả* (fruit, round) as in (3). While *chiếc* (individual) is used to individuate the noun *thuyền* (boat) as in (4), *bờ* (shore) goes with the noun *biển* (sea) as in (5) to indicate a specific place near the sea. The classifier *gốc* (foot) occurs with the noun *cam* (orange) as in (6), so it individuates the noun and indicates the foot of an orange tree. In (7), *thứ* (type) is used with the noun *bánh* (cake) to emphasize a specific kind of cake that is referred to in the preceding context. The classifier *hòn*

⁸ Even though the classifier *con* (animate) in the current study occurs with a noun indicating a non-living thing, I am going to refer to it by its usual designation “animate”. Also, I am not explaining the apparent cross-over here of *con* (animate) between animate and inanimate classification, and this is left for future research.

(round) individuates the noun ‘gem’ and classifies it as a small round object as in (8). It is surprising that the ninth most frequent inanimate classifier is *con* (animate) as it is the general classifier in the animate non-human classifier type recognized by all researchers. However, it occurs with inanimate nouns such as *con sông* (CL river) as in (9). This classifier appears 36 times with nouns indicating non-living things in this corpus, so this phenomenon is quite common.

- (1) Tôi sang xem *cái* cầu cao bên Hà Tây.
I come see CL(inani.) bridge high in Ha Tay
‘I came to see the high bridge in Ha Tay.’ (N1.174)⁹
- (2) nay *cây* kiếm muốn theo ta đánh giặc
now CL(tree, long) sword want follow me fight enemy
‘now the sword wants to fight the enemy with me’ (N2.280)
- (3) thấy *quả* bầu hổ tưởng là đá
see CL(fruit, round) gourd tiger think be stone
‘seeing the gourd, the tiger thought it was a stone’ (N2.69)
- (4) Hai ông bà thấy một *chiếc* thuyền lạ
Two he she see one CL(individual) boat strange
‘Both of them saw a strange boat’ (N2.312)
- (5) Đây là *bờ* biển cháu ạ,
Here is CL(shore) sea you ah
‘This is the seashore!’ (N2.313)
- (6) Tại dưới *gốc* cam có hai chum vàng
At under CL(foot) orange have two jar gold
‘Under the foot of the orange tree there are two jars of gold’ (N1.88)
- (7) chọn hai *thứ* bánh ấy đem lễ Trời, Đất, cùng Tiên Vương.
choose two CL(type) cake that bring kowtow Heaven Earth with Late Emperor
‘chose those two types of cakes to kowtow Heaven and Earth, with Late Emperor’ (N2.08)
- (8) *Con* quạ bèn thả ra một *hòn* ngọc,
CL(animate) raven then release out one CL(round) gem
‘The raven then released a gem’ (N1.28)

⁹ The examples given here in this chapter are labelled starting with (1). These examples are from the corpus for the current study. (N1.174) in (1) means that the example is from the Narrative Corpus (N), Book 1 and page number 174.

- (9) gặp một *con* sông rất sâu
 meet one CL(animate) river very deep
 ‘met a very deep river’ (N2.114).

The examples of the nine most frequent inanimate classifiers in the narrative corpus have been analysed. The distribution of these classifier types is discussed in the next section.

4.2.2 Overall distribution of frequent classifiers in the Narrative Corpus

As mentioned in 4.2.1, nine classifiers are the most frequent in the Narrative Corpus, each of them accounting for about 2% or more. Table 5 shows the overall distribution of these nine most frequent inanimate classifiers in the Narrative Corpus with the number of occurrences and percentages, which will be the focus of this section. The remaining 183 inanimate classifiers, which occur less than two percent each, are grouped into the ‘others’ category. Many of them rarely occur in the corpus. The findings of these classifiers are not reported in this section, but their distribution will be analysed in section 4.2.3. As the list of the classifiers in the ‘others’ category is too long, the table has been extended with all these infrequent inanimate classifiers in the Appendix A for reference.

Table 5: Distribution of frequent CLs in the Narrative Corpus

Inanimate CLs	No. of occurrences	%
<i>cái</i> (inanimate)	404	22.27
<i>cây</i> (tree, long object)	180	9.92
<i>quả</i> (fruit, round)	67	3.69
<i>chiếc</i> (individual)	59	3.25
<i>bờ</i> (bank, shore, fence)	46	2.54
<i>gốc</i> (root, foot)	45	2.48
<i>thứ</i> (kind, sort)	40	2.21
<i>hòn</i> (round, CL stone)	38	2.09
<i>con</i> (animate)	36	1.98
‘others’	913	49.56
Total	1828	100.00

As Table 5 shows, the most frequent classifier *cái* (inanimate) occurs 404 times, accounting for 22% of all the tokens found in the corpus. The second most frequent classifier *cây* (tree, long object) appears 180 times, approximately 10%. The classifiers *quả* (fruit, round object) and *chiếc* (individual) with 67 and 59 tokens respectively account for 3 percent each. The other five most frequent classifiers, which are *bờ* (bank, shore, fence), *gốc* (root, foot), *thứ* (type), *hòn* (round), and *con* (animate), occur less often, approximately 2 percent each. These nine classifiers altogether account for a half of all the tokens found in the corpus.

The general classifier *cái* (inanimate) appears with 184 different inanimate nouns in the Narrative Corpus, both concrete and abstract nouns. The nouns combining with this classifier may refer to a big thing such as ‘a boat’ as in (10a) or to a small thing such as ‘a toothpick’ as in (10b), and to abstract concepts such as ‘greed’ or ‘trick’ as in (10c-d).

- (10) a. ghép thành một *cái* thuyền lớn.
 join into one CL(inani.) boat big
 ‘joined into a big boat’ (N2.180)
- b. tôi đeo *cái* tăm xỉa răng.
 I make CL(inani.) toothpick
 ‘I am making the toothpick’ (N1.102)
- c. Rùa tuy ghét Khi vì *cái* thói tham lam
 Turtle though hate Monkey because CL(inani.) habit greedy
 ‘though Turtle hates Monkey because of his greed’ (N2.152)
- d. anh ta mới lập ra *cái* mẹo
 he then set up CL(inani.) trick
 ‘he then set up a trick’ (N1.128).

The classifier *cái* (inanimate) combines with a wide variety of nouns as a single classifier, and co-occurs with other specific classifiers in the double classifier construction in the Narrative Corpus. As double classifiers rarely occur in this corpus, they will be analysed in section 4.2.3 with other infrequent inanimate classifiers.

The second most frequent classifier *cây* (tree, long object) in this corpus mainly occurs with different kinds of trees or plants such as một *cây* lim (a tree of ironwood) as in (11a). In addition to the explicit indication of trees or plants, this classifier goes with other nouns to indicate

the ‘long’ objects such as *cây cột đèn* (CL light pole), *cây gậy* (CL cane), *cây kiếm* (CL sword), or *cây đàn* (CL musical instrument) as in (11b-e). The objects that these nouns refer to are considered as ‘long objects’, or at least in the Vietnamese perceptions, they are long.

- (11) a. *ngả một cây lim đẽo thuyền.*
 cut down one CL(tree, long) ironwood make boat
 ‘cut down an ironwood tree to make a boat’ (N2.376)
- b. *những cây cột đèn lớn*
 PL CL(tree, long) pole lamp big
 ‘big light poles’ (N2.284)
- c. *cầm cây gậy trong tay*
 hold CL(tree, long) cane in hand
 ‘holding the cane in the hand’ (N2.342)
- d. *trao hai cây kiếm ấy cho Lét, Le*
 give two CL(tree, long) sword DEM for Let Le
 ‘gave those two swords to Let and Le’ (N2.282)
- e. *mà chỉ mang theo cây đàn kì diệu.*
 but just take with CL(tree, long) musical instrument magical
 ‘but just took the magical musical instrument with him’ (N2.265).

As the third most frequent, *quả* (fruit, round object) mainly goes with nouns indicating different kinds of fruit such as *hai quả bầu* (two gourds) as in (12a). It also appears with other nouns such as *một quả núi đá* (a rocky mountain) or *quả trứng gà* (the chicken egg) as in (12b-c).

- (12) a. *Con biếu cha mẹ hai quả bầu.*
 I offer parents two CL(fruit, round) gourd
 ‘I offer you (parents) two gourds.’ (N2.190)
- b. *anh lại gặp một quả núi đá.*
 he again meet one CL(fruit, round) mountain rock
 ‘he met a rocky mountain again.’ (N2.114)
- c. *Anh hỏi quả trứng gà trong ổ cạnh chòi*
 He ask CL(fruit, round) egg chicken in net next to tent
 ‘He asked the chicken egg in the net next to the tent’ (N2.112).

As the fourth most frequent, *chiếc* (individual) appears quite often with 27 different inanimate nouns including nouns indicating means of transport such as một *chiếc* thuyền (a boat) as in (4) given in 4.2.1. It combines with nouns indicating food such as bốn *chiếc* bánh (four cakes) as in (13a); kitchen tools or containing items such as *chiếc* thoi (shuttle) and *chiếc* sọt (crate) as in (13b-c). It occurs with nouns referring to weapons such as *chiếc* kiếm (sword) as in (13d); personal items một *chiếc* giày (a shoe) as in (13e); or natural phenomenon một *chiếc* cầu vồng (a rainbow) as in (13f). It is noted that all the nouns appearing with *chiếc* (individual) are concrete nouns and none of them is an abstract noun.

- (13) a. Trên bàn có bốn *chiếc* bánh
 on table have four CL(individual) cake
 ‘On the table there are four cakes’ (N1.42)
- b. già sẽ tìm cho con *chiếc* thoi khác.
 I (old) will find for you CL(individual) shuttle other
 ‘I will find you another shuttle.’ (N2.389)
- c. đan xong bảy *chiếc* sọt rất đẹp.
 make complete seven CL(individual) crate very beautiful
 ‘made seven very beautiful crates.’ (N2.187)
- d. mượn *chiếc* kiếm thần
 borrow CL(individual) sword magic
 ‘borrowed the magic sword’ (N2.226)
- e. Tấm đánh rơi một *chiếc* giày.
 Tam drop off one CL(individual) shoe
 ‘Tam dropped a shoe.’ (N2.161)
- f. như một *chiếc* cầu vồng mọc lên từ mặt biển
 like one CL(individual) rainbow rise up from surface sea
 ‘like a rainbow arising from the surface of the sea’ (N2.113) .

Both the classifiers *bờ* (bank, shore, fence) and *gốc* (root, foot) appear quite often in the corpus. However, *gốc* (root, foot) is used with only nouns indicating different types of trees as in (6) repeated here, whereas *bờ* (bank) goes with a number of nouns indicating ‘the land alongside different structures of water geographically’ such as *bờ biển* (seashore) as in (5). It also combines

with other nouns such as *bờ ao* (the bank of the pond), *bờ giếng* (the bank of the well), *bờ ruộng* (the bank of the field), *bờ vực* (the bank of the abyss), *bờ rào* (CL fence), *bờ cõi* (CL border), *bờ suối* (the bank of the stream), or *bờ sông* (the bank of the river) as in (14).

(6) Tại dưới *gốc* cam có hai chum vàng

At under CL(foot) orange have two CL(jar) gold

‘Under the foot of the orange tree, there are two jars of gold’ (N1.88)

(14) sống tận trên *bờ* sông Mê Kông.

live as far as on CL(bank) river Mekong

‘live as far as near the bank of Mekong river.’¹⁰ (N2.312).

The classifier *thứ* (kind, sort) goes with many different nouns since it emphasizes the type of the ‘thing’ that the noun refers to. It appears with such nouns as *gỗ* (woods), *quần áo* (clothes), *gạo* (rice), *mía* (sugar cane), *cà* (eggplant), *đồ ăn* (edible), *tráng miệng* (dessert). As in (7) to be repeated here, it occurs with *bánh* (cake) to indicate the types of cakes that are highly valued. It is used to emphasize the ‘precious type of rice’ that makes the speaker so surprised at as in (15).

(7) chọn hai *thứ* bánh ấy đem lễ Trời, Đất, cùng Tiên Vương.

choose two CL(kind) cake that bring kowtow Heaven Earth with Late Emperor

‘chose those two types of cakes to kowtow Heaven and Earth, with Late Emperor’ (N2.08)

(15) Bạn làm thế nào lấy được *thứ* thóc nếp quý này, hả bạn?

You do how get positive CL(type) rice sticky precious this QUES you

‘How did you get this type of precious sticky rice, my friend?’ (N2.155).

The next most frequent classifier *hòn* (round) in the corpus occurs with six different inanimate nouns including *hòn ngọc* (CL gem) as in (8) in 4.2.1, *hòn than* (CL coal), *hòn đá* (CL stone), *hòn núi* (CL mountain), *hòn đất* (CL land), and *hòn đảo* (CL island) as in (16a-e). In addition to the function of individualization, *hòn* (round) not only indicates something round and

¹⁰ As explained above, the classifier *bờ* (bank, shore, fence) usually combines with a noun to indicate the land alongside different structures of water geographically. It does not indicate the noun such as the sea or the river itself. It classifies these nouns and categorizes them to indicate the bank/shore of something such as *bờ biển* (the shore of the sea), *bờ sông* (the bank of the river), *bờ giếng* (the bank of the well), *bờ ruộng* (the bank of the water field), or *bờ vực* (the bank of the valley). It cannot be a noun because in Vietnamese it cannot stand on its own.

small, but also refers to something big such as một *hòn* núi (a mountain) or một *hòn* đảo (an island) as in (16c) and (16e).

(16) a. và đem theo *hòn* than cháy đỏ hồng.

and carry with CL(round) coal burn red pink

‘and carried with a red burning piece of coal’ (N1.203)

b. Nó vào rừng bứt dây buộc một *hòn* đá thật to, tròng vào cổ,

It went into forest get string bind one CL(round) stone really big tie to neck

‘It went into the forest, got strings to bind a really big stone, and tied to the neck’ (N2.69).

c. Sau nghe nói có một *hòn* núi cao

Then hear say have one CL(round) mountain high

‘Then heard that there exists a high mountain’ (N1.85)

d. không có một *hòn* đất ném qua

not have one CL(stone) land throw through

‘there is not a piece of land to be thrown’ (N2.358)

e. dạt vào một *hòn* đảo hoang vu

flown to one CL(round) island deserted

‘flown to a deserted island’ (N2.362).

It is interesting to find that *con* (animate), the general animate non-human classifier, appears 36 times with nine different inanimate nouns in the corpus. It usually goes with nouns indicating long roads/paths or flows such as *con sông* (river) as in (9) in 4.2.1, *con suối* (stream), *con đường* (road/path) as in (17a-b). It also appears with nouns indicating weapons or tools such as *con dao* (knife), *con kiếm* (sword) as in (17c-d) or *con thoi* (shuttle) as in (17e), and even parts of human body such as *con mắt* (eye) as in (17f). This animate classifier occurs quite often with nouns indicating non-living things in the corpus.

(17) a. ngắm tất cả những *con* suối chảy qua buôn của nhà vua

look at all PL CL(ani.) stream flowing through village of King

‘looked at all the streams flowing through the King’s village’ (N2.114)

b. bèn cho mở một *con* đường rộng

then allow open one CL(ani.) road wide

‘then allowed to open a wide road’ (N1.183)

- c. Cô liền lấy *con* dao
 She immediately get CL(ani.) knife
 ‘She immediately got the knife’ (N2.215)
- d. bằng *con* kiếm gần gãy của mình.
 by CL(ani.) sword nearly broken of self
 ‘by his nearly broken sword.’ (N2.281)
- e. lội xuống giếng tìm *con* thoi
 get down pond look for CL(ani.) shuttle
 ‘got down into the pond to look for the shuttle’ (N2.388)
- f. *con* mắt tráo trung
 CL(ani.) eye showy
 ‘showy eyes’ (N1.17).

Above is the analysis of the most frequent inanimate found in the narrative corpus. The remaining classifier types occur much less often, which will be analysed in the next section.

4.2.3 Overall distribution of infrequent classifiers in the Narrative Corpus

For the remaining 183 inanimate classifiers which are less frequent in the Narrative Corpus, their distribution differs from one another. As mentioned above, each of these classifiers occur less than 2 percent in the corpus. As they are less frequent than the classifiers analysed in 4.2.2, they are grouped into ‘infrequent classifiers’ although some of them are not really rare. For example, *dòng* (long flow, style, line) appears 22 times in the corpus as in (18). The classifiers *đám* (procession, mass, patch) as in (19) or *hạt* (seed, small round object) as in (20) occur 20 times each in this corpus.¹¹

(18) nhưng đã bị *dòng* sông ngăn cách.

but PAST PASS CL(flow) river separate

‘but was separated by the river.’ (N2.227)

(19) anh ta đi ngang qua một *đám* cỏ,

he go across past one CL(patch) grass

‘he went across a patch of grass’ (N1.187)

¹¹ As mentioned in 4.2.2, the distribution of all the classifiers found in the corpus is shown in Appendix A.

(20) rồi *hạt* ngọc tự nhiên biến đi mất.
 then CL(round) pearl suddenly disappear
 ‘then the pearl suddenly disappeared.’ (N1.32).

However, many other classifiers appear less than 10 times each in this corpus. Most of these classifiers go with one or two inanimate nouns only. Specifically, forty-eight classifiers occur only two or three times each such as *bức* (mail) and *quyển* (volume) as in (21-22). Fifty-four other classifiers are rarely used in the corpus as each of them occurs only once such as *tờ* (sheet) and *viên* (small, round object) as in (22-24).

(21) viết cho anh một *bức* thư
 write for him one CL(mail) mail
 ‘wrote a mail to him’ (N1.20)

(22) nhưng chỉ biết lếu lảo dăm ba *quyển* sách
 but just know slovenly a few CL(volume) book
 ‘but just knew a few books in a slovenly way’. (N1.61)

(23) đem ra mỗi người một *tờ* giấy,
 bring out every person one CL(sheet) paper
 ‘brought out a sheet of paper for each person,’ (N1.152)

(24) cầm *viên* đá đi vào trong thành.
 hold CL(small, round) stone go into in citadel
 ‘held the stone going into the citadel.’ (N2.317).

The data shows that a large number of inanimate classifiers have been identified with a limited number of tokens in this corpus. Many of them are mensural classifiers which rarely appear including *chén* (cupful), *chai* (bottleful) as in (25)-(26). These words are nouns indicating objects which are used very often by Vietnamese people. They are like measure words, although in these cases, *chén* (cupful) and *chai* (bottleful) are used as mensural classifiers. They individuate the mass nouns *cơm* (cooked rice) in (25) and *mật ong* (honey) in (26).

(25) đòi ăn cả một *chén* cơm
 ask for eat all one CL(bowlful) rice
 ‘asked for a bowlful of rice’ (N2.100)

(26) và đổ vào một *chai* mật ong
 and pour in one CL(bottleful) honey
 ‘and poured in a bottleful of honey’ (N2.102).

Other nouns such as *vườn* (garden) are also used as mensural classifiers as in (27). The classifier *vườn* (garden) is used with the noun *cà* (eggplants) to indicate *một vườn cà* (a garden of eggplants) in this example.

(27) giống một *vườn* cà rất tốt.
 grow one CL(garden) eggplant very good
 ‘grow a very good garden of eggplants.’ (N1.193)

Clearly, a great variety of mensural classifiers are used with mass nouns in Vietnamese, resulting in a higher number of classifiers in the corpus. For instance, three different classifiers *nắm* (closed handful), *dúm* (handful), *túm* (handful) appear in the corpus to indicate ‘a handful of something’ such as *nắm thóc* (a handful of unprocessed rice) or *một dúm muối* (a handful of salt) as in (28a-b).

(28) a. Tám bốc cho gà *nắm* thóc.
 Tam get for chicken CL(handful) rice
 ‘Tam got a handful of rice for the chicken.’ (N2.160)
 b. thường chỉ ăn một *dúm* muối
 usually just eat one CL(handful) salt
 ‘usually eat just a handful of salt’ (N1.192).

Furthermore, many nouns can go with a number of classifiers, which leads to a higher number of classifiers identified in the narrative corpus. For instance, the noun ‘stone’ occurs with ten different classifiers including *hòn* (round), *viên* (round), *cục* (small piece), *tảng* (big stone), *phiến* (flat stone) and even human classifier *thằng* (human, male, young) in the corpus. However, the distribution of these classifiers with the same noun differs. As discussed earlier in this section, as one of the most frequent classifiers in the corpus, *hòn* (round) appears with the noun ‘stone’ most often. Next, *viên* (small, round), *cục* (piece), *tảng* (big piece), and *phiến* (flat stone) occur more often than the other classifiers do. In addition to the function of individuation, each of these classifiers appears to add more meaning to the noun ‘stone’. In (16b) repeated here for reference,

in the presence of *hòn* (round), the noun *hòn đá* (CL stone) is presupposed not to be too big or small so that the tiger in this story can get it and tie it to his neck.

- (16) b. Nó vào rừng bứt dây buộc một hòn đá thật to, tròng vào cổ,
 It went into forest get string bind one CL(round) stone really big tie to neck
 ‘It went into the forest, got strings to bind a really big stone, and tied to the neck’ (N2.69).

Similarly, in *viên đá* (CL stone) and *cục đá* (CL stone), *viên* (small, round) and *cục* (piece) indicate that the stone is so small that somebody can hold it in their hands as in (24) above or can pick it up as in (29a). In contrast, the classifier *tảng* (big stone) as in (29b) indicates that it is a very big stone which is even ‘bigger than a house’. The classifier *phiến* (flat stone) as in (29c) indicates that it is a big and flat stone that people can sit on. Surprisingly, the noun *đá* (stone) occurs with a human classifier *thằng* (human, male, young) as in (29d). This is a special phenomenon in which a human classifier goes with inanimate nouns *đá* (stone) and *tre* (bamboo). In fact, this example is what a rabbit said to a turtle in a story in the corpus while the turtle is moving so slowly. I believe that in this case the writer used personification when choosing this classifier for the noun ‘stone’ and ‘bamboo’.

- (29) a. đã nhặt lại cục đá
 PAST pick up again CL(piece) stone
 ‘picked up the stone again’ (N2.318)
- b. bằng những tảng đá to hơn cái nhà.
 by PL CL(big piece) stone big over CL(inani.) house
 ‘by the stones bigger than a house’ (N2.363)
- c. rồi lên ngồi trên một phiến đá.
 then get sit on one CL(flat) stone
 ‘then sat on a stone’ (N2.187)
- d. Như thế thà không đi như thằng đá, thằng tre.
 So then not go like CL(human, m, y) stone CL(human, m, y) bamboo
 ‘so slow, then it is better not to go like a stone or a bamboo tree.’ (N2.135).

As mentioned in 4.2.1, only three double classifiers including *cái con* (inani., animate), *cái đám* (inani., patch), and *cái vị* (inani., taste, type) appear in the Narrative Corpus with five tokens

altogether. In this double classifier construction, *cái* (inanimate) co-occurs with a specific classifier. The first construction *cái con* (inani., animate) appears with the noun *rẫy* (mountain field) as in (30a). It is interesting to note that the combination of *cái* (inanimate) and the general animate classifier *con* (animate) precedes a noun indicating a non-living thing such as *rẫy* (mountain field). The second double classifier *cái đám* (inanimate, patch) occurs with the noun *cỏ* may (grass) to indicate ‘a patch of grass’ as in (30c), in which *đám* (patch) individuates the noun ‘grass’ while *cái* (inanimate) seems to emphasize the noun. Similarly, *cái vị* (inanimate, taste, type) appears with the noun *thuốc* (medicine) as in (30e), in which *vị* (taste, type) individuates the noun while *cái* (inanimate) appears to emphasize the noun ‘type of medicine’. However, in other cases when *cái* (inanimate) goes with the noun, it individuates that noun. In the doubling construction, the appearance of *cái* (inanimate) is optional, so it is called the “extra” *cái* (inanimate) by previous researchers.

- (30) a. yêu thương *cái con* *rẫy* *mỡ màu* *bắp, lúa*
 love CL(inani.) CL(ani.) field fertile corn rice
 ‘love the fertile rice and corn field (in the mountain)’ (N2.282)
- b. vì *cái* *rẫy* *rộng nó lạnh*.
 because CL(inani.) mountain field large it cold
 ‘because the mountain field is large, it is cold.’ (N2.282)
- c. Anh như *cái đám* *cỏ* *may giữa đường*.
 You like CL(inani.) CL(patch) grass in path
 ‘You are like a patch of grass in the path.’ (N1.104)
- d. anh ta đi ngang qua một *đám* *cỏ*,
 he go across past one CL(patch) grass
 ‘he went across a patch of grass,’ (N1.187)
- e. *cái vị* *thuốc* *quý hóa cứu người* *sống được*
 CL(inani.) CL(taste, type) medicine valuable save human live get
 ‘the valuable medicine that saves human life’ (N1.74)
- f. *vị* *thuốc* *này thật sự là quý hóa*
 CL(type) medicine this really be valuable
 ‘this type of medicine is really valuable’ (N1.75).

In this double classifier construction, *cái* (inanimate) appears to be used for emphasis while the specific classifier individuates the nouns it precedes. Thus, *cái* (inanimate) can be omitted, leaving the specific classifier going with the noun. The evidence in the corpus shows that *đám* (patch) and *vị* (taste, type) appear with the nouns ‘grass’ and ‘medicine’ without the presence of *cái* (inanimate) as in (30d) and (30f). Therefore, *cái* (inanimate) is considered “extra” in this case. However, for the combination of *cái con* (inani., animate) as in (30a), only *cái* (inanimate) is found to occur with the noun *rẫy* (mountain field) in the corpus as in (30b). This means, *cái* (inanimate) can co-occur with *con* (animate) before combining with noun as in (30a) or can appear with the noun on its own as in (30b). The use of *cái* (inanimate) appears to need more exploration, which will be further investigated and discussed in section 5.3.

Above is the analysis of the nine most frequent classifiers and a number of infrequent inanimate classifiers found in the Narrative Corpus. The next section will summarize the main findings from this corpus.

4.2.4 Summary

In brief, with the examination of the 1828 classifier tokens in the Narrative Corpus, 192 inanimate classifier types have been identified in the corpus. The data shows big differences in the frequency and distribution of these classifiers. Although the overall frequency of classifier use in this corpus is 160 per 10,000 words, the frequency of different classifiers greatly differs. As the general inanimate classifier, *cái* (inanimate) is the most frequent. This classifier also co-occurs with other specific classifiers to generate double classifiers although this construction rarely appears in the narratives. In the Narrative Corpus, a number of inanimate classifiers appear frequently with different nouns while others are less frequent or rare with a limited number of nouns. This reveals that the capability of combination with nouns depends on the properties of each classifier. However, clearly, the results of the study ascertain that a Vietnamese noun can go with different classifiers, and one classifier can occur with many different nouns. This finding supports the claim made by researchers including D. H. Nguyen (1957), Thompson (1965), and P. P. Nguyen (2002). These are the main findings from the Narrative Corpus. The next section will present the major findings from the Online Newspaper Corpus.

4.3 Findings from the Online Newspaper Corpus

The major findings of classifier use regarding frequency and distribution in the Online Newspaper Corpus are to be reported and analysed in this section. The frequency of inanimate classifiers is presented in 4.3.1, the overall distribution of frequent inanimate classifiers in 4.3.2, and the distribution of infrequent inanimate classifiers in 4.3.3.

4.3.1 Frequency of classifiers in the Online Newspaper Corpus

With the analysis of 10063 inanimate NPs in the Online Newspaper Corpus of 135,900 words, a total of 2472 classifier tokens have been found. The overall frequency of inanimate classifiers in this corpus is 182 per 10,000 words, a little higher than the rate in the Narrative Corpus, 160 per 10,000 words. However, the frequency of different classifiers greatly varies in the corpus. One hundred fifty-three inanimate classifier types have been identified in the corpus, with 151 single and two double classifiers. The frequency of the ten most frequent inanimate classifiers in the online newspaper corpus is shown in Table 6.

Table 6: Frequency of CLs in the Online Newspaper Corpus

Inanimate CLs	No. of occurrences	Frequency (No. of CLs per 10,000 words)
<i>sự</i> (event)	277	20.38
<i>cuộc</i> (life, strike, match)	187	13.76
<i>cái</i> (inanimate)	180	13.25
<i>chiếc</i> (individual)	144	10.60
<i>việc</i> (activity)	116	8.54
<i>bộ</i> (set)	68	5.00
<i>vụ</i> (catastrophe)	61	4.49
<i>con</i> (animate)	58	4.27
<i>dòng</i> (river, line)	57	4.19
<i>căn</i> (unit of house)	55	4.05
‘others’	1269	93.38
Overall	2472	181.90

All the remaining 143 inanimate classifiers which occur less often are put in the ‘others’ category in Table 6. I will not look at the frequency rates of these classifiers because they are low although the total number of tokens is 1269 altogether. However, their overall distribution will be analysed later in section 4.3.3. Table 6 shows the ten most frequent inanimate classifiers in the online newspaper corpus including *sự* (event), *cuộc* (life, strike, match), *cái* (inanimate), *chiếc* (individual), *việc* (activity), *bộ* (set), *vụ* (catastrophe), *con* (animate), *dòng* (river, line), and *căn* (unit of house). Different from the narrative genre, *sự* (event) is the most frequent classifier in the online newspapers with the frequency of 20 per 10,000 words, 277 tokens. It has the function of nominalizing verbs, mainly stative and adjectival verbs it precedes. With 187 tokens, *cuộc* (life, strike, match) is the second most frequent classifier with the rate of over 13 per 10,000 words. Different from the Narrative Corpus, in the Online Newspaper Corpus, *cái* (inanimate) is just the third most frequent with 180 tokens having the similar frequency as *cuộc* (life, strike, match). The next common classifier *chiếc* (individual) has the frequency of 10 per 10,000 words with 144 tokens found in the corpus, followed by *việc* (activity) having the frequency of 8 per 10,000 words with 116 tokens. The other five classifiers *bộ* (set), *vụ* (catastrophe), *con* (animate), *dòng* (river, line), and *căn* (unit of house) are less frequent with the number of tokens and frequencies shown in Table 6. Examples of these frequent classifiers in the Online Newspaper Corpus are not given here, but will be analysed in the next section.

4.3.2 Overall distribution of frequent classifiers in the Online Newspaper Corpus

The overall distribution of the ten most frequent classifiers out of 153 inanimate classifiers identified in the Online Newspaper Corpus is presented in this section. Each of them accounts for more than 2% of all the tokens in the corpus. As shown in Table 7, the most frequent classifier *sự* (event) with 277 tokens accounts for about 11% of all the inanimate classifier tokens in the corpus. Next, *cuộc* (life, strike, match) and *cái* (inanimate) with 187 and 180 tokens respectively account for about 7% each. The fourth frequent classifier *chiếc* (individual) in the corpus accounts for approximately 6%. The other six classifiers occur less often with the numbers of occurrences and percentages shown in Table 7. These ten most frequent classifiers altogether account for nearly a half of all the inanimate classifier tokens found in the corpus. All the remaining 143 inanimate classifiers are put in the ‘others’ category in Table 7 because they appear less often, accounting for less than 2% each. They are too long to be listed here in this table. Although these 143

classifiers appear with 1269 tokens altogether in the corpus, each of them has a limited number of occurrences. Table 7 shows the overall distribution of the ten most frequent inanimate classifiers in the corpus with their numbers of occurrences and percentages. For details of the distribution of all the inanimate classifiers in the Online Newspaper Corpus, please see Appendix B.

Table 7: Distribution of frequent CLs in the Online Newspaper Corpus

CLs	No. of occurrences	%
sự (event)	277	11.21
cuộc (life, strike, match)	187	7.56
cái (inanimate)	180	7.28
chiếc (individual)	144	5.83
việc (activity)	116	4.69
bộ (set)	68	2.75
vụ (catastrophe)	61	2.47
con (animate)	58	2.35
dòng (river, line)	57	2.31
căn (house, disease)	55	2.22
‘others’	1269	51.33
Overall	2472	100.00

These frequent inanimate classifiers are analysed with examples for illustration in this section. As reviewed in 2.4.4, Vietnamese classifiers have the nominalization function (H. T. Nguyen 2004). As the most frequent in the Online Newspaper Corpus, *sự* (event) is one of the classifiers functioning as a nominalizer. This classifier occurs with 192 different stative and adjectival verbs in the corpus. As in (31), it nominalizes the verb *phát triển* (develop) and turns it into the noun *sự phát triển* (development).

- (31) *Sự* *phát triển* của mạng xã hội
 CL(action) develop of network social
 ‘The development of a social network’. (O28.1892)¹²

¹² (O28.1892) in (31) means that the example is from the VONC (O), article number 28, and token number 1892.

The group of classifiers functioning as nominalizers are common in this corpus, but it has not yet received much attention from researchers. I will discuss it further later in section 5.4. The second most frequent classifier *cuộc* (life, strike, match) also performs the function of nominalization. It goes with 55 different verbs or nominals in the corpus including *cuộc* đình công (strike), *cuộc* sống/*cuộc* đời (life), *cuộc* họp (meeting), *cuộc* bầu cử (election), *cuộc* chiến (fight), and *cuộc* thi (contest). This classifier is frequent in the corpus since it can go with a wide variety of nominals or verbs indicating the process of some activity or happening in a certain period of time. It individuates nouns and/or nominalizes verbs indicating a process such as *cuộc* phỏng vấn (CL interview) as in (32).

- (32) Trong *cuộc* phỏng vấn với Korea Times
 During CL(meeting) interview with Korea Times
 ‘During the interview with Korea Times’ (O3.142)

As the third and fourth most frequent classifiers in the corpus, *cái* (inanimate) and *chiếc* (individual) occur with a large variety of nouns in the corpus. The general classifier *cái* (inanimate) appears with 66 different nouns or nominals in this corpus. It goes with concrete nouns such as ‘*cái* máy nướng bánh’ (the toaster) as in (33a) and with abstract nouns such as ‘*cái* suy nghĩ’ (the thought) as in (33c). It even goes with non-classified nouns such as ‘*cái* màu tím thẫm’ (the colour of dark purple) as in (33b).

- (33) a. với *cái* máy nướng bánh mì sáu trăm năm mươi nghìn ở nhà
 with CL(inani.) machine toast bread six hundred fifty thousand at home
 ‘with the toaster valued at 650,000 VND at home’ (O55.5414)
- b. *cái* màu tím thẫm của nước biển
 CL(inani.) colour purple dark of water sea
 ‘the dark purple colour of the sea water’ (O33.2555)
- c. tôi thấy sợ *cái* suy nghĩ về việc con người như món hàng
 I find scared CL(inani.) thought about issue CL(ani.) human like CL(item) goods
 ‘I am scared of the thought about humans as items’ (O57.5644).

However, the classifier *chiếc* (individual) appears with concrete nouns such as ‘*chiếc* huy chương’ (the medal) as in (34). It is found that with 144 tokens, this classifier goes with 60 different

nouns in the corpus, but they are all concrete nouns. Thus, it is assumed that *chiếc* (individual) can go with concrete nouns, but cannot combine with abstract nouns.

- (34) *chiếc* huy chương ấy là *niềm* tự hào vô giá
 CL(individual) medal that be CL(sentiment) proud invaluable
 ‘that medal is an invaluable pride’ (O11.635).

The next frequent classifier in this corpus is *việc* (activity). Like *sự* (event), the classifier *việc* (activity) also functions as a nominalizer in Vietnamese as claimed by Hoang (1996). However, almost all the verbs that are classified and nominalized by *việc* (activity) are action verbs. For example, *việc* (activity) nominalizes the verb ‘chăm sóc’ (take care of) and turns it into the nominal phrase *việc chăm sóc* (taking care of) as in (35). This classifier goes with 95 different verbs in the corpus and its primary function is nominalization and individuation. This phenomenon is interesting. As mentioned earlier in this section, this will be further discussed in section 5.4.

- (35) *Việc* chăm sóc một gia đình lớn không dễ dàng.
 CL(activity) take care of one family big not easy
 ‘Taking care of a big family is not easy.’ (O4.208)

The next most frequent classifier in this corpus is *bộ* (set). It classifies a number of nouns indicating a complete set of something or a collection of something. As in (36), it classifies and individuates the noun ‘hồ sơ tài chính’ (finance profile). This classifier appears with 20 different nouns in the corpus including *bộ hồ sơ* (profile), *bộ ria mép* (moustaches), *bộ não* (brain), *bộ lông* (set of fur), *bộ cốc* (set of cups), *bộ quần áo* (set of clothes), *bộ trang bị* (set of equipment), *bộ sưu tập* (collection), *bộ ảnh* (set of pictures), *bộ luật* (code), *bộ phim* (movie), *bộ nhớ* (memory), *bộ vi xử lý* (microprocessor), *bộ lưu trữ* (storage), *bộ khung* (framework).

- (36) để có một *bộ* hồ sơ tài chính hợp lệ,
 to have one CL(set) profile finance appropriate
 ‘to have a set of appropriate financial profile,’ (O67.6635)

The next frequent classifier *vụ* (catastrophe) appears with 23 different nouns in the corpus. It usually classifies and individuates nouns indicating mostly unexpected bad incidents such as *vụ hỏa hoạn* (fire) as in (37a). It is found to classify and individuate other nouns including *vụ cướp* (robbery), *vụ trộm cắp* (theft), *vụ đột nhập* (break-in), *vụ án* (case), *vụ bắn* (shooting), *vụ máy bay*

roi (plane crash), vụ cháy tàu (ship fire), vụ tai nạn (accident), and vụ khủng hoảng (CL crisis) as in the example in (37b).

(37) a. đã bị vụ hoả hoạn năm 1697 thiêu rụi phần lớn
 PAST PASS CL(catastrophe) fire year 1697 destroy part big
 ‘was mostly destroyed by the fire in 1697’ (O85.7573).

b. Đây là vụ khủng hoảng nghiêm trọng nhất trong lịch sử 104 năm của hãng.
 Here be CL(catastrophe) crisis serious most in history 104 year of firm
 ‘This is the most serious crisis in the firm’s history of 104 years.’ (O104.8584)

It is interesting to find that like in the Narrative Corpus, *con* (animate), an animate non-human classifier, appears with a number of different inanimate nouns in this corpus. It classifies and individuates nouns indicating roads or rivers such as *con đường* (road), *con lộ* (road), *con phố* (street), *con ngõ* (alley), *con dốc* (slopes), *con sông* (river), *con mương* (ditch), *con kênh* (canal), *con sóng* (wave), *con nước* (tide); or nouns indicating boats or vehicles such as *con thuyền* (boat), *con xe* (vehicle); or *con chữ* (letter), *con số* (number); and even nouns indicating parts of a human body such as *con mắt* (eye), *con tim* (heart). This classifier combines with the noun *sóng* (wave) as in (38a) while it appears with the noun *chữ* (letter) as in (38b). It also goes with the noun ‘number’ as in (38c) and ‘audience’s hearts’ as in (38d).

(38) a. giữa những con sóng cao ngút của đại dương.

among PL CL(ani.) wave towering of ocean
 ‘among the towering waves of the ocean.’ (O33.2554)

b. mang con chữ đến các vùng sâu vùng xa,

bring CL(ani.) letter to PL area deep area far
 ‘bring the letters/writing to remoted areas,’ (O50.4759)

c. và con số này gia tăng tịnh tiến

and CL(ani.) number this increase constantly
 ‘and this number increases constantly’ (O26.1702)

d. khiến cho hàng triệu con tim khán giả phải thổn thức

make for row million CL(ani.) heart audience have to thrill
 ‘made millions of audience’s hearts thrilled’ (O82.7424).

The classifier *dòng* (flow, line) appears with 21 different inanimate nouns including nouns indicating rivers or a flow of water including *dòng chảy* (flow of running), *dòng nước* (flow of water), *dòng sông* (river) as in (39a). It also goes with nouns indicating a flow of something such as *dòng kiều hối* (remittance inflows), *dòng tiền* (flow of money) as in (39b). In addition, it is used with nouns indicating a line of products such as *dòng xe* (line of vehicles), *dòng máy bay* (line of airplanes) as in (39c). Also, it appears with nouns indicating lines of words such as *dòng giới thiệu* (line of introduction), *dòng nhắn gửi* (message), and *dòng chữ* (line of words) as in (39d).

- (39) a. các *dòng* sông vẫn cứ lần lượt chết lâm sàng,
 PL CL(flow) river still in turn die clinically
 ‘the rivers are still dying clinically one after the other,’ (O53.5138)
- b. nhìn thấy *dòng* tiền lớn đổ vào hệ thống ngân hàng thương mại
 see CL(flow) money large pour into system banking commercial
 ‘see that the large flow of money put into the commercial banking system’ (O16.905)
- c. buộc phải ngưng sản xuất *dòng* máy bay này,
 have to stop produce CL(line) airplane this
 ‘have to stop producing this line of airplanes,’ (O104.8583)
- d. Những *dòng* chữ này
 PL CL(line) word this
 ‘These lines of words’ (O129.9901).

The tenth most frequent classifier *căn* (house, disease) appears quite often in the corpus though it combines with a limited number of nouns. This classifier usually goes with nouns indicating ‘houses, rooms, apartments, villas or tents’ and with nouns indicating diseases. Specifically, it is used with the noun *nhà* (house) as in (40a), and the noun *bệnh ung thư* (the disease of cancer) as in (40b).

- (40) a. nó chưa một lần về *căn* nhà đó.
 he not yet one time come back CL(house) house that
 ‘he has not yet come back to that house once.’ (O39.3332)
- b. đi về thế giới khác do *căn* bệnh ung thư.
 to go world another due to CL(disease) disease cancer
 ‘to go to another world due to the disease of cancer.’ (O70.6763).

In sum, the uses of the ten most frequent inanimate classifiers in the Online Newspaper Corpus have been analysed. It is interesting to see that many of these frequent classifiers differ from those in the Narrative Corpus. However, these two corpora have three common classifiers *cái* (inanimate), *chiếc* (individual), and *con* (animate). These classifiers are quite conventional and can combine with a great variety of inanimate nouns. It is worth noting that other classifiers including *sự* (event), *cuộc* (life, strike, match), *việc* (activity), *bộ* (set), and *vụ* (catastrophe) are common in the Online Newspaper Corpus although they are not frequent in the Narrative Corpus. Specifically, *sự* (event) and *việc* (activity) rarely occur in the Narrative Corpus with only six and two tokens respectively. These findings reveal evidence of variation in classifier use across these two genres, especially the distribution of frequent classifiers. This comparison is discussed in chapter 5. Apart from these frequent classifiers, the other 143 inanimate classifiers, which appear less often in this corpus, are a substantial part of the Vietnamese classifier system. Their distribution is presented in the next section with the analysis of some of them.

4.3.3 Overall distribution of infrequent classifiers in the Online Newspaper Corpus

These 143 infrequent inanimate classifiers altogether account for a half of all the classifier tokens found in the corpus. Each of them is less than 2%. They are called ‘infrequent classifiers’ in order to separate from the ten most frequent classifiers analysed in 4.3.2. To be exact, they are just less frequent classifiers. Although they appear less often than the ten most frequent classifiers, their distribution greatly varies. A number of them are more frequent than the others as many of them are rare. Specifically, out of these 143 classifiers, 45 of them appear 10 to 48 times each in this corpus. These 45 classifiers include *loại* (kind, sort) with 48 occurrences, *chuyến* (trip) 46 occurrences, *cây* (tree) 45, *bức* (picture) 43, *ngôi* (house) 41, *số* (amount) 39, *trận* (match, fight) 28, and *con* (anger, wind) 18. Examples of some of these classifiers in the Online Newspaper Corpus will be given and analysed in this section.

As a type classifier, *loại* (kind, sort) goes with a large variety of concrete nouns in the corpus including flowers, trees, seeds, leaf, presents, tests, pollutions, masks, dictionaries, vehicles, vegetables and different kinds of food. This classifier is used to emphasize the type of thing that the noun refers to (Emeneau 1951). As in (41), it emphasizes the certain sorts of ‘unqualified’ medical masks that are made by a workshop.

(41) đã phát hiện các *loại* khẩu trang y tế do xưởng này
 PAST discover PL CL(sort) mask medical by workshop this
 ‘discovered the sorts of medical masks made by this workshop’ (O72.6867)

The classifier *chuyến* (trip) appears 46 times in the corpus. It classifies and individuates ten different nouns which are mainly ‘flight, journey, business trip, voyage, overseas trip, shipment, or tour’. It also nominalizes the verb bay (fly) and individuates the nominal *chuyến bay* (flight) as in (42).

(42) trên *chuyến* bay tới Iran
 on CL(trip) fly to Iran
 ‘on the flight to Iran’ (O18.1038)

Similarly, with 43 tokens found in the corpus, *bức* (picture, wall) mainly combines with nouns indicating ‘painting, picture, photo, drawing, mail, or wall’. It classifies and individuates the nouns such as *bức họa* (picture) as in (43).

(43) *Bức* họa bị tấn công có tên “Bust of a Woman”
 CL(picture) picture PASS attack have name “Bust of a Woman”
 ‘The picture that was attacked is named “Bust of a Woman”’ (O79.7269).

The classifier *ngôi* (house), with 41 tokens found in the corpus, classifies and individuates a limited number of nouns. It usually appears with nouns indicating ‘house, school, village’ such as *ngôi trường* (CL school), *ngôi sao* (CL star), and *ngôi làng* (CL village) as in (44).

(44) *Ngôi* làng xinh đẹp này
 CL(house) village beautiful this
 ‘this beautiful village’ (O74.7128).

However, *số* (amount), which appears 39 times in the corpus, classifies and individuates 20 different nouns including ‘money, funds, medicine, masks, goods, materials, and salt’. This classifier goes with the noun *tiền* (money) to indicate an ‘amount of money’ as in (45).

(45) họ vẫn nhận được một *số* tiền nhỏ hơn
 they still receive get one CL(amount) money small less
 ‘they still get a smaller amount of money’ (O04.203)

The six classifiers analysed above are among the subgroup of 45 classifiers which appear more often than the remaining classifiers. The other subgroup of 98 inanimate classifiers rare. Each of them appears less than ten times. Among this subgroup, 66 inanimate classifiers appear only one to three times each, such as *quyển* (volume) as in (46).

- (46) tôi mới chỉ đọc một lần từ quyển sách lớp 1
 I just only read one time from CL(volume) book grade 1
 ‘I just read only once from the Grade 1 book’ (O67.6474).

The two double classifiers identified in this corpus are *cái phần* (section, part) with two tokens and *cái bộ* (set) one token. They appear with the noun ‘movie’ as in (47) and the noun ‘perfection’ as in (48). In this construction, the specific classifiers classify and individuate the nouns while *cái* (inanimate) is to be used for emphasizing the nouns as prior researchers argue (Diep 2005; H. T. Nguyen 2004; Simpson and Ngo 2018). The use of double classifiers and the function of each classifier in the combination will be further discussed in 5.2.2.

- (47) ngồi xem cái bộ phim hay đọc cuốn sách đó
 sit watch CL(inani.) CL(set) movie or read CL(volume) book that
 ‘sit to watch that movie or read that book’ (O58.5853)
- (48) để đưa cái phần hoàn hảo lên ảnh ảo
 to bring CL(inani.) CL(part) perfect to image virtual
 ‘to bring the part of perfection to the virtual image’ (O110.8888).

In sum, a large number of actual inanimate classifier types including 151 single and two double classifiers have been identified in the Online Newspaper Corpus. The distribution of these classifiers found in this corpus greatly differs. However, it is different from the narratives, *sự* (event) and *cuộc* (life, strike, match) are the most frequent, with higher frequencies than *cái* (inanimate) in this corpus. A large number of classifiers which are infrequently used in the corpus combine with a limited number of nouns. Double classifiers also rarely appear in this corpus. A number of classifiers functioning as nominalizers have been found in the corpus. These classifiers appear quite frequently in the online newspaper corpus although they are rarely used in the narrative corpus. This shows evidence for variation in classifier use in these two genres. These

classifiers will be discussed in section 5.4. The next section presents the major findings from the Spoken Corpus.

4.4 Findings from the Spoken Corpus

In this section, I will present the findings from the Spoken Corpus. As described in section 3.1.3, this corpus consists of twenty-two talk show episodes with the total duration of 14 hours and the word count of 151,000 words. Forty-six Vietnamese native speakers belong to three age groups of 14 older speakers (over 50 years old), 18 middle-aged speakers (between 30 and 50 years old), and 14 younger speakers (under 30 years old). Section 4.4.1 presents the frequency of inanimate classifiers in the corpus. The overall distribution of frequent and infrequent inanimate classifiers is analysed in 4.4.2 and 4.4.3 respectively. Exceptional cases that appear in this corpus are discussed in section 4.4.4.

4.4.1 Frequency of inanimate classifiers in the Spoken Corpus

With the examination of 8911 noun phrases in the Spoken Corpus, the study has found 4326 classifier tokens including 3879 single and 403 double classifier tokens, and 44 exceptional cases. One hundred thirty-four actual classifier types have been identified in the corpus, which includes 48 double and 86 single classifier types, excluding 44 exceptional classifier tokens. The frequency of inanimate classifiers in the Spoken Corpus is 286 per 10,000 words, which is much higher than the rates in the Narrative and Online Newspaper Corpora. The frequency of the nine most frequent classifiers with the rate of over 3 per 10,000 words in the spoken corpus is shown in Table 8. All the remaining less frequent inanimate classifiers, having a frequency of less than 3 per 10,000 words each, are put in the ‘others’ category in the table. Their frequency will not be discussed as each of them accounts for less than one per cent of all the tokens in the corpus.

As Table 8 shows, *cái* (inanimate) is the most frequent at a really high rate of 176 per 10,000 words in the spoken corpus. This number of tokens *cái* (inanimate) in Table 8 and 9 are only the tokens in which *cái* (inanimate) appears as a single classifier, excluding those combining with other classifiers or other *cái* (inanimate) forms. The tokens of *cái* (inanimate) in the combination with other classifiers or *cái* (inanimate) forms are temporarily put in the ‘others’ category in these tables. The specific classifiers *bài* (song, lesson, text) and *cuộc* (life, strike, match) are far less frequent, having the frequency of 13 per 10,000 words. Next, the frequency

rates of the double classifier *cái sự* (inanimate, event) and the single classifier *sự* (event) are 9 and 8 per 10,000 words respectively. The other four classifiers *chiếc* (individual), *tình* (relationship), *con* (animate), and *đám* (procession, patch, mass) are less frequent at a rate of three to four per 10,000 words. The distribution of these frequent inanimate classifiers will be analysed in detail in section 4.4.2 although examples of these nine most frequent classifiers are given and analysed in this section.

Table 8: Frequency of CLs in the Spoken Corpus

Classifiers	No. of tokens	Frequency (No. of CLs per 10,000 words)
cái (inanimate)	2658	176.98
bài (song, lesson, text)	204	13.51
cuộc (life, strike, match)	201	13.31
cái sự (inanimate, event)	144	9.54
sự (event)	129	8.54
chiếc (individual)	65	4.30
tình (relationship)	61	4.04
con (animate)	58	3.84
đám (procession, patch, mass)	47	3.11
‘others’	759	50.31
Overall	4326	286.49

As the most frequent classifier in the Spoken Corpus, *cái* (inanimate) appears with the noun ‘script’ in (49).

- (49) bắt đầu viết một cái kịch bản
begin write one CL(inani.) script
‘begin to write a movie script’ (S2.5108)¹³

¹³ (S2.5108) in (49) means that the example is from the Spoken Corpus (S), talk show episode number 2, and token number 5108.

The classifier *bài* (unit of song, lesson) nominalizes the verb ‘hát’ and individuates the nominal ‘*bài hát*’ (song) as in (50). This classifier is one of the classifiers with the function of nominalization found in this corpus.

- (50) *Bài hát* đó giới thiệu về rất nhiều *món ăn*
 CL(unit of song) sing that introduce about very many CL(dish) eat
 ‘That song introduces a lot of dishes’ (S17.458)

As analysed in 4.3.2, *cuộc* (life, strike, match) is a classifier with the nominalization function. This classifier is also frequent in the spoken corpus. It nominalizes the verb *trò chuyện* (talk) it precedes and individuates it as in (51).

- (51) đang theo dõi *cuộc* trò chuyện với ông ĐĐC
 PROG watch CL(talk) talk with Mr. ĐĐC
 ‘(are) watching the talk with Mr. ĐĐC’ (S9.8875)

The single classifier *sự* (event) nominalizes the verb *mong muốn* (desire) as in (52). This classifier co-occurs with *cái* (inanimate) in the double classifier *cái sự* (inanimate, event), which precedes *thành công* (succeed) as in (53). In this double classifier, *sự* (event) nominalizes the verb while *cái* (inanimate) emphasizes the noun/nominal. This means that *sự* (event) can appear either as a single classifier or with *cái* (inanimate) in the doubling construction *cái sự* (inanimate, event). This doubling construction will be discussed in section 5.2.2.

- (52) bởi vì đây nó là *sự* mong muốn kết nối của khán giả
 because here this is CL(event) desire connect of audience
 ‘because this is the audience’s desire to connect’ (S2.5148)

- (53) một phần của *cái sự* thành công đó
 one part of CL(inani.) CL(event) succeed that
 ‘a part of that success’ (S2.5238).

The sixth most frequent classifier in this corpus is *chiếc* (individual). It classifies and individuates the noun *hộp* (box) it precedes as in (54).

- (54) Và đây *chiếc* hộp bí mật của chúng tôi.
 And here CL(individual) box secret of us
 ‘And here is our secret box.’ (S14.3662).

It is interesting to find that the general classifier *con* (animate) is frequent and appears with many different inanimate nouns in the Spoken Corpus. This classifier classifies and individuates the noun *đường* (road) as in (55).

- (55) Đại học không phải là *con* đường duy nhất
 University not be CL(road) road only
 ‘University is not the only road’ (S6.7680)

One further frequent classifier in the corpus, *đám* (procession), appears with the noun *cưới* (wedding) as in (56). It individuates the noun and indicates the procession of a wedding.

- (56) mà thấy *đám* cưới vui, ở quê í
 but find CL(procession) wedding joyful at home village eh
 ‘but (I) find that weddings in home villages are joyful’ (S16.272)

In sum, the overall frequency of inanimate classifiers in the Spoken Corpus has just been reported with examples for illustration. The overall distribution of these classifiers is presented in the next section.

4.4.2 Overall distribution of frequent classifiers in the Spoken Corpus

Although the distribution of each classifier found in the narrative and online newspaper corpora differs, the distribution of classifiers shows a great difference in the spoken corpus. Out of the 134 inanimate classifiers identified in the spoken corpus, *cái* (inanimate) is the most frequent, with 2658 tokens, accounting for over 61%. The remaining 133 classifiers altogether are just about 38% of all the tokens in the corpus. The overall distribution of the nine most frequent classifiers in the corpus is shown in Table 9. Following *cái* (inanimate), the classifiers *bài* (song, lesson, text) and *cước* (life, strike, match) are frequent with 204 and 201 tokens respectively, accounting for 4% each. Both the doubling construction *cái sự* (inanimate, event) and *sự* (event) are quite frequent, with 144 and 129 tokens respectively, each accounting for approximately 3 per cent. This means *sự* (event) in both single and doubling constructions is used quite frequently in the spoken corpus. The data shows that *cước* (life, strike, match) and *sự* (event) are frequent in the online newspaper and spoken corpora, but not in the narrative corpus. The other four classifiers *chiếc* (individual), *tình* (relationship), *con* (animate), and *đám* (procession, patch, mass) account for over 1% each. All the other remaining 125 inanimate classifiers, accounting for under 1% each,

are grouped into the ‘others’ category in Table 9. As the distribution of all the inanimate classifiers in the Spoken Corpus is a long list, it is put in Appendix C for reference.

As Table 9 shows, the distribution of *cái* (inanimate) shows a big difference in the use of classifiers in the spoken corpus compared to the narrative and online corpora. It is used far more frequently in comparison with the distribution of other inanimate classifiers. Specifically, thirty classifiers occur two or three times each while forty other classifiers appear only once each. The distribution of these infrequent classifiers will be analysed in section 4.4.3. This big difference in the distribution of classifiers depends on several factors, such as the number of nouns that the classifier can go with and the number of occurrences of those noun(s) in the corpus depending on the content of the discourse. The most frequent classifier, *cái* (inanimate), appears with over a thousand different nouns/nominals, either concrete or abstract, in the spoken corpus. The use of this classifier will be discussed in section 5.3.

Table 9: Overall distribution of frequent CLs in the Spoken Corpus

Classifiers	No. of occurrences	%
<i>cái</i> (inanimate)	2658	61.46
<i>bài</i> (song, lesson, text)	204	4.72
<i>cuộc</i> (life, strike, match)	201	4.65
<i>cái sự</i> (inanimate, event)	144	3.33
<i>sự</i> (event)	129	2.98
<i>chiếc</i> (individual)	65	1.50
<i>trình</i> (relationship)	61	1.41
<i>con</i> (animate)	58	1.34
<i>đám</i> (procession, patch, mass)	47	1.09
'others'	759	17.55
Overall	4326	100.00

The classifier *bài* (song, lesson, text) appears with a limited number of nouns although it is the second most frequent in this corpus. It combines with thirteen different nouns including *bài thơ* (poem), *bài hát* (songs), *bài học* (lesson), *bài ca* (song), *bài phỏng vấn* (interview), *bài giới thiệu* (introduction), *bài viết* (writing), *bài báo* (article), and *bài phát biểu* (speech). It classifies

and individuates the noun *thơ* (poem) as in (57a). It nominalizes the verbs *hát* (sing) and *học* (study) and individualizes the nominals *bài hát* (a song) and *bài học* (the lesson) as in (57b-c). Also, it nominalizes the verb ‘*phát biểu*’ (speak) and turning it into the noun ‘speech’ as in (57d).

- (57) a. cũng làm được độ khoảng gần 200 *bài* *thơ*
 also write get about nearly 200 CL(poem) poem
 ‘also wrote about nearly 200 poems’ (S7.7873)
- b. thì thực sự đó là một *bài* *hát* rất tuyệt vời.
 then actually that be one CL(song) sing very wonderful
 ‘then actually that is a very wonderful song.’ (S13.2700)
- c. Bây giờ mình đến với *bài* *học* thứ 4.
 now we come with CL(lesson) study fourth
 ‘Now let us come to the fourth lesson.’ (S15.3894)
- d. em đã có một *bài* *phát biểu* rất xúc động,
 I PAST have one CL speak very thrilling
 ‘I had a very thrilling speech,’ (S6.7606)
- e. sẽ có những *bài* của Tường
 will have PL CL(unit of song) of Tuong
 ‘there will be Tuong’s songs’ (S13.2675)
- f. mà không biết sao mọi người cứ thích em hát *bài* này.
 but not know why every human still like I sing CL(unit of song) this
 ‘but (I) don’t know why everyone still likes to have me sing this song.’ (S13.2691).

The classifier *bài* (song, lesson, text) is frequent since the nouns it precedes are repeatedly used in the corpus due to the content and topic of the talk. Moreover, this classifier can occur with possessives and plural morphemes with the omission of the noun when the noun is identified in the preceding context as in (57e). In this case, because the noun ‘song’ is previously mentioned, *bài* (song, lesson, text) occurs with the possessive Tuong’s (her) and the plural morpheme ‘*những*’ in the absence of the noun. The classifier also goes with demonstratives with the omission of the noun identified in the context as in (57f). When the noun ‘song’ is mentioned in the context, this classifier combines with the demonstrative *này* (this) in the absence of the noun.

The third most frequent classifier *cuộc* (life, strike, match) appears with eleven different nouns or nominals in the corpus. It individuates the nouns such as *cuộc đời* (CL life), *cuộc điện thoại* (CL phone calls), or *cuộc tình* (CL love) as in (58a-c). However, it also combines with verbs, nominalizing and individuating them as *cuộc chơi* (CL game), *cuộc dạo chơi* (CL outing), *cuộc gặp* (CL meeting), *cuộc họp* (CL meeting), *cuộc phỏng vấn* (CL interview), *cuộc trò chuyện* (CL talk), or *cuộc thi* (CL competition) as in (58d-j). Interestingly, this classifier also functions as a nominalizer in Vietnamese.

- (58) a. mỗi người chỉ có một *cuộc đời* thôi
 every person just have one CL life only
 ‘everyone has only one life’ (S2.5058)
- b. mà nhận được 30 *cuộc điện thoại* của các nhà hảo tâm gọi đến
 but receive get 30 CL phone call of PL human generous call to
 ‘but received 30 phone calls from the generous people’ (S7.7937)
- c. sau khi một *cuộc tình* kết thúc
 after one CL love end
 ‘after one/a love ends’ (S15.4211)
- d. cô ấy muốn là mình làm chủ *cuộc chơi* của mình.
 She want that she master CL game of herself
 ‘she wants to master her own game.’ (S14.3842)
- e. với em nó là một *cuộc dạo chơi* hoàn toàn không có chủ định gì cả.
 for me it be one CL outing totally not have intention any all
 ‘for me it is totally an outing without any intention at all.’ (S12.2346)
- f. đây sẽ là *cuộc gặp* định mệnh
 here will be CL meet fateful
 ‘this will be a fateful meeting’ (S5.7339)
- g. *Cuộc họp* không suôn sẻ đến như vậy
 CL meeting not smooth to so
 ‘The meeting was not so smooth’ (S9.8538)
- h. đã thực hiện một *cuộc phỏng vấn* lần thứ hai
 PASS carry out one CL interview time second
 ‘carried out an interview for the second time’ (S6.7778)

i. đang theo dõi *cuộc* trò chuyện với ông ĐDC

PROG watch CL talk with Mr. DDC

‘(are) watching the talk with Mr. DDC’ (S9.8875)

j. đại diện Việt Nam tham dự các *cuộc* thi quốc tế.

represent Vietnam participate PL CL competition international

‘representing Vietnam to participate the international competitions.’ (S20.5933).

In addition, *cuộc* (life, strike) co-occurs with *cái* (inani.) in the doubling construction 25 times in the corpus as in (59). While *cuộc* (life, strike) nominalizes and individuates the noun, *cái* (inani.) emphasizes it. The use of double classifiers will be discussed in section 5.2.3.

(59) để mình biết là à *cái* *cuộc* thi đó, *cuộc* kiểm tra đó,

for me know that ah CL(inani.) CL competition that CL test that

‘for me to know that that competition, that test,’ (S20.5935)

With 129 tokens, *sự* (event) nominalizes and individuates 86 different verbs or adjectival verbs in the Spoken Corpus. As in (60a), the verb ‘practice’ is nominalized and individuated by *sự* (event). The adjectival verbs ‘khác biệt’ (different) and ‘tự tin’ (confident) are also nominalized and individuated by this classifier as in (60b-c).

(60) a. để xem anh đầu tư cho *sự* tập luyện của mình như thế nào

For see him invest for CL(event) practice of self how

‘to see how he invested in his practice’ (S11.2068)

b. Đây là một *sự* khác biệt rất là rõ rệt,

Here be one CL(event) different very be obvious

‘This is an obvious difference,’ (S16.45)

c. Bạn hãy kể về *sự* tự tin của bạn.

you let tell about CL(event) confident of you

‘You please tell about your confidence.’ (S10.1627).

With 65 tokens, *chiếc* (individual) appears with twelve different nouns in this corpus including *chiếc áo* (costume), *chiếc đũa* (chopstick), *chiếc ghế* (chair), *chiếc nón* (bamboo hat), *chiếc thẻ* (card) as in (61a-e). This classifier also occurs with numerals as a pro-form in the absence of the noun when the noun is previously mentioned as in (61f). In this case, the hearer can use the

preceding context to figure out what noun this classifier refers to. In (61f), *chiếc* (individual) appears with the numeral *một* (one) with the omission of the noun, so the hearer must refer to the preceding context to find out what noun is omitted. In this case, *chiếc* (individual) refers to one of the sandals because the noun ‘sandals’ is mentioned at the beginning of the sentence. It is surprising to find that this specific classifier co-occurs with *cái* (inani.) as in (61g). Like other double classifiers, *chiếc* (individual) performs its lexical semantic function of individuating the noun ‘xe Hải Âu’ (Hai Au vehicle) while *cái* (inani.) emphasizes the noun. The nouns that appear with *chiếc* (individual) and *cái chiếc* (inanimate, individual) constructions in the corpus are all concrete nouns, but not abstract nouns. This evidence suggests that *chiếc* (individual) and *cái chiếc* (inani., individual) can go with concrete nouns only, not with abstract nouns.

- (61) a. sẽ khoác lên mình *chiếc* áo cử nhân
 will put on body CL(individual) costume bachelor
 ‘will put on the bachelor’s costume’ (S6.7603)
- b. có một *chiếc* đũa thần
 there one CL(individual) chopstick magic
 ‘there is a magic chopstick’ (S9.8500)
- c. Đây là *chiếc* ghế xứng đáng dành cho em.
 Here be CL(individual) chair deserve set for you
 ‘This is the chair that you deserve.’ (S17.571)
- d. Đây là *chiếc* nón, là món quà của bạn.
 Here be CL(individual) conical hat be CL gift of you
 ‘This is a conical hat, a gift for you.’ (S17.663)
- e. Không phải một người chỉ có một *chiếc* thẻ.
 not right one person just have one CL(individual) card
 ‘It is not that a person just has only one card.’ (S9.8804)
- f. Mua dép thì thường về chỉ đi được một *chiếc*,
 Buy sandals then often back just go get one CL(individual)
 ‘Buying sandals, then (I) just can wear one,’ (S6.7862)
- g. các nghệ sĩ còn đi trên những *cái chiếc* xe Hải Âu đó đấy,
 PL artist also go on PL CL(inani.) CL(individual) vehicle Hai Au that then
 ‘the artists also went on those HaiAu cars then,’ (S11.2193).

The classifier *tình* (relationship) is quite frequent in the corpus although it goes with a limited number of nouns. It usually appears with the verbs *yêu* (love) or *yêu thương* (love, affection) and nominalizes them into the nouns ‘a love’ or ‘love for something’ as in (62). As this noun is used repeatedly in the corpus, the classifier *tình* (relationship) appears quite often despite the limited number of nouns it can classify.

- (62) Anh suy nghĩ gì về một *tình* yêu bị phản đối?
 you think what about one CL(relationship) love PASS deprecate
 ‘What do you think about a deprecated love?’ (S17.754)

Interestingly, *con* (animate) is also one of the most frequent classifiers in the Spoken Corpus as it is in the Narrative and Online Newspaper Corpora. It appears with seven different nouns in this corpus including *con đường* (CL road), *con đường sắt* (railway), *con tàu* (CL ship), *con thuyền* (boat), *con số* (CL figure), *con mắt* (eye), and *con tim* (CL heart) as in (63a-h). These nouns are used repeatedly in the corpus, resulting in the frequent use of this classifier. It goes with the noun ‘đường’ (road, path) with literal and figurative meanings, which indicates a ‘real’ road or railway as in (63c) and a road in figurative meaning like a ‘career road’ as in (63a). It is used to describe a ‘real’ train, *con tàu* (CL train), as in (63d) or to a metaphorical boat *con thuyền* (CL boat) as in (63e). This classifier even appears with parts of a human body like ‘eyes’ or ‘heart’ such as *con mắt* (CL eyes), *con tim* (CL heart) in (63g-h). It also co-occurs with *cái* (inanimate) for emphasizing the noun *cái con đường nghệ thuật* (CL CL road arts) as in (63b). This is a very interesting phenomenon as two general classifiers of the two types, inanimate and animate non-human, combine in the doubling construction. This construction will be further discussed in 5.2.2.

- (63) a. và chả có *con* đường nào trải đầy hoa hồng.
 and not have CL(ani.) road which spread full flower rose
 ‘and no road is full of roses.’ (S12.2389)
 b. em chính thức chọn *cái con* đường nghệ thuật.
 I officially choose CL(inani.) CL(ani.) road arts
 ‘I officially choose the arts road.’ (S20.5606)
 c. Tính *con* đường sắt trước
 plan to build CL(ani.) way rail before
 ‘plan to build the railway before’ (S9.8508)

- d. hay tính *con* tàu trước,
or plan to build CL(ani.) train before
'or plan to build the train before,' (S9.8775)
- e. *con* thuyền của thầy Ngô Mạnh Cường
CL(ani.) boat of teacher Ngo Manh Cuong
'the boat of Teacher Ngo Manh Cuong' (S7.8076)
- f. tôi lấy *con* số của Đài Loan.
I take CL(ani.) figure of Taiwan
'I'll take the figure of Taiwan.' (S9.8744)
- g. bằng *con* mắt của đứa trẻ thì năm nay mình hòa đồng hơn với các con.
through CL(ani.) eye of CL child then year this I sociable more with PL children
'through the eyes of a child, I am more sociable with the children this year' (S20.5760)
- h. ông đã chiếm được *con* tim của một triệu thính giả.
He PAST gain get CL(ani.) heart of one million audience
'he gained the heart of a million audience.' (S9.8860).

The ninth most frequent classifier in the Spoken Corpus *đám* (procession) appears quite often even though it just classifies and individuates the two nouns *cưới* (wedding) and *mây* (cloud/Icloud) as in (64a) and (64c). It also co-occurs with *cái* (inanimate) in the classifier doubling construction *cái đám* (inani., procession) as in (64b). In this construction, the classifier *đám* (procession) classifies and individuates the noun 'wedding' while *cái* (inanimate) emphasizes the noun. These nouns are repeatedly used in the corpus due to the topic of the talk, so the classifier appears quite often. However, in the Narrative and Online Newspaper Corpora, this classifier also combines with a number of other nouns.

- (64) a. Chúng ta sẽ gặp một người vừa đi *đám* cưới về.
we will meet one person just go CL(procession) wedding back
'We will meet a person who has just been back from the wedding ceremony.' (S16.402)
- b. tôi thấy những *cái* *đám* cưới như vậy người ta lại nhớ hơn
I find PL CL(inani.) CL(procession) wedding such people again miss more
'I find that they miss such weddings more' (S16.209)

- c. nó đẩy được lên đám mây hết
 it push get up CL(patch) cloud all
 ‘it is all put in iCloud’ (S21.6148).

In summary, among the nine most frequent classifiers in the Spoken Corpus analysed above, *cái* (inanimate) is the most frequent with nearly two thirds of the tokens found in the corpus. Although *cái* (inanimate) is frequent across the three genres, it is especially frequent in the spoken corpus, compared to the other classifiers within the same corpus as well as *cái* (inanimate) in the other two written corpora. It is noticeable that *cuộc* (life, strike, match) and *sự* (event) are also quite frequent in the spoken corpus, which is similar as the online newspaper. It appears that these two genres have the two frequent classifier types in common, which are different from the narratives. The distribution of infrequent classifiers in this corpus is reported in the next section.

4.4.3 Overall distribution of infrequent classifiers in the Spoken Corpus

Besides the most frequent classifiers, the remaining 125 classifiers appear far less often in the Spoken Corpus while a large number of specific classifiers are rarely used. In fact, these classifiers can go with only one or two nouns which rarely appear in the corpus, resulting in a small number of occurrences. For instance, some infrequent classifiers including *điếu* (cigarette), *điệu* (dance), and *vũng* (puddle) occur only once or twice each in the corpus. The classifier *điếu* (cigarette) goes with the noun *thuốc* (cigarette) as in (65a). The classifier *điệu* (dance) nominalizes the verb *nhảy* (dance) and individuates the nominal *một điệu nhảy* (a dance) as in (65b). The classifier *vũng* (puddle) classifies and individuates the noun *nước* (water) as in (65c).

- (65) a. được cuộn lại như *điếu* thuốc
 PASS roll in like CL cigarette
 ‘(was) rolled like a cigarette’ (S9.8831)
- b. Xin cảm ơn một *điệu* nhảy
 Please thank one CL(dance) dance
 ‘Thank you for a dance’ (S10.1470)
- c. nhưng mà khi rút ra khỏi *vũng* nước
 but then when get out of CL(puddle) water
 ‘but then when got out of the puddle’ (S11.1997).

In brief, the distribution of inanimate classifiers in the Spoken Corpus varies greatly since a number of classifiers can go with a great variety of nouns while other classifiers can combine with only one or two nouns. This results in the variety and diversity of Vietnamese inanimate classifiers. In this corpus, 134 actual inanimate classifiers are found, including 86 single 48 double classifiers, excluding 44 exceptional cases, which will be discussed in the next section.

4.4.4 Exceptional cases in the Spoken Corpus

Prior researchers including H. T. Nguyen (2013) claim that the construction of *cái cái* (inanimate) never occurs in Vietnamese, and it has never been analysed in previous studies to date. However, this construction appears 22 times in the Spoken Corpus as in (66a) and (66c). There are two ways to view this construction. In the first view, *cái cái* (inani.) is analyzed as a ‘fused’ construction such as *cái sự* (inani., event). This construction appears similar as *cái sự* (inani., event), in which *sự* (event) has its own function of nominalization and individualization while *cái* (inani.) is used as an emphatic as claimed by H. T. Nguyen (2004, 2013) and Simpson and Ngo (2018). From this view, the *cái* (inani.) positioning right before the noun/nominal performs its own function of individualization and classification as usual. The other *cái* (inani.) may be treated as an emphatic as in *cái sự* (inani., event) and other double classifiers. This view sounds plausible. However, when examining all the 22 tokens of *cái cái* (inani.) construction, only five of them can be interpreted in this way as in the example in (66c). In (66c), the nominal ‘suy nghĩ’ (thinking) is nominalized and individuated by the *cái* (inani.) positioning right before the nominal while the other *cái* (inani.) functions as an emphatic. In fact, the remaining 17 tokens of *cái cái* (inani.) construction cannot be interpreted in this way because they appear with unclassified nouns such as *sở thích* (interest) in (66a-b). This noun can follow the plural morpheme ‘những’ without a classifier as in (66b). That means these types of nouns do not require a classifier to be classified and individuated. In this case, *cái* (inani.) as an emphatic may appear with these nouns optionally. Thus, *cái cái* (inani.) construction in this case should be interpreted as in the second view.

In the second view, *cái cái* (inani.) construction is analyzed differently from other doubling constructions. In this view, the *cái* (inani.) positioning right before the noun/nominal performs its own function of an emphatic while the other *cái* (inani.) form can be treated as a filler. The noun *sở thích* (interest) is an unclassified noun in Vietnamese because it can appear without a classifier as in (66b). In (66a), the *cái* (inani.) before the noun *sở thích* (interest) performs its own function

of an emphatic while the other *cái* (inani.) form can be treated as a filler. I assume that the other *cái* (inani.) form occurs in this case due to speech production when the speaker is trying to find the appropriate word for their expressions.

(66) a. đó không phải là *cái cái* sở thích của mình.

that not right be CAI CL(inani.) taste of me

‘That is not my taste.’ (S22.994)

b. Bom có những sở thích cực kì buồn cười.

Bom have PL taste extremely funny

‘Bom has extremely funny tastes.’ (S3.6710)

c. đủ nhận thức thì *cái cái* suy nghĩ

enough knowledge then CAI CL(inani.) thinking

‘with enough knowledge, then the thinking/thought’ (S20.5915)

It is similar as other cases in which several *cái* (inanimate) forms is repeated before the noun/nominal. For classified nouns, the *cái* (inanimate) positioning right before the noun would perform the function of classification, individualization and/or nominalization while one *cái* (inanimate) form functions as an emphatic and the additional *cái* (inanimate) forms as fillers. For unclassified nouns, the *cái* (inanimate) positioning right before the noun performs the function of an emphatic while the additional *cái* (inanimate) forms function as fillers. It is found that thirteen cases in which *cái cái cái* (inanimate) occur with nouns in the corpus as in (67). Similarly, three cases in which *cái cái cái cái* (inanimate) occur in the corpus as in (68). Also, one case *cái cái cái cái cái* (inanimate) and one case *cái cái cái cái cái cái* (inanimate) occur in the corpus. I believe that the extra *cái* (inanimate) occurs in these cases are just repetitions or ‘fillers’ when the speaker is thinking of the appropriate word for their expressions. These extra *cái* (inanimate) cannot be treated as double or triple classifiers as they do not carry any semantic functions in the noun phrases. In (67-68), one *cái* (inanimate) is functioning as a classifier to classify and individuate the noun khác biệt (difference) and lịch trình (agenda) respectively while all the other extra *cái* (inanimate) are considered as repetitions.

(67) Đây, *cái cái cái* khác biệt này tôi cho rằng là không nên

That CAI CAI CL(inani.) difference this I suppose that be not should

‘That, this difference, I suppose, should not’ (S16.59)

(68) Tuy nhiên là *cái cái cái cái* lịch trình thì rất là dày đặc

However, be CAI CAI CAI CL(inani.) agenda then very be tight

‘However, the agenda is very tight’ (S20.5729).

In addition, there is one case in which *cái cái cái* (inanimate) co-occurs with the classifier *sự* (event). In this case, the classifier *sự* (event) co-occurs with one classifier *cái* (inanimate) to make a perfect double classifier *cái sự* (inanimate, event), which is quite common in the Spoken Corpus. In this double classifier, *sự* (event) nominalizes the adjective verb *đồng điệu* (harmonious) while *cái* (inanimate) emphasizes the noun as well as the definiteness of the noun as in (69). That means the other two extra *cái cái* in this case might be repetitions or ‘filler’.

(69) Em muốn là nó sẽ lan tỏa được *cái cái cái* *sự* đồng điệu ở đó qua nhiều cách

I want be it will spread get CAI CAI CL(inani.) CL harmony in that via many way

‘I want it spread the harmony via many ways’ (S20.5581)

These constructions sound unusual and may be viewed as exceptional cases, especially when three or more *cái* (inani.) forms co-occur. In cases where additional forms of *cái* (inani.) are repeated in a sequence, I assume that when they appear with a classified noun, one *cái* (inani.) retains its grammatical/semantic property and one *cái* (inani.) performs as an emphatic while the additional *cái* forms function as fillers, perhaps to regulate discourse. When additional *cái* (inani.) forms appear with an unclassified noun, one *cái* (inani.) performs as an emphatic while all the additional *cái* forms may function as fillers. The case in which additional forms of *cái* (inani.) are repeated in a sequence reinforces the view that they do not function as prototypical classifiers, but instead fillers – perhaps to regulate discourse. Moreover, as these cases just appear in the spoken corpus, I assume that additional *cái* (inani.) forms may be repeated unintentionally as fillers, perhaps to regulate discourse. It sounds plausible because the repetition of some other function words is also found as probably fillers as in the examples given below. It is important to assign these additional *cái* (inani.) forms a function at the level of interactive discourse. That is why I just temporarily put CAI instead of assigning them a function at this step.

Another exception is a case in which *cái cái cái* (inani.) co-occurs with the classifier *niềm* (sentiment) followed by the double classifier *cái sự*, then the noun *thôi thúc* (urge) as shown in (70). In this case, I assume that the first two extra *cái cái* before *cái niềm* (inani., sentiment) are

just ‘fillers’, then the speaker finds that the double classifier *cái niềm* (inani., sentiment) is not what he/she wants to say because it is not the appropriate one for the noun. The double classifier *cái sự* (inani., event) is thus used instead as it is the right choice for the noun. It seems that the speaker is thinking while speaking and begins the sentence without really knowing what they will use with it. That is why a number of extra function words are repeatedly used before the actual words uttered.

(70) thì *cái cái cái* *niềm cái* *sự* thôi thúc của em nó lại càng nhiều hơn
 then CAI CAI CAI CL CL(inani.) CL urge of self it again more many more
 ‘then my urge is increasing more’ (S20.5965)

One last exceptional case is that in the double classifier *cái viên* (inani., small, round), *viên* (small, round) is repeated before the noun is uttered as in (71). This might be because the speaker is hesitating while choosing the appropriate noun for their expression, so he/she uttered the specific classifier twice unintentionally. In this case, I assume that the speaker wants to use the double classifier *cái viên* (small, round) instead of *cái viên viên* (small, round) as in (71).

(71) chú Tuấn cũng nói là *cái viên viên* gạch đầu tiên nó dễ lắm
 Mr. Tuan also say that CL(inani.) CL(brick) CL brick first that easy very
 ‘Mr. Tuan also says that the first brick is very easy’ (S20.5615)

These are the exceptional cases in the Spoken Corpus. I do not know which classifier group they should be put into. Thus, I just want to mention them, but I am not going to include them in any other groups of classifiers in the study. The combination of *cái cái* (inanimate) or *cái cái cái* (inanimate) is not found in the other written discourse corpora. I argue that this is repetition since this phenomenon has been found with other function words as in (72), but not with content words. In (72a), apart from the repetition of *cái* three times, the word ‘hơi’ (rather) is repeated three times in the same utterance. Similarly, in (72b), the numeral ‘nhiều’ (much) is repeated twice before the repetition of *cái*. Thus, the repetition of function words in oral speech is more likely to be due to the speaker hesitation. They are assumed to be treated as ‘fillers’ in the oral speech.

(72) a. Nó là *cái cái cái* *hơi* *hơi* *hơi* hoài nghi.
 It be CAI CAI CL(inani.) rather rather rather suspicious
 ‘It is the rather rather rather suspicious.’ (S20.5689)

- b. Mọi người đặt cho mình nhiều nhiều *cái cái cái* kì vọng.
 Everyone impose for me much much CAI CAI CL(inani.) ambition
 ‘Everyone imposes much ambition onto me.’ (S20.5722)

These exceptions sound very interesting, but these will be left for future study. In sum, I have reported and analysed the major findings from the Spoken Corpus. I will summarize the main findings from the three corpora in the next section.

4.5 Summary

In brief, the chapter has reported the primary findings of the study mainly regarding frequency and distribution of inanimate classifiers in the three corpora. Clearly, the study identifies variation in the frequency and overall distribution of classifiers across the three genres. It is interesting that the classifier frequency in the spoken corpus is much higher than the rates in the other two written corpora. More interestingly, the results show that the classifier frequency declines among younger speakers compared to older and middle-aged groups of speakers in the Spoken Corpus. The differences in classifier frequency across the three genres as well as among the three age groups of speakers in the spoken corpus will be discussed in section 5.1.1.

Moreover, the findings show differences in the overall distribution of classifiers across the three genres. The distribution of different inanimate classifiers varies greatly within the genre as well as across the three genres. As one of the most frequent inanimate classifiers in the three corpora, *cái* (inanimate) is the most frequent in the narrative and spoken corpora, but not the most frequent in the online newspaper. It is especially frequently used with almost two thirds of all the tokens found in the Spoken Corpus. As it appears with a variety of different word classes in various structures, the usage of *cái* (inanimate) will be further explored and discussed in section 5.3. Different from the narratives, both *sự* (event) and *cuộc* (life, strike) appear often in the online newspaper and spoken corpora. They both perform the function of a nominalizer in Vietnamese. In addition to these two classifiers, the data reveals a number of other inanimate classifiers with the nominalization function used in the online newspaper and spoken corpora. This group of classifiers will be explored and discussed in section 5.4 in chapter 5.

Furthermore, double classifiers appear quite often and in a large number in the Spoken Corpus while they are rarely used in the written corpora. This construction is special, which will

be discussed in section 5.2.2 in chapter 5. The results of the study also reveal variation in the number of classifiers and classifier constructions used across the three genres. These findings will be analysed and discussed in chapter 5.

Chapter 5

Discussion

In this chapter, the use of inanimate classifiers in the three corpora will be compared and discussed. An overview of Vietnamese inanimate classifiers in the study is discussed in section 5.1. Section 5.2 looks at the differences in classifier use in the spoken corpus versus the written corpora. As one of the most noticeable differences in spoken and written discourse, the use of double classifiers and their lexical semantic functions are analysed in 5.2.2. Then section 5.3 analyses and discusses the use of *cái* (inanimate) since it is very frequent, especially in the Spoken Corpus. Section 5.4 is an analysis on the classifiers functioning as nominalizers. Finally, section 5.5 discusses Vietnamese classifier constructions and the definiteness of the noun.

5.1 Overview of inanimate classifier use in the corpora

The results of this study show the differences in the use of inanimate classifiers in the three corpora with respect to frequency, distribution, and number of actual classifiers. This section will begin with discussions on the frequency of classifier use in 5.1.1. Then the overall distribution of classifiers is discussed in 5.1.2, and the number of actual classifiers in 5.1.3. Section 5.1.4 summarizes the main differences in classifier use in the three corpora of this study.

5.1.1 Frequency of inanimate classifier use

5.1.1.1 Frequency of classifier use in the three corpora

The data shows that there are differences in frequency of classifiers across the three genres and among different age groups in the spoken corpus. The frequency of classifiers across the genres is discussed in this section, and the next section will discuss the classifier frequency among different age groups. The rates of classifier use in the Narrative, Online Newspaper, and Spoken Corpora are 160, 182, and 286 classifiers per 10,000 words respectively, as shown in Figure 1. The discrepancy in the classifier frequency in the Narrative and Online Newspaper corpora is small although the data shows a big difference between the spoken corpus and these two written discourse corpora. Specifically, the classifier frequency in the spoken corpus is far higher than that in the two written discourse corpora.

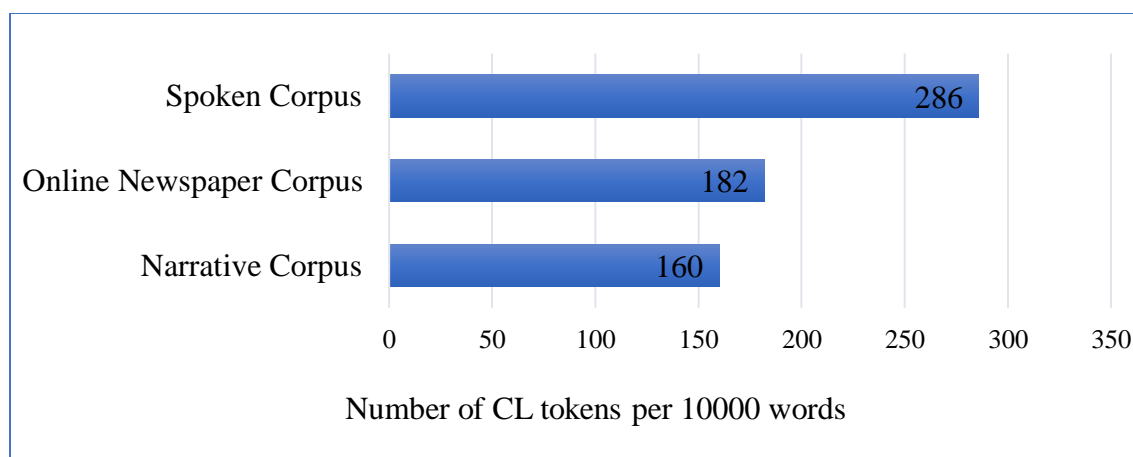


Figure 1: Frequency of classifier use in the three corpora

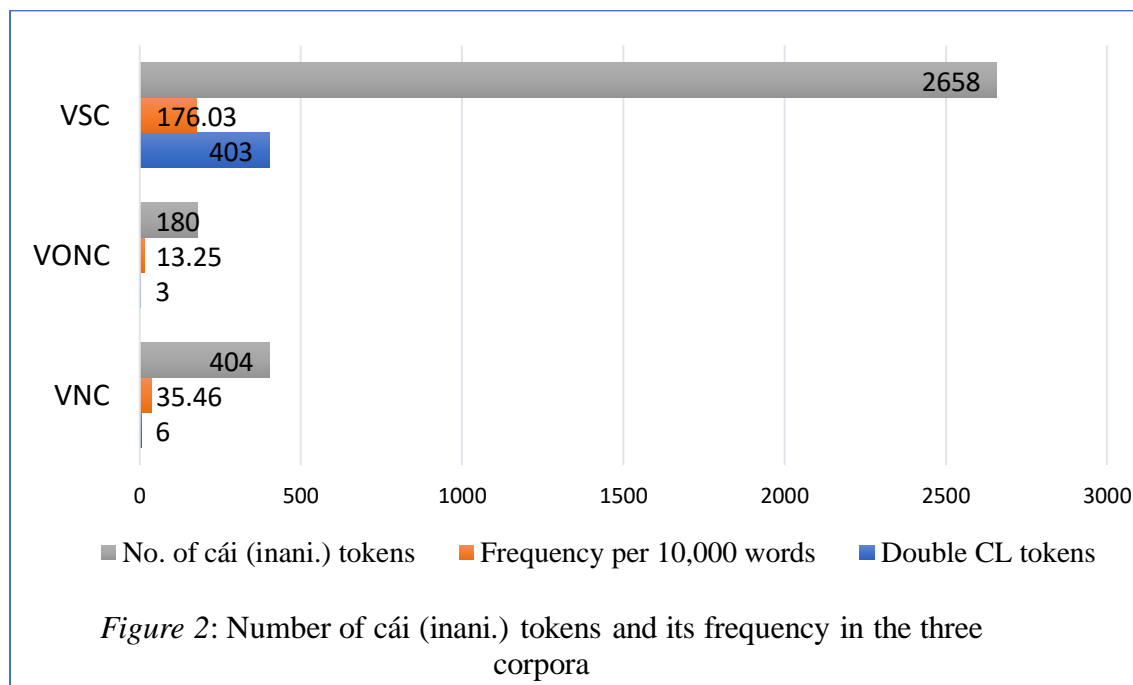
The differences in frequency of classifiers across the three genres can be attributed to several reasons. The first possible reason might be the genre effects because different genres have different characteristics and are used for various purposes of communication. Based on the framework for situational analysis discussed in Biber and Conrad (2009:40), from a situational perspective, the three genres differ in several key respects including participants, relations among participants, channel, production circumstances, setting, communicative purposes, and topics. With respect to participants, narratives and newspapers are alike, having an author addressing an un-enumerated number of readers with no interaction or personal relations, while in the spoken corpus, all participants are speakers addressing one or more addressees with active interaction and probably personal relations. Thanks to the same written channel, the text in both narratives and newspapers has been carefully planned, revised, and edited, while the discourse in the talk shows may have been planned but cannot be revised or edited due to the spoken channel. Despite the unknown setting of both narratives and newspapers, the folktales written over sixty years ago appear to be historical and conventional, while the newspaper articles are current and contemporary because they have been e-published for the last two years. For the talk shows, the setting is different, in which all participants are physically together in a studio to discuss events, thoughts, or opinions about certain topics in the immediate context. The face-to-face conversation in the talk shows requires direct interaction between two or more people who are together at the same place. This could explain why the appearance of multiple *cái* (inanimate) forms in 44 exceptional cases in the spoken corpus analysed in section 4.4.4, while this phenomenon has never occurred in the other two written corpora. For the communicative purposes, these three genres all

convey information though the topics and the content of the text and discourse differ. The audiences of these genres are also different. The main audience of the narrative are children, while the major audience of the newspaper and the talk shows are adults. The situational differences of the genres in these respects explain the variation in frequency of classifiers and in distribution of classifiers among the three genres which is discussed in section 5.1.2.

As the narrative and online newspaper genres are similar in respects including participants, relations among participants, channel, production circumstances, the difference in classifier frequency in these two genres is small. In contrast, the discrepancy in classifier frequency in the spoken and the two written discourse genres is substantial. The higher classifier frequency in the conversations over written texts can be attributed to the genre effects and the content of the texts or discourse. Since classifiers are required for classified nouns with referents in Vietnamese to classify and individuate the nouns/nominals (Emeneau 1951; D. H. Nguyen 1957), the appropriate classifier is used with whatever classified nouns appear in the texts or discourse. Thus, the appearance of a classifier depends on the occurrence of the nouns. For communicative purposes, the texts and discourse in these three genres convey different kinds of messages within various topics. This means, different kinds of nouns are used to convey the information of the folktales, the newspaper articles, and the talk shows. Therefore, the use of inanimate classifiers depends on the appearance of the nouns in the texts or discourse. The differences in classifier frequency across the three genres are due to the genre effects and the content of the texts and discourse.

Due to the differences in characteristics of various genres, the distribution of classifiers is used differently. Specifically, *cái* (inanimate) is very frequent in the spoken corpus, while it is less frequent in the narrative, especially in the newspaper corpus. Figure 2 shows that 2658 *cái* (inani.) tokens have been identified in the spoken corpus, while 404 *cái* (inani.) tokens were found in the narrative and 180 *cái* (inani.) tokens in the newspaper corpus. Comparing the frequency of *cái* (inani.), the spoken corpus has the rate of 176 *cái* (inani.) tokens per 10,000 words, while the rates in the narrative and online newspaper corpora are 35 and 13 per 10,000 words respectively. Similarly, double classifiers in which *cái* (inani.) is constructed with a specific classifier combining with classified nouns appear quite often in the spoken corpus, but rarely in the other two written corpora. Specifically, we identified 403 double classifier tokens in the spoken corpus, while only

three and six double classifier tokens were found in the online newspaper and narrative corpora respectively.



Noticeably, the data reveals that 1804 out of 2658 *cái* (inanimate) tokens in the Spoken Corpus appear with many different non-classified nouns or optional-classifier nouns functioning as emphatics, as analysed in previous research (D. H. Nguyen 1957; H. T. Nguyen 2004, 2013; Simpson and Ngo 2018). However, *cái* (inanimate) does not appear with non-classified nouns with this function in the other two written corpora. Because the use of *cái* (inanimate) in this case is not obligatory, its use totally depends on the speaker's choice and intention. The frequent use of *cái* (inanimate) with non-classified nouns as emphatics in spoken Vietnamese, but not in written language, may lead to the higher classifier frequency in conversations over written texts. Furthermore, it has been observed that *cái* (inanimate) used as emphatics in this case always has a phonological stress. This stress could be considered as one of the "paralinguistic devices" for emphasis as analysed by Biber and Conrad (2009:86). Vietnamese speakers seem to make use of this paralinguistic device when using *cái* (inanimate) to emphasize the noun in order to attract the listener's attention and/or focus on certain things in their speech. This paralinguistic device cannot be used in writing since the author can use "typographic devices" such as bold face, underscoring, or capital letters to indicate emphasis instead (Biber and Conrad 2009:86). This makes sense when *cái* (inanimate) is not used as emphatics in the written corpora of the study. As the use of *cái*

(inanimate) in this case is dependent on the speaker's choice, the differences in classifier frequency across genres may also be due to individual speakers. I argue that Vietnamese speakers use *cái* (inanimate) with non-classified nouns for emphasis to attract the listener's attention and/or focus on certain things in their speech, but not in writing. The use of *cái* (inanimate) will be further explored and discussed in section 5.3.

As three different genres and time periods were chosen for investigation in this study, not only differences in language use across genres but also language change over time may be observed. The folktales which were written over 60 years ago reveal the more conventional language use, while the concurrent newspapers and talk shows reflect the contemporary language use. Thus, we can observe language change at work with the corpora of this study. However, there is just a small discrepancy between the classifier frequency in the narrative and newspaper corpora although the folktales and the newspapers are over sixty years apart. Thus, I assume that the differences in classifier frequency in these corpora are mainly due to the genre effects rather than language change over time generally. The frequent use of *cái* (inanimate) as an emphatic - a novel function can be seen as some kind of language change in Vietnamese as this new function is frequently used in spoken language, but not in written language. This use of *cái* (inanimate) may also indicate that the formality of genres might affect the frequency and choice of classifiers. Although the newspaper and spoken corpora are contemporary, *cái* (inanimate) is very frequent in the spoken, but far less frequent in the newspaper corpus. As it has been observed, the distribution of classifiers in the corpora of this study reveals more evidence for language change in progress in Vietnamese, which is discussed in section 5.1.2. It is noted that the data also shows differences in the classifier frequency among speakers of different age groups in the Spoken Corpus, which is discussed in the next section.

5.1.1.2 Frequency of classifier use among different age groups

Interestingly, the results show the difference in frequency of classifier use among three different age groups in the Spoken Corpus including older speakers of over 50 years old, born in 1968 or earlier, middle-aged speakers aged between 30 and 50, born in 1969 to 1987, and younger speakers under 30 years old, born in 1988 or later. As mentioned in the methodology, in this study, I only analyze data in the spoken corpus within the variationist framework as the age of the speakers in the talk shows, who are well-known by Vietnamese community, could be identified in

social media, while the age of the writers of the folktales and newspapers could not be found. Figure 3 shows the frequency of classifier use among these age cohorts with the totals of tokens. The totals of classifier tokens among these groups of speakers are different, but it is not important to compare them because of the different word counts in the speech of these groups. The classifier frequencies among older, middle-aged, and younger speakers are 304, 291, and 269 per 10,000 words respectively. This shows that the use of classifiers among these age groups follows the pattern that the older the speakers are, the more classifiers appear in their speech. Although the difference in the frequency of classifier use among these three groups is not very big, it clearly shows the tendency that younger speakers may use classifiers less often than older speakers do in spoken Vietnamese.

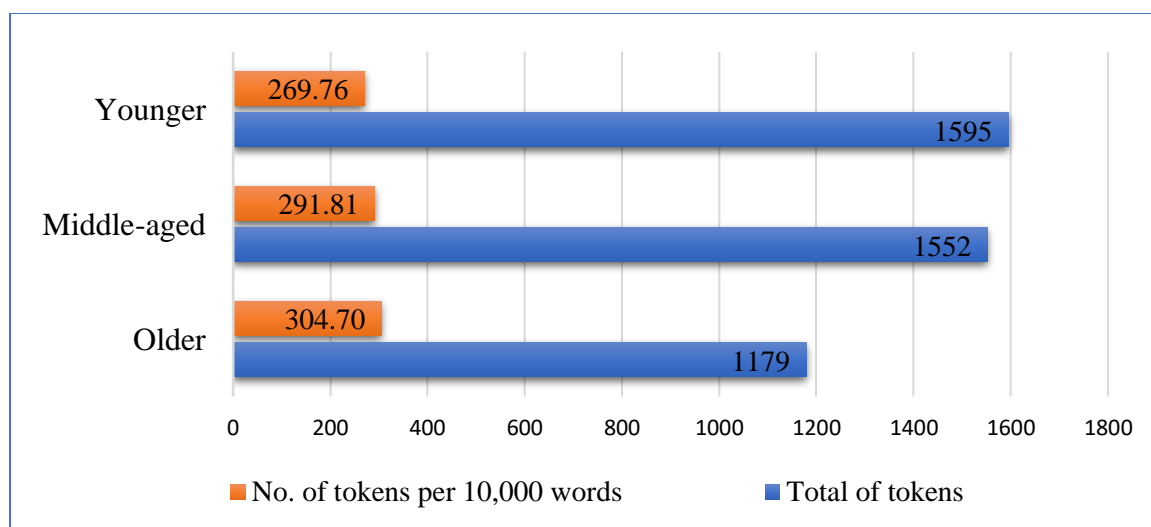
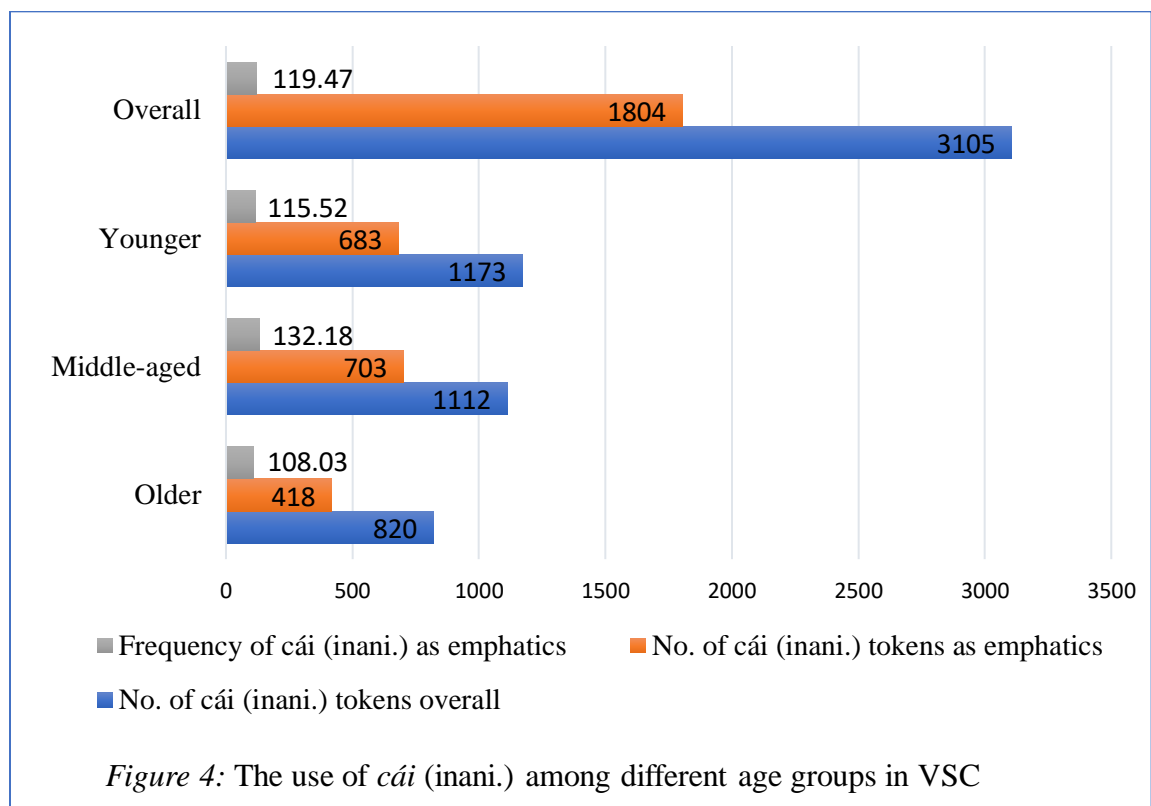


Figure 3: Frequency of CL use among different age groups in VSC

This is a very interesting finding because the data shows a decline in classifier use by younger age group despite a significant increase overall in classifier use in the Spoken Corpus compared to the other two written corpora. However, as explained in the previous section, the classifier frequency differs across genres probably due to the genre effects, the content of the texts and discourse, and individual effects. Specifically, the use of *cái* (inanimate) with non-classified nouns as emphatics in the spoken corpus may lead to the higher classifier frequency in conversations over written texts as its use is not obligatory.

In contrast, the data of the study shows that the use of *cái* (inanimate) as emphatics with non-classified nouns among the three different age groups does not affect the classifier use pattern

among them. Figure 4 shows the totals of *cái* (inanimate) tokens and the number of *cái* (inanimate) tokens used as emphatics among the different age groups in the spoken corpus as well as its frequency per 10,000 words. The number of *cái* (inanimate) tokens in this figure consists of *cái* (inanimate) tokens as a free-standing classifier and as a part of double classifiers including 44 cases of multiple *cái* (inanimate) forms. As Figure 4 shows, the frequencies of *cái* (inanimate) used as emphatics among middle-aged and younger speakers, 132 and 115 per 10,000 words respectively, are a little higher than that among older speakers, 108 per 10,000 words. This data does not support the hypothesis that the use of *cái* (inanimate) with non-classified nouns as emphatics may be a possible reason for the decline in classifier use among younger age group compared to older age group although it appears to be a reason for the higher classifier frequency in spoken Vietnamese compared to written language. This also means the use of other classifiers among younger speakers may decline.



In Vietnamese, classified nouns with specific referents do require a classifier. However, the data shows some evidence that among younger speakers in the spoken corpus, some classified nouns sometimes do not appear with a classifier as they are supposed to. As shown in the examples

in (73), the classified nouns ‘áo’ (sweater) in (73a-b) and ‘ghế’ (chair) in (73c-d), which have their own specific referents, do not occur with a classifier they require. As analysed in section 4.4.2 and also in previous research (Emeneau 1951; Thompson 1965), these two nouns usually require *cái* (inanimate) or *chiếc* (individual) as in (61a) and (61c) repeated here for comparison. This means, in (73), younger speakers dropped the classifier required for the classified nouns, and this phenomenon takes place at times among younger speakers in the spoken corpus.

- (73) a. bảo là Nguyễn Anh cởi áo ra cho Tường.
 tell that Nguyen Anh take sweater off for Tuong
 ‘(they) told Nguyen Anh (me) to take off my sweater for Tuong’ (S13.2649)
- b. áo của bạn sẽ không ấm bằng của Tường
 sweater of you will not warm as of Tuong
 ‘your sweater will not be as warm as Tuong’s (mine)’ (S13.2650)
- c. Anh thích ngồi ghế nào?
 You like sit chair which
 ‘Which chair do you like to sit on?’ (S17.568)
- d. thì em ngồi ghế còn lại.
 then I sit chair remain
 ‘then I will sit on the remaining chair.’ (S17.569)
- (61) a. sẽ khoác lên mình chiếc áo cử nhân
 will put on body CL(individual) costume bachelor
 ‘will put on the bachelor’s costume’ (S6.7603)
- c. Đây là chiếc ghế xứng đáng dành cho em.
 Here be CL(individual) chair deserve set for you
 ‘This is the chair that you deserve.’ (S17.571).

The omission of a constituent in language use in general and a classifier in this case can be considered as “language simplification” (Honeyfield 1977:431). In the examples in (73), the speakers simply removed the classifier in their speech when this omission does not affect the meaning of the noun phrase or obstruct the listener’s understanding. It takes place within the noun phrase although it looks like “morphological simplification” (Atkinson et. al. 2018:2818). It seems that in this case, the speakers want to simplify their language use by omitting the classifier for the

purpose of shortening the noun phrase for faster speech production. I assume that the lower classifier frequency among younger speakers might be because they want to simplify language use by removing the classifier in their speech when it is possible. That means, the speaker may reduce classifiers in their speech whenever the reduction does not impact the overall coherence of the text. In my view, there are no internal motivations for this hypothesis. This might be the case of individual-level simplification in conversation due to the advantage of simplification such as “gains of speed” as Chandrasekar et. al. (1998:1044) suggest. Thus, the lower classifier frequency among younger speakers compared to older speakers indicates that there may be language change in progress in spoken Vietnamese. To be specific, linguistic simplification may be some change in language use regarding classifiers in spoken language in progress via apparent time. However, this issue would need further investigation in a larger scale corpus study with more speakers as this hypothesis could not be tested in the data of this study.

Due to the nature of the talk shows, the question that patterns of classifier use might be shaped by dyadic relationships could be raised. In fact, the speakers in the spoken corpus differ by age, but they sound quite close to the other participants in the talk shows. Most of the participants in the talk shows know each other, so generally they are quite friendly to each other. There seems no difference in the dyadic relationship among the speakers in the talk shows despite their age discrepancy. Thus, dyadic relationship is not assumed to have impacts on the frequency of classifiers in their speech. Also, this means that the formality of these talk show episodes appears to be similar. Therefore, the difference in classifier frequency among the three age groups is not related to the formality level of the discourse although it might be a factor that affects the choice of classifiers across different genres. This will be discussed in section 5.1.2.

In sum, the decline in classifier frequency among younger speakers compared to older speakers may be due to the content of the discourse, individual effects, and language simplification. However, the data of this study is not enough for testing this hypothesis. In fact, the difference in classifier frequency among the different age groups of speakers is not as substantial as the discrepancy across the three genres, especially between the spoken and the two written corpora. With respect to frequency, there is not enough data for claiming language change over time although language change in progress regarding distribution of inanimate classifiers has been observed in the corpora of this study. This will be discussed in the next section.

5.1.2 Overall distribution of classifiers in the three corpora

The data shows differences in the overall distribution of classifiers in the three corpora. As shown in Table 3 in section 4.1.2, repeated here for easy reference, most of the frequent inanimate classifiers are differently distributed within the genres and across the different genres. It is noticeable that as one of the most common classifiers, *cái* (inanimate) is very frequently used in the Spoken Corpus. Interestingly, *cái* (inanimate) is constructed with a specific classifier to generate a double classifier. Specifically, the frequent co-occurrence of *cái* (inanimate) and *sự* (event) makes the double classifier *cái sự* (inanimate, event) become one of the most frequent classifiers with 144 tokens in the Spoken Corpus. The study has found that six of the ten most frequent classifiers in the Narrative Corpus overlap with those among the ten ‘core’ classifiers claimed by Löbel (2000) including *cái* (inanimate), *cây* (tree, long object), *quả* (fruit, round), *chiếc* (individual), *hòn* (round), and *con* (animate). It seems that the finding of the most frequent inanimate classifiers in the Narrative data is close to the findings of previous research. This sounds reasonable because the data in previous studies are primarily based on narratives and constructed or elicited utterances. On the contrary, the online newspaper and spoken corpora have only three most frequent classifiers in common with the narrative data and Löbel (2000)’s core classifiers including *cái* (inanimate), *chiếc* (individual), and *con* (animate) although their distribution differs across the three genres. The Spoken Corpus has 2658 tokens of *cái* (inanimate) as a single classifier, excluding the number of its tokens in double classifiers and multiple *cái* (inanimate) forms, while the Narrative Corpus has 404 *cái* (inanimate) tokens, and the Online Newspaper Corpus has 180 *cái* (inanimate) tokens. This means that the distribution of *cái* (inanimate) greatly differs across the three genres. This difference is the most noticeable in the uses of classifiers in the three corpora of the study, which is examined and discussed extensively in section 5.3.

Similar as *cái* (inanimate), the overall distribution of *chiếc* (individual) varies across the three genres as shown in Table 3. With 268 tokens altogether in the three corpora, it appears with 90 different inanimate nouns including common nouns as well as proper nouns indicating vehicles or airplanes specifically, such as *chiếc Mazda* (Mazda), *chiếc Boeing* (Boeing). It also goes with some English words such as ‘smartphone’ and ‘Ipad’. However, all the nouns that *chiếc* (individual) appears with in the corpora are concrete nouns, not an abstract noun. This evidence

supports the argument made by Tran (2018) that the classifier *chiếc* (individual) can combine with a wide variety of concrete nouns, but not with abstract nouns.

Table 3: Overall distribution of frequent CLs in the three corpora

Narrative Corpus (115,000 words)		Online Newspaper Corpus (135,900 words)		Spoken Corpus (151,000 words)	
CLs	No. of tokens	CLs	No. of tokens	CLs	No. of tokens
<i>cái</i> (inanimate)	404	<i>sự</i> (event)	277	<i>cái</i> (inanimate)	2658
<i>cây</i> (tree, long)	180	<i>cuộc</i> (life, strike)	187	<i>bài</i> (song, lesson)	204
<i>quả</i> (fruit, round)	66	<i>cái</i> (inanimate)	180	<i>cuộc</i> (strike, life)	201
<i>chiếc</i> (individual)	59	<i>chiếc</i> (individual)	144	<i>cái sự</i> (inani., event)	144
<i>bờ</i> (bank, shore)	46	<i>việc</i> (activity)	116	<i>sự</i> (event)	129
<i>gốc</i> (root)	45	<i>bộ</i> (set)	68	<i>chiếc</i> (individual)	65
<i>thứ</i> (type)	40	<i>vụ</i> (catastrophe)	61	<i>tình</i> (relationship)	61
<i>hòn</i> (round)	38	<i>con</i> (animate)	58	<i>con</i> (animate)	58
<i>con</i> (animate)	36	<i>dòng</i> (river, line)	57	<i>đám</i> (procession)	47
‘others’ (<2% each)	913	‘others’ (<2% each)	1324	‘others’ (<2% each)	759
Total	1828	Total	2472	Total	4326

Interestingly, the classifier *con* (animate) appears with nouns indicating non-living things quite often in the three corpora. It goes with 30 different nouns in the three corpora including nouns indicating roads/paths such as *con đường* (road) either with literal or connotational meaning, *con lộ* (road), *con ngõ* (lane), *con phố* (street), *con dốc* (slope), *con đường sắt* (railway) and indicating rivers or the like such as *con sông* (river), *con suối* (stream), *con mương* (ditch), *con sóng* (wave). These nouns also indicate vehicles such as *con thuyền* (boat), *con tàu* (ship), *con xe* (car); or tools such as *con dao* (knife), *con rựa* (machete), *con kiếm* (sword), *con ác* (a part of loom), *con thoi* (shuttle); and even nouns indicating parts of a human body such as *con mắt* (eye) or *con tim* (heart). So far, it is not clear why the general classifier *con* (animate) can go with these inanimate nouns. However, it is beyond the focus of this study and would be left for future research.

As mentioned in the previous paragraph, the other three most frequent classifiers *cây* (tree, long object), *quả* (fruit, round), and *hòn* (round) in the Narrative Corpus are in Löbel (2000)'s list of core classifiers although they are not frequent in the online newspaper and spoken corpora. As analysed in section 4.2.2, *cây* (tree, long object) is used with nouns indicating all kinds of trees or plants including bamboo trees, apple trees, flower plants, and even grass as well as other long objects such as swords or lamp poles. Similarly, *quả* (fruit, round) appears with nouns indicating all kinds of fruit including starfruit and peppers as well as round objects such as balls or eggs. The classifier *hòn* (round) appears with such nouns as *hòn ngọc* (CL gem), *hòn than* (CL coal), *hòn đá* (CL stone), *hòn núi* (CL mountain), and *hòn đảo* (CL island). These three classifiers go with a limited number of inanimate nouns. However, due to the repeated use of these nouns in the narrative corpus, they become common regardless of the number of inanimate nouns the classifier can combine with.

Furthermore, *việc* (activity) and *vụ* (catastrophe) are among the most frequent classifiers in the Online Newspaper Corpus but have not yet received much attention from prior researchers. With the function of nominalization, *việc* (activity) accounts for over 4% with 116 tokens in the online newspapers. It nominalizes 95 different verbs, mainly action verbs, in the corpus. The classifier *vụ* (catastrophe), which was recognized in previous studies, classifies and nominalizes 23 verbs or nouns in the Online Newspaper Corpus. This classifier is less frequent and classifies fewer nouns than *việc* (activity) does. With the nominalization function, these two classifiers appear quite often in the newspapers although they do not occur often in the narrative and spoken corpora. The variation in the uses of these classifiers shows evidence that the distribution of classifiers depends on genres and content of the text or discourse. The findings about the group of classifiers functioning as nominalizers is further discussed in section 5.4.

The most interesting finding about the distribution of inanimate classifiers in Vietnamese across the three genres lies in the uses of *sự* (event) and *cuộc* (strike, life) in addition to the use of *cái* (inanimate). While *sự* (event) and *cuộc* (strike, life) are frequent in the online newspaper and spoken corpora, they rarely appear in the narrative. Their distribution differentiates the use of classifiers in the two contemporary corpora from the narrative. The number of tokens and percentage of these three classifiers in the three corpora are shown in Table 10 for comparison. Table 10 shows that *sự* (event) appears only six times in the Narrative Corpus. However, as the

most frequent in the Online Newspaper Corpus, it appears 277 times with 192 different stative or adjectival verbs, while it goes with 86 different stative verbs or adjectival verbs in 129 tokens in the Spoken Corpus. This classifier reveals its substantial use in the Vietnamese online newspaper and spoken corpora with its ability to combine with a rich variety of stative and adjectival verbs.

*Table 10: Distribution of *cái* (inani.), *sự* (event), and *cuộc* (strike, life)*

CLs	Narrative Corpus (115,000 words)		Newspaper Corpus (135,900 words)		Spoken Corpus (151,000 words)	
	No. of tokens	%	No. of tokens	%	No. of tokens	%
<i>cái</i> (inani.)	404	22.2	180	7.5	2658	61.4
<i>cuộc</i> (life, strike)	14	0.77	187	7.5	201	4.65
<i>sự</i> (event)	6	0.33	277	11.2	129	2.98
<i>cái sự</i> (inani., event)	0	0	0	0	144	3.33

The distribution pattern of *cuộc* (life, strike, match) is somewhat similar as the pattern of *sự* (event). While *cuộc* (life, strike, match) rarely appears in the narrative with only 14 tokens, it is the second most frequent classifier, with 187 tokens in the newspaper, and 201 tokens in the spoken corpus. The data shows that *cuộc* (life, strike, match) combines with a smaller number of verbs/nominals than *sự* (event) does. While *cuộc* (life, strike, match) appears with 55 different verbs/nominals in the online newspaper, it goes with 11 different verbs/nominals in the spoken corpus. They are mainly action verbs, so *cuộc* (life, strike) not only individuates but also nominalize these verbs in Vietnamese, which is similar as *sự* (event). These two classifiers have not much been attended to in previous studies although it was recognized by researchers including D. H. Nguyen (1957), and H. T. Nguyen (2004, 2013).

I assume that the asymmetrical distribution of *sự* (event) and *cuộc* (life, strike, match) across genres is attributed to the genre effects and the content of the texts and discourse. The classifier *sự* (event) is primarily used with stative verbs or adjectival verbs to express concepts, perceptions, or feelings of humans, while *cuộc* (life, strike, match) normally combine with action verbs to indicate processes of some activities in the contemporary newspapers and talk shows. However, in the narrative, folktales are stories with simple content about animals and stuff as their main readers are children. The sophisticated expressions as in newspapers or talk shows hardly

ever appear in folktales. Thus, the difference in the content of the discourse affects the distribution of classifiers. Furthermore, as a means of communication to convey information and arguments, language plays its role for the effectiveness of communicative purposes. In this case, to express various activities, processes, perceptions, behavior, and feelings, the speaker needs to find an appropriate classifier for their nouns/nominals. To be specific, *sự* (event) and *cuộc* (life, strike), functioning as nominalizers, are used in the newspapers and talk shows to convey the speaker's/writer's more sophisticated information and/or ideas precisely. In contrast, the writers of the folktales do not need these classifiers for their expressions in their writing. This means, changes in language use are needed for the communicative purpose. In brief, the asymmetrical distribution of these classifiers across genres may be due to the genre effects and the content of the discourse. From the communicative perspective, *sự* (event) and *cuộc* (life, strike) can be considered as important classifiers that make nouns for communicating more sophisticated expressions in the current social contexts. Their frequent use in the two concurrent corpora, newspaper and spoken, but not in the narrative, reveals diachronic variation in classifier use. This shows evidence that language change regarding choice of classifiers may be in progress over time in Vietnamese.

Furthermore, the data shows that *sự* (event) appears more often in the newspaper than in the spoken corpus. However, *sự* (event) is also constructed with *cái* (inanimate) quite often in the spoken corpus. This double classifier appears 144 times as shown in Table 10 although it does not occur in the newspaper corpus. The functions of *cái* (inanimate) and *sự* (event) in this doubling construction will be discussed in section 5.2.2. It is clear that the nature of this double classifier in combination with nouns/nominals depends on the function and combinability of *sự* (event). That means, technically, the double classifier *cái sự* (inanimate, event) can go with whatever noun or nominal that *sự* (event) can combine with. To be specific, *cái sự* (inanimate, event) appears with 102 different nouns/nominals in the Spoken Corpus. Despite the fact that *sự* (event) appears often as a single classifier and in the doubling construction *cái sự* (inanimate, event) in the spoken corpus, it is far less frequent than *cái* (inanimate). In contrast, in the online newspaper corpus, *sự* (event) is the most frequent with 277 tokens, while *cái* (inanimate) is the third frequent with 180 tokens. Thus, *cái* (inanimate) is not as frequent in the newspaper as it is in the spoken corpus as a free-standing classifier and as a part of the double classifier construction.

Interestingly, the study has found that *cái* (inanimate) is far more frequent in the talk shows, while *sự* (event) is more preferably used than *cái* (inanimate) in the newspapers. It has been observed that a number of verbs and adjectival verbs, such as chia sẻ (share), hy sinh (sacrifice), hiểu biết (know), thành công (succeed), khác biệt (different), and tự tin (self-confident), appear with *sự* (event) in the newspaper, while they go with *cái* (inanimate) in the spoken corpus. For instance, it is found that the verb thành công (succeed) is classified, nominalized and individualized by *cái* (inanimate) as in (74a) in the spoken corpus, and by *sự* (event) as in (74b). Similarly, the adjectival verb khác biệt (different) is classified, nominalized and individualized by *cái* (inanimate) as in (75a) in the spoken corpus, and by *sự* (event) as in (75b).

- (74) a. *cái* thành công của bộ phim nó không phải là ở kịch bản
 CL(inani.) succeed of CL(set) movie it not right be in transcript
 ‘the success of the movie does not lie in the transcript’ (S2.5236)
- b. *Sự* thành công của điện ảnh Hàn Quốc
 CL(event) succeed of cinema South Korea
 ‘The success of South Korean cinema’ (O54.5211)
- (75) a. Và một *cái* khác biệt khác mà mình thấy cần phải bàn nhiều hơn,
 and one CL(inani.) different other which I find need must discuss much more
 ‘And the other difference which I find needs to be discussed more,’ (S16.53)
- b. chính là điểm tạo nên *sự* khác biệt ở Đại học Duy Tân.
 just be feature create CL(event) different at University Duy Tan
 ‘(that) is the feature to create difference at the University of Duy Tan.’ (O114.9072)

This difference in the choice of classifiers for the same nouns in these two concurrent corpora suggests that the choice of classifiers might be affected by the different formality level of the genres. As people have different ways of expression, and one person may express the same idea in different ways when addressing different audiences, variation takes place (Heylighen and Dewaele 1999). The spoken and newspaper genres differ as the writer primarily focuses on communicating information rather than developing a personal relationship as interlocutors do (Biber and Conrad 2009). The speakers/writers in the spoken and online newspaper corpora address different audiences and have different communicative purposes. The participants in the talk shows interact within a small group of people with direct interpersonal interaction, while the

writers of the newspaper address a large group of public people generally with no direct interaction. The speakers in the conversations in the spoken corpus usually exchange their opinions or ideas related to certain topics, while the writers of the newspaper normally reports events, including factual and opinion reports, in many areas. According to Heylighen and Dewaele (1999), in a text with higher formality, the formal, non-deictic category of words including nouns and adjectives would be more frequent, while the deictic category including pronouns would decline with increasing formality of speech. The higher frequency of nouns/nominals which are nominalized by *sự* (event) in the online newspaper corpus compared to the spoken corpus makes the formality of the newspaper higher than the spoken corpus. It is worth noting that *sự* (event) does not appear with a numeral or a demonstrative with the omission of the noun when the noun is identified in the context. That means, this classifier usually goes with nominals, and is not used as a deictic word in the corpora. Unlike *sự* (event), *cái* (inani.) appears quite often with demonstratives or numerals with the omission of the noun when the noun is identified in the preceding context in the spoken corpus. Also, *cái* (inani.) combining with some other morphemes is used as pronouns or pro-forms quite often in the spoken corpus. Thus, the frequent use of *cái* (inani.) as a deictic category can be seen as evidence that the formality in the spoken corpus is lower than that of the newspaper. In short, the formality of the genres may affect the choice of classifiers in Vietnamese or the choice of classifiers may be dependent on the formality of the genres.

In sum, as I do not investigate classifiers in newspapers of sixty years ago or so in this study, I cannot compare the use of classifiers in newspapers diachronically. However, to some extent, the high frequency of *sự* (event) and *cuộc* (life, strike) in the online newspaper and spoken corpora compared to the narrative reveals some diachronic variation or change in the choice of classifiers in Vietnamese. This change in language use can be considered as a response to the current changes of the world, especially in communication. The differences in the choice of *cái* (inanimate) and *sự* (event) in the spoken and newspaper show synchronic variation in classifier choice in these two concurrent corpora. Furthermore, the different choice of *cái* (inanimate) and *sự* (event) in these genres suggests that the formality affects the choice of classifiers. These two classifiers *sự* (event) and *cuộc* (life, strike) are discussed together with other classifiers functioning as nominalizers in section 5.4. As a certain classifier is used with certain nouns in Vietnamese, due to the different content of the text or discourse, the number of classifier types in the three corpora differs. This will be discussed in the next section.

5.1.3 Number of actual inanimate classifiers in the corpora

The data shows differences in the number of actual classifiers identified in the three corpora. Figure 5 shows the actual number of inanimate classifier types with distinctive portions of single and double classifiers in the three corpora. The Narrative Corpus has the largest number of actual inanimate classifier types (192), while the Online Newspaper Corpus has 153 inanimate classifier types. The Spoken Corpus has the lowest number of inanimate classifier types (134) even though 48 double classifiers are counted as different classifiers from the single ones that use the same morphemes. Without the 48 double classifiers included, the number of classifier types in the Spoken Corpus is only 86. As shown in Figure 5, the number of double classifiers is very small in the narrative and online newspaper corpora, only three and two respectively.

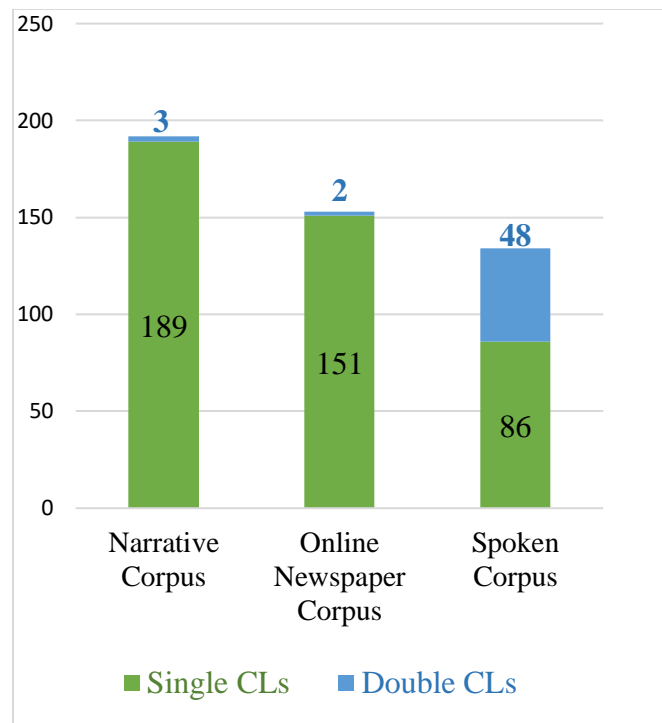


Figure 5: Number of actual inanimate CLs in the three corpora

This discrepancy in the number of inanimate classifier types across the three genres may be due to the characteristics of the three genres, especially the content of texts or discourse. In Vietnamese, a number of nouns can go with a certain classifier, while many nouns can combine with several different classifiers (D. H. Nguyen 1957). This means there would be no other choice of classifiers for a number of certain nouns. However, there would be several different choices of

classifiers for other nouns. Thus, the content of texts or discourse would have great influence in the use of classifiers, depending on the nouns used in the discourse. On the contrary, as discussed in section 5.1.2, the classifier choice is also affected by the formality of the genre and individual speakers. In short, the genre effects, the content of texts or discourse, individual effects, and the formality of the genres may lead to the differences in the number of classifiers across the three genres. In addition, the data shows that the number of mensural classifiers results in the different numbers of classifier types used in the three corpora.

The results of the study show that the Narrative Corpus has the biggest number of mensural classifiers compared to the other two corpora. Specifically, 69 mensural classifiers are identified in the Narrative Corpus while 42 and 17 mensural classifiers are found in the Online Newspaper and Spoken Corpora respectively. Clearly, this big number of mensural classifiers in the narrative leads to the highest number of inanimate classifier types used in the corpus. The data shows that the more mensural classifiers are used in the corpus, the higher number of classifier types it is. This means, the difference in the number of classifiers in each of the corpora is attributed to the number of mensural classifiers. According to Grinevald (2000:64), in numeral classifier systems with a large number of classifiers, the majority of classifiers are in fact “mensural classifiers” while the number of “true classifiers” is very limited. This may be true for the Vietnamese classifier system and can explain the reason why the number of classifier types in Vietnamese is high. To make it clear, I would distinguish mensural classifiers from true or proper classifiers. True or proper classifiers and mensural classifiers are the two semantic subtypes of numeral classifiers. Proper classifiers are the ones which always perform their own function of classification and individualization and/or nominalization (D. H. Nguyen 1957; H. T. Nguyen 2004, 2013). In contrast, mensural classifiers are measure words which are used as classifiers to individuate mass stuff (Grinevald 2000:64).

A wide variety of mensural classifiers are used in the Narrative Corpus since the folktales tell stories about human life of different minority ethnic groups in rural and mountainous areas. They use a number of words indicating something similar to a ‘basket’, for instance, *thúng* (large round basket), *bỏ* (tube-shaped basket), *giá* (small basket), *giỏ* (small basket with handles), and *bị* (sedge bag) as mensural classifiers for the noun ‘rice’ although they may be made of different materials and in various shape and size. While *thúng* (large round bamboo-made basket) is a large

round basket, *bồ* (tube-shaped basket) is smaller but has a tube-like shape. However, *giá* (small basket) is small and shallow. These three objects are all made of bamboo or the like, and have no handles. Both *giỏ* (small basket with handles) and *bị* (sedge bag) are small and usually have one or two handles. *Giỏ* (small basket with handles) is made of bamboo, but *bị* (sedge bag) is made of sedge. Below are the examples of mensural classifiers and proper classifiers for easy comparison and contrast. The classifiers *thúng* (large round basket) as in (76a), *bị* (sedge bag) as in (76b), and *giá* (small basket) as in (76c) are mensural. They appear with mass nouns, such as *gạo* (rice) in these examples. As mensural classifiers, they are to measure the quantity of the thing that the noun they precede refers to.

- (76) a. *cứ giao cho tôi một thúng gạo.*
 let give for me one CL(large round basket) rice
 ‘let’s give me one basketful of rice.’ (N1.177)
- b. *lén để lại cho con một bị gạo*
 secret leave back for child one CL(sedge bag) rice
 ‘secretly left a sedge bagful of rice for the child’ (N2.186)
- c. *làm lật giá gạo đổ cả xuống ao*
 make turn over CL(small basket) rice pour whole into pond
 ‘turned the small bagful of rice over into the pond’ (N1.36)

Furthermore, synonyms of measuring words coming from different dialects or subdialects are used as mensural classifiers in the Narrative Corpus. For example, *bát* (bowlful), *chén* (bowlful), and *giáp* (big bowlful) indicating the same thing are used as mensural classifiers for the noun ‘water, rice, or soup’. However, *bát* (bowlful) comes from the Northern Vietnamese dialect, while *chén* (bowlful) comes from the Southern Vietnamese dialect, and *giáp* (big bowlful) from the Vietnamese subdialect spoken by a minority ethnic group of people living in Northern mountainous areas. The use of synonyms of measuring words from different dialects of Vietnamese as mensural classifiers makes their number higher, resulting in the higher number of inanimate classifiers in the Narrative Corpus. On the contrary, in the newspaper and spoken corpora, fewer mensural classifiers appear. Especially, in the spoken corpus, speakers normally communicate opinions, ideas, and concepts with abstract nouns due to the characteristics and topics of the talk shows, so a very small number of measure words are used as classifiers. As any

measure words can become mensural classifiers (D. H. Nguyen 1957), the number of mensural classifiers can increase. Measure words can function as nouns in other cases.

In contrast, proper/true classifiers cannot stand on their own or function as nouns in any case. The classifiers *chiếc* (individual) and *cuộc* (life, strike, match) are two of the many other proper classifiers identified in the corpora. As proper classifiers, they usually classify, individuate and/or nominalize the noun or nominal they precede and make them countable as *chiếc* bánh (CL cake) in (77a) and *cuộc* thi (CL contest) in (77b).

- (77) a. Trên bàn có bốn *chiếc* bánh
On table have four CL(individual) cake
'On the table there are four cakes' (N1.42)
- b. Cả ba *cuộc* thi chúng đều bị thua.
All three CL contest they all PASS lose
'All the three contests they lost.' (N2.231).

In short, this study has found 297 actual inanimate classifier types altogether including 49 double classifiers counted as different classifiers and added up. Excluding 49 double classifiers in which *cái* (inanimate) is constructed with another classifier, 248 inanimate classifier types identified in the three corpora of this study is the higher number of classifiers claimed to date. Each of the corpora has a different number of classifiers. However, 57 classifiers overlap in the three corpora, while 68 other classifiers overlap within two of the three corpora, out of which 64 classifiers overlap in the Narrative and Online Newspaper Corpora and four overlapped in the Online Newspaper and Spoken Corpora. As it is a long list, the actual inanimate classifier types used in the three corpora are put in Appendix D for comparison. As the highest number of actual classifiers including human and animate types reported by Nguyen (2002) is 195, the total of classifiers in the inanimate category found in this study is much higher than those claimed in previous research. That means there are even more classifiers of the human and animate categories, which have not yet been examined in this study. This high number of classifier types identified in this study may be due to the fact that 110 mensural classifiers altogether are used in the three corpora. This evidence supports Grinevald (2000)'s claim that the majority of classifiers in a numeral classifier system with a large number of classifiers are mensural classifiers. However, regardless of the mensural classifier type, the number of proper classifiers in the corpora is still

higher than what has been claimed in previous research. Clearly, as analysed in (77), *chiếc* (individual) and *cuộc* (life, strike, match) are always “proper classifiers” with the function of classification, individualization and/or nominalization in Vietnamese. They cannot be head nouns in any case as they are not independent by themselves. This ascertains that there are more “proper classifiers” used in Vietnamese than only three general classifiers recognized by Cao (1998). The issue of number of actual classifiers is beyond the focus of this study, so I will not further investigate it, but leave it for future research.

In sum, the study has identified a large number of both proper classifiers and mensural classifiers in the Vietnamese classifier system. A number of inanimate classifiers found in this study are not in the list of classifiers claimed in previous research (Emeneau 1951; Nguyen 1957; Thompson 1965; Nguyen 2002). This is plausible because previous studies are not corpus-based, but mainly based on the basic vocabulary or constructed and/or elicited utterances. More recent studies did not investigate all classifiers or report the number of classifiers in Vietnamese as they focus on analysing some issues of classifiers (H. T. Nguyen 2004; T. B. N. Nguyen 2013). This means that a corpus-based study reveals more classifiers than in traditional studies. Together with the fast-changing world nowadays, language change might be in progress as a response to the needs of human communication. Non-corpus studies might have certain limits because they cannot reflect the realistic picture of how language is actually in use or predict language change in apparent time. The findings of this corpus-based study reveal that the Vietnamese classifier system is highly complex with various classifiers, in which a number of different classifiers can go with the same noun. This large set of classifiers in a language allows a great number of choices as Adams (1986) argues. This means, the Vietnamese classifier system with a larger set of classifiers allows more choices, regardless of mensural or proper classifiers. The choice of a classifier basically depends on a particular shape-related or other property of the referent which the speaker wants to focus on as Aikhenvald (2000) and Behrens (2003) claim. Additionally, as discussed in section 5.1.2 and in this section, the finding of this study reveals that the differences in the uses of classifiers may be due to the genre effects, the content of the texts and discourse, the individual speakers. The choice of classifiers is also dependent on the formality of different genres. In short, Vietnamese speakers have a wide choice of classifiers for a noun since the Vietnamese classifier system has a large number of classifiers. Although the exact number has not been determined, it

is definitely to be over 200. Corpus-based studies have identified more classifiers in use in Vietnamese than traditional research does.

5.1.4 Summary

In sum, the findings of the study show variation in classifier frequency and distribution in the three corpora. The classifier frequency in the Spoken Corpus is much higher than the rates in the Narrative and Online Newspaper Corpora. This may be due to the genre effects, the content of the texts and discourse, and individual speakers. In addition, the study has found that the frequent use of *cái* (inanimate) as emphatics with non-classified nouns in the spoken corpus, but not in the written corpora, results in the higher frequency in spoken Vietnamese than in written language. Vietnamese speakers make use of *cái* (inanimate) as a paralinguistic device for emphasis to attract the listener's attention and/or focus on certain nouns in their speech.

Interestingly, there is a decline in classifier frequency among younger speakers compared to older speakers in the Spoken Corpus despite an increase in classifier use in the spoken over the written corpora. The decline in classifier frequency among younger age group may be due to the content of the discourse and individual effects. It is interesting that the use of *cái* (inanimate) with non-classified nouns as emphatics does not affect the classifier use pattern among different age groups. On the contrary, the finding that the required classifiers before classified nouns with specific referents are sometimes omitted among younger speakers leads to the assumption that language simplification may be employed for gains of speed when the omission does not impact the coherence of the discourse. However, the data of this study is not enough for testing this hypothesis, so it needs further investigation in a larger corpus study on spoken Vietnamese.

The overall distribution of classifiers differs within each corpus and among the three corpora. The results reveal more frequent classifiers in the Narrative Corpus overlap with those in Löbel (2000)'s list of ten core classifiers. The distribution of inanimate classifiers in the Narrative Corpus is closer to previous research, while the Online Newspaper and Spoken Corpora have more frequent classifiers in common. Specifically, *sự* (event) and *cuộc* (life, strike) are used very frequently in the online newspaper and spoken corpora, but rarely in the narrative. Their distribution in the two concurrent corpora compared to the narrative reveals some change in the choice of classifiers in Vietnamese, and this can be seen as diachronic variation. In addition, *cái*

(inanimate) is preferably used in the spoken corpus, while *sư* (event) is used with the same nouns in the newspaper. Moreover, *cái* (inanimate) is used as the deictic category including pronouns or pro-forms when combining with different words or morphemes in the spoken corpus, while *sư* (event) just appears in the non-deictic category with nominals in the newspaper corpus. This finding reveals that the choice of classifiers may be dependent on the formality of the genres. This difference in classifier choice is seen as synchronic variation in classifier use between genres. In short, beyond the possible reasons of genre effects and the content of the texts and discourse, the formality of the genre is also an important factor that may affect the use of classifiers regarding frequency and distribution.

The number of actual inanimate classifier types identified in the study is 248 excluding 49 double classifiers, in which 110 mensural classifiers altogether are used in the three corpora. This number is much higher than those claimed in prior research because this is a corpus study while previous studies are mainly on constructed and/or elicited utterances. This provides evidence that corpus-based studies reveal more interesting findings on actual language use than traditional research. The data shows that comparing classifier use in written and spoken discourse shows interesting findings, which is discussed in the next section.

5.2 Comparing classifier use in written and spoken discourse

This section briefly discusses the major differences in classifier use in written and spoken discourse. The differences in classifier frequency and distribution in these modes are discussed in section 5.2.1. As one of the main differences in classifier use in spoken and written discourse, the use of double classifiers and their lexical semantic functions are analysed in 5.2.2. Section 5.2.3 summarizes the main points in the section.

5.2.1 Differences in classifier use in written and spoken discourse

The classifier frequency in written versus spoken discourse is briefly discussed in 5.2.1.1. The distribution of classifiers in the two modes of discourse is compared in 5.2.1.2.

5.2.1.1 Frequency of classifier use in written versus spoken discourse

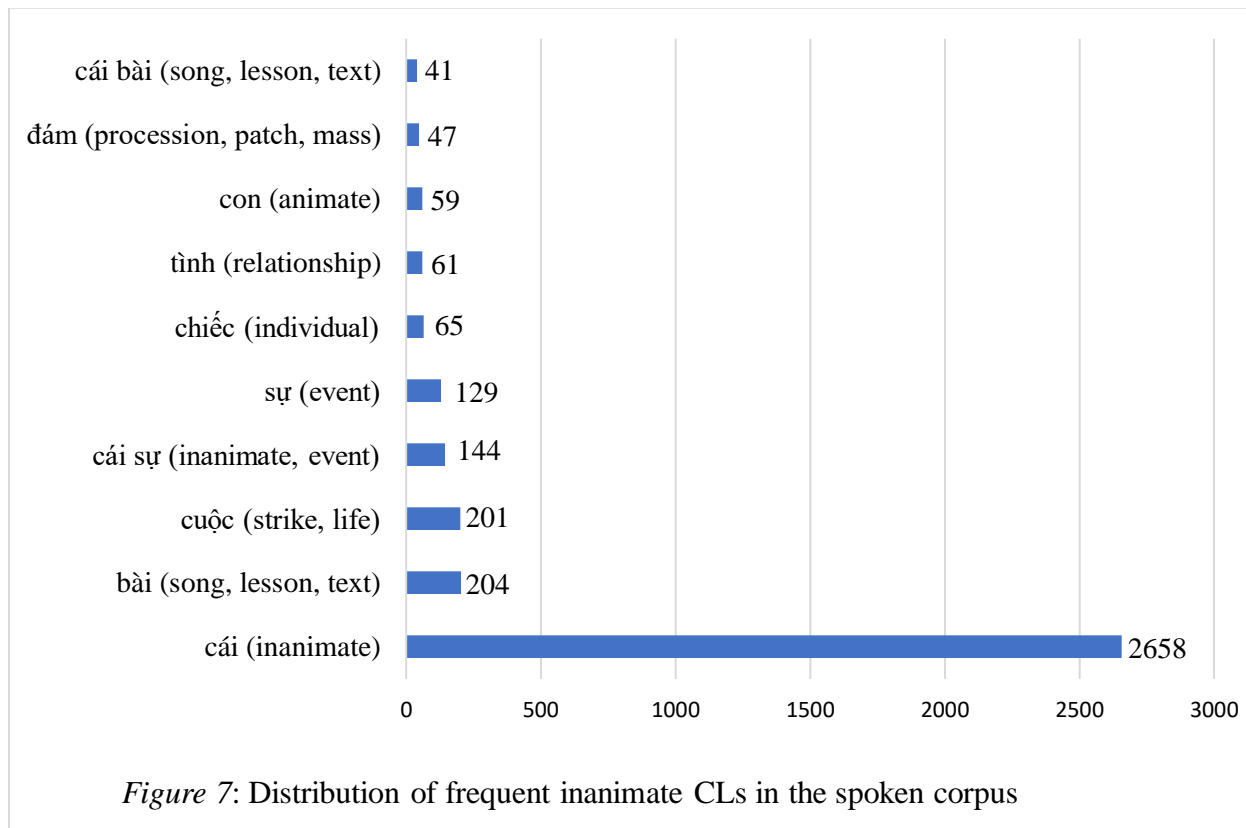
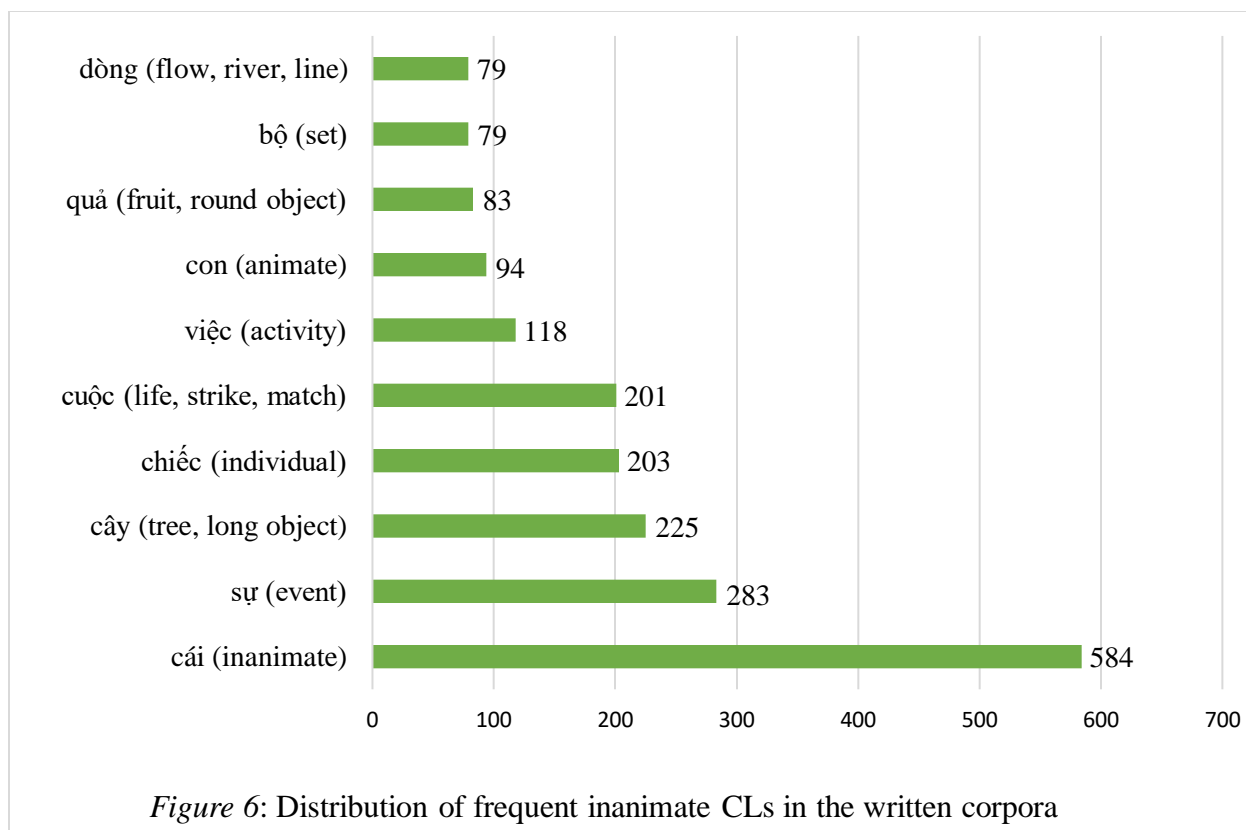
The results show a substantial difference in classifier frequency in the Spoken Corpus in comparison with the two written corpora. In fact, the rate of classifier use in the Spoken Corpus is

286 per 10,000 words while it is 172 per 10,000 words in the two written corpora taken together. This means that classifiers are more frequently used in Vietnamese conversations than in written texts. This is discussed in section 5.1.1. Clearly, this difference may be due to the genre effects, the content of the texts or discourse, and individual speakers. Furthermore, the finding shows that the higher classifier frequency in conversations over written texts may be attributed to the frequent use of *cái* (inanimate) with non-classified nouns as emphatics in spoken Vietnamese. A classifier is required for classified nouns in Vietnamese, but not required for non-classified nouns. However, *cái* (inanimate) appears with non-classified nouns as an emphatic and constructed with another classifier before classified nouns in the classifier doubling construction very often in conversations. Thus, the appearance of *cái* (inanimate) as emphatics is optional, depending on the speaker's intention. The optional use of *cái* (inanimate) as emphatics with non-classified nouns is further discussed in section 5.3.4, and the use of *cái* (inanimate) constructed with a classifier in the classifier doubling construction is analysed and discussed in section 5.2.2. I argue that Vietnamese speakers prefer to use *cái* (inanimate) for emphasizing certain nouns to attract the listener's attention and focus on what they are saying in their speech rather than in their writing. They make use of the phonological stress on *cái* (inanimate) as a paralinguistic device of emphasis in speech (Biber and Conrad 2009) since *cái* (inanimate) in this case was observed to have a phonological stress (D. H. Nguyen 1957; H. T. Nguyen 2004, 2013).

Together with a substantial difference in frequency, the distribution of classifiers in spoken and written discourse clearly differs. These differences will be discussed in the next subsection.

5.2.1.2 Distribution of classifiers in written and spoken discourse

Among the ten most frequent inanimate classifiers in written and spoken discourse, five of them overlap including *cái* (inanimate), *cuộc* (life, strike, match), *sự* (event), *chiếc* (individual), and *con* (animate). Three out of these five classifiers including *cái* (inanimate), *chiếc* (individual), and *con* (animate) are common and got attended to in previous studies, while the other two classifiers: *cuộc* (life, strike, match) and *sự* (event) are not. This is because these two classifiers are becoming more frequent in the contemporary corpora, while they rarely appear in the narrative, which were written over sixty years ago, as discussed in section 5.1.2. However, the distribution of all these frequent classifiers greatly differs within the genre as well as in spoken and written discourse as shown in Figure 6 and Figure 7.



The finding shows that the most noticeable differences in the use of classifiers in spoken and written discourse are the frequent use of *cái* (inanimate) and double classifiers in the conversations over the written texts. All the most frequent inanimate classifiers in the written corpora are single classifiers, while in the spoken corpus, two of them are double classifiers, which are *cái sự* (inanimate, event) and *cái bài* (inanimate, lesson, song, text). The use of *cái* (inanimate) in the three corpora is discussed extensively in section 5.3. The use of double classifiers in the Spoken Corpus is discussed in the next section.

5.2.2 Double classifiers and their lexical semantic functions

The data shows that a variety of double classifiers appear quite often in the spoken corpus while a few of them with few tokens are used in the written corpora. Specifically, 48 different double classifiers are identified with 403 tokens, accounting for 9% of all the tokens in the Spoken Corpus, while only 5 double classifiers are found with 9 tokens altogether in the written corpora. The phenomenon of two classifiers co-occurring is a distinctive feature in Vietnamese classifier system. In the construction of the double classifiers identified in this study, *cái* (inanimate) always precedes a specific classifier before combining with a classified noun, except for the *cái con* (inani., animate) combination. The data reveals four out of the five double classifiers found in the written corpora overlap with those in the Spoken Corpus. Therefore, 49 double classifiers are identified with 412 tokens altogether in this study.

It is interesting find that most of the specific classifiers in the classifier doubling construction also appear as single classifiers in the spoken corpus. Specifically, forty-one specific classifiers in this construction are used in the single form, while seven others are not found in the single form including *khoản* (amount), *khúc* (section), *làn* (wave), *mảnh* (piece), *sợi* (thread, string), and *set* (set)¹⁴. Technically, all these classifiers can be used as single classifiers. In fact, they are found in the Narrative and Online Newspaper Corpora although they co-occur with *cái* (inanimate) in the classifier doubling construction in the Spoken Corpus.

Most of the double classifiers do not appear often, accounting for less than one percent of all the tokens each, except *cái sự* (inanimate, event). To be specific, *cái sự* (inanimate, event) is

¹⁴ *set* (set) is a borrowing from English. In fact, the ‘pure’ Vietnamese classifier *bộ* (set) which has the same function as *set* (set) appears in the Spoken Corpus.

the most frequent double classifier with 144 tokens, accounting for over 3% of all the tokens in the Spoken Corpus. Following it, *cái bài* (inani., song, lesson, text) appears far less often with 41 tokens, accounting for approximately one percent. Table 11 lists all the double inanimate classifiers in the Spoken Corpus with their numbers of occurrences and percentages.

Table 11: Distribution of double classifiers in the Spoken Corpus

Double classifiers	No. of occurrences	%
cái sự (inanimate, event)	144	3.49
cái bài (inani., song, lesson, text)	41	0.99
cái cuộc (inani., strike, life)	26	0.63
cái con (inani., animate)	16	0.38
cái phần (inani., part)	15	0.36
cái đám (inani., procession)	14	0.34
cái bộ (inani., set)	13	0.31
cái niềm (inani., sentiment)	12	0.29
cái việc (inani., activity)	11	0.27
cái khoảng (inani., period)	10	0.24
cái nỗi (inani., feeling)	8	0.19
cái chuyến (inani., trip)	8	0.19
cái món (inani., dish)	7	0.17
cái mối (inani., relationship)	6	0.15
cái ánh (inani., glow)	6	0.15
cái buổi (inani., session)	5	0.12
cái chiếc (inani., individual)	4	0.10
cái dòng (inani., line)	4	0.10
cái tính (inani., quality)	4	0.10
cái cơn (inani., anger, wind)	4	0.10
cái bước (inani., step)	4	0.10
cái cú (inani., blow)	3	0.07
cái cuốn (inani., volume)	3	0.07

cái ngôi (inani., unit of house)	3	0.07
cái quả (inani., fruit, round object)	3	0.07
cái sợi (inani., thread)	3	0.07
cái căn (inani., unit of house)	2	0.05
cái giấc (inani., sleep)	2	0.05
cái quãng (inani., section)	2	0.05
cái quyển (inani., volume)	2	0.05
cái bản (inani., script)	1	0.02
cái bức (inani., picture)	1	0.02
cái cánh (inani., door)	1	0.02
cái cây (inani., tree, long)	1	0.02
cái chặng (inani., section)	1	0.02
cái điều (inani., affair)	1	0.02
cái đoạn (inani., section)	1	0.02
cái đồng (inani., money)	1	0.02
cái đốt (inani., knot)	1	0.02
cái khúc ((inani., part)	1	0.02
cái làn (inani., wave)	1	0.02
cái mảnh (inani., piece)	1	0.02
cái ngọn (inani., top part)	1	0.02
cái nụ (inani., smile)	1	0.02
cái set (inani., set)	1	0.02
cái tấm (inani., picture, degree)	1	0.02
cái tờ (inani., sheet)	1	0.02
cái túi (inani., bag)	1	0.02
Total	403	9.68

The double classifier *cái sự* (inani., event) appears with 102 different stative or adjectival verbs in the Spoken Corpus. In this combination, *sự* (event) nominalizes and individuates the nominals while *cái* (inani.) emphasizes the nouns. The double classifier *cái sự* (inani., event)

precedes the verbs tấn công (attack) and chấp nhận (accept) as in (78a-b). The classifier *sự* (event) nominalizes and individuates these verbs while *cái* (inani.) emphasizes the nominals *sự tấn công* (CL attack) and *sự chấp nhận* (CL accept). Therefore, *cái* (inani.) in this double classifier construction is used as an emphatic. A numeral may not appear in the classifier phrase as in (78a) or may appear as in (78b). Similarly, the adjectival verbs đam mê (passionate) and khác biệt (different) are nominalized and individualized by *sự* (event), and then emphasized by *cái* (inani.) as in (78c-d). The plural morpheme ‘những’ as a numeral precedes this classifier phrase.

- (78) a. chống lại *cái* *sự* tấn công của 4.0
 oppose to CL(inani.) CL(event) attack of 4.0
 ‘opposed to the attack of 4.0’ (S21.6367)
- b. Đây là một *cái* *sự* chấp nhận
 That be one CL(inani.) CL(event) accept
 ‘That is the one acceptance’ (S2.4997)
- c. Và bọn em cần phải học tập những *cái* *sự* đam mê
 And we need have to learn PL CL(inani.) CL(event) passionate
 ‘And we have to learn the passions’ (S3.6530)
- d. *cái* *sự* khác biệt đấy là thấy rất rõ.
 CL(inani.) (event) different that be see very clear
 ‘that difference has been seen vary clearly’ (S16.52)

Furthermore, the data shows that the double classifier *cái sự* (inani., event) appears in a combination of cũng là (also be) or gọi là (called) preceding the nouns or verbs. It goes with cũng là (also be) and the noun nhiệm vụ chung (the common task) as in (79a), while it combines with gọi là (called) and the noun nội lực (internal force) and possessor của em (your) as in (79b) or the verb sụp đổ (collapse) as in (79c). This double classifier also appears with a clause such as mình đã quá cuồng nhiệt đến như thế (we are so frenetic) as in (79d). In brief, this double classifier can combine with a variety of stative or adjectival verbs as well as in the construction of cũng là (also be) or gọi là (called) plus nouns/verbs or even clauses. Its combination possibility is quite diverse.

- (79) a. vì những *cái* *sự* cũng là nhiệm vụ chung
 because PL CL(inani.) CL(event) also be task common
 ‘because of the also common tasks’ (S11.1746)

- b. Những ai có nghe thì thấy *cái* *sự* gọi là nội lực của em
 PL who have listen then find CL(inani.) CL(event) call be internal force of me
 ‘Whoever listens then will find your called ‘internal force’.’ (S14.3832)
- c. là cả một *cái* *sự* gọi là sụp đổ
 be all one CL(inani.) CL(event) call be collapse
 ‘is the one called collapse’ (S4.6899)
- d. chính vì *cái* *sự* mình đã quá cuồng nhiệt đến như thế
 just because CL(inani.) CL(event) we PAST too frenetic to so
 ‘it is because we are so frenetic’ (S8.8257)

The double classifier *cái bài* (inani., song, lesson, text) appears less often than *cái sự* (inani., event). The classifier *bài* (song, lesson, text) combines with a limited number of verbs or nouns. However, these nouns are used repeatedly in the corpus, so this classifier appears quite often. The specific classifier *bài* (song, lesson, text) nominalizes and/or individuates the verbs hát (sing) and học (learn) into the nouns *bài hát* (a/the unit of song) and *bài học* (a/the lesson) as in (80a) and (80c). The presence of *cái* (inani.) in this case is used to emphasize the noun. According to D. H. Nguyen (1957), and H. T. Nguyen (2004, 2013), the use of *cái* (inani.) in the doubling construction emphasizes the noun and makes it definite. However, there is no evidence to show whether the appearance of *cái* (inani.) in this case is involved in making the nouns definite or not although the NPs in (80a)-(80b) are definite because a demonstrative appears in (80a) and a defining clause that postmodifies the noun is used in (80c). Whether a numeral is present or absent does not influence the lexical semantic function of *cái* (inani.) in this construction. Like a single classifier, a double classifier can appear with a numeral with the omission of the noun when the noun is identified in the preceding context. As in (80b) and (80d), the noun ‘song’ is previously mentioned, so it is omitted in these phrases. While *cái bài* (inani., song, lesson, text) goes with the numeral một (one) in the absence of the noun as in (80b), it combines with the demonstrative này (this) with the omission of the noun as in (80d). Therefore, when a noun is identified in the context, double classifiers can occur with numerals or demonstratives without the presence of the noun.

- (80) a. *Cái bài hát* đấy thì Hòa Minzy thể hiện được hết
 CL(inani.) CL(unit) sing that then Hoa Minzy perform get all
 ‘That song, then Hoa Minzy can perform it all’ (S17.459)

- b. Nhưng mà một *cái bài* để người ta nghe đi nghe lại
 But then one CL(inani.) CL(unit) for people listen to listen back
 ‘But the one song for people to listen repeatedly’ (S13.2674)
- c. đó là hai *cái bài* học mà anh muốn chia sẻ.
 that be two CL(inani.) CL(unit) learn which I want share
 ‘Those are the two lessons that I want to share.’ (S15.3891)
- d. Mai một nha phôi lại *cái bài* này cho các bé đi thi The Voice nhá.
 Tomorrow ah mix again CL(inani.) CL(unit) this for PL kid go compete the Voice yay
 ‘You’ll mix this song again for the kids to participate in The Voice in future.’ (S14.3784)

In addition, most of the double classifiers found in the corpus are infrequent, accounting for less than one per cent each. Eight of them appear 10 to 25 times each including *cái cuộc* (inani., life, strike, match), *cái con* (inani., animate), *cái phần* (inani., part, section), *cái đám* (inani., procession), *cái niềm* (inani., happiness, joy), *cái bộ* (inani., set), *cái việc* (inani., activity), and *cái khoảng* (inani., unit of time, area). The remaining thirty-eight double classifiers appear less than 10 times each, out of which eighteen occur only once each. Despite different distribution, the double classifiers in the corpora have the same construction in which *cái* (inani.) is constructed with a specific classifier. In this construction, the specific classifier performs its own function of classification, individualization and/or nominalization, while *cái* (inani.) functions as an emphatic.

It is necessary to clarify one point that *set* (set), an English word, is combined with *cái* (inani.) in the classifier doubling construction *cái set* (inani., set) as in (81a). It is interesting that a borrowed English word could be used as a grammatical item - a classifier - in Vietnamese. Because the English word ‘set’ has the same meaning as *bộ* (set) in Vietnamese, it is used as a classifier instead of the ‘pure’ Vietnamese classifier. In fact, *set* (set) does not appear as a single classifier in the corpus. In contrast, the ‘pure’ Vietnamese classifier *bộ* (set) is found as a single classifier as in (81b) and as a part of the double classifier *cái bộ* (inani., set) as in (81c). In (81b), *bộ* (set) goes with the noun *đồ* (clothes), preceded by the numeral *một* (one) to indicate ‘a very courteous suit’, which is an indefinite noun phrase. However, *cái set* (inani., set) in (81a) and *cái bộ* (set) in (81c) combine with the noun *đồ* (clothes), followed by the demonstrative *đó* (that) or ‘đấy’ (that) respectively to indicate ‘that suit’. These nouns are definite due to the presence of the demonstrative ‘that’, regardless of the appearance of *cái* (inani.) in the combination. The double

classifiers *cái bộ* (set) and *cái set* (set) are similar because the English word ‘set’ borrowed for the later combination is used to perform the same function as *bộ* (set) in Vietnamese.

- (81) a. đúng *cái* *set* đồ đó
 right CL(inani.) CL(set) clothes that
 ‘right that set of clothes’ (S13.2976)
- b. một tân sinh viên mặc một *bộ* đồ rất lịch thiệp
 one new student wear one CL(set) clothes very courteous
 ‘a new student wears a very courteous suit’ (S3.6558)
- c. mà Tường mặc đúng *cái* *bộ* đồ đấy vô
 but Tuong wear just CL(inani.) CL(set) clothes that on
 ‘but Tuong puts on just that suit’ (S13.2977).

Similarly, other double classifiers appear with the nouns which the specific classifier in the combination usually goes with. The specific classifiers in the doubling construction perform the same functions as they do when appearing as single classifiers. For example, *cái* (inani.) co-occurs with *quyển* (volume) preceding the noun ‘book’ in (82a) or with *ngôi* (school) preceding the noun ‘school’ in (82b). The numeral may appear as in (82a) but may not as in (82b). However, in these cases, both of the nouns are definite. The definiteness of the noun may be due to the presence of the defining clause following the noun as in (82a), or previous mention in the context as in (82b).

- (82) a. Nhờ những *cái* *quyển* sách mẹ mua khó hơn rất rất là nhiều
 Thanks PL CL(inani.) CL(volume) book mother buy difficult more very very be much
 ‘Thanks to the books that my mother bought are much more difficult’ (S10.1499)
- b. Từ khi mà bước chân về *cái* *ngôi* trường
 Since when that step foot back CL(inani.) CL(house) school
 ‘Since coming back to the school’ (S7.8098)

The only combination that does not follow the pattern of *cái* (inani.) constructed with a specific classifier is *cái con* (inani., animate) because *con* (animate) is not considered as a specific classifier, but a general animate classifier in Vietnamese. It is interesting that *con* (animate), which is an animate classifier, is constructed with nouns indicating non-living things, especially in the combination with *cái* (inani.). In fact, *con* (animate), one of the ten most frequent classifiers in all

the three corpora, appears with a number of different nouns indicating non-living things, which has been analysed in 4.3.2 and 4.4.2. Technically, the combination *cái con* (inani., animate) can be used with the nouns that *con* (animate) appears with. This double classifier appears 16 times in the Spoken Corpus but only once in the Narrative Corpus as in (30a) in section 4.2.3. As analysed in 4.2.3, this is an extraordinary case when *cái* (inani.) co-occurs with *con* (animate). In the Spoken Corpus, *cái con* (inani., animate) mainly appears with the nouns *đường/ đường nghệ thuật* (road/arts road) and *số* (number) as in (83a) and (83c). These nouns are found to be with the single classifier *con* (animate) as in (83b) and (83d). When combining with *con* (animate) only, the noun may be definite as in (83b) due to the defining clause postmodifying the noun or may be indefinite as in (83d). However, when going with the double classifier *cái con* (inani., animate), the noun is always definite as in (83a) and (83c). The previous mention in the context determines the definiteness of the noun in (83a), while the demonstrative *này* (this) and possessive *của em* (of mine) makes the noun definite in (83c). In this case, *con* (animate) is obligatory because it performs the function of individualization, while *cái* (inani.) is optional and functions as an emphatic. There is no evidence to ascertain that *cái* (inani.) in the double classifier construction makes the nouns definite as H. T. Nguyen (2004, 2013) and Simpson and Ngo (2018) argue. However, the nouns in this case is always definite in the presence of a demonstrative, a defining clause, and/or the noun is determined by the context, i.e. by previous mention.

- (83) a. và mình đã quyết định đi trên *cái con* đường nghệ thuật
 and I PAST decide go on CL(inani.) (ani.) road arts
 ‘and I have decided to go on the arts road’ (S20.5605)
- b. đây chính là *con* đường mà tôi lựa chọn.
 here just be CL(ani.) road which I choose
 ‘this is just the road that I choose.’ (S20.5597)
- c. cho nên là *cái con số* này của em mới 4/6
 so be CL(inani.) CL(ani.) number this of me just 4/6
 ‘so that this number of mine is 4/6’ (S15.4167)
- d. tôi không quan tâm tới *con số*
 I not concern to CL(ani.) number
 ‘Number is of no concern to me’ (S11.2045).

In sum, the phenomenon of co-occurrence of two inanimate classifiers is quite common in spoken Vietnamese although it is rarely used in written language. It appears that Vietnamese speakers prefer to use *cái* (inanimate) before another classifier and classified nouns for emphasizing the nouns with the purpose of attracting the listener's attention and/or focus on certain things in their speech, while the other classifier performs its own function of classification, individualization, and/or nominalization. This use of *cái* (inanimate) in the double classifier construction for emphasis is dependent on the speaker's intention. It is similar as its use with non-classified nouns as emphatics in spoken Vietnamese. The addition of *cái* (inanimate) in this construction functioning as emphatics seems very helpful in speech because it always receives a phonological stress as claimed in prior research (Diep 2005; H. T. Nguyen 2004, 2013). Therefore, double classifiers are used often in spoken Vietnamese, but not in written language.

5.2.3 Summary

There are a number of major differences in classifier use in written versus spoken discourse in Vietnamese. The first and most important difference is that the classifier frequency in spoken language is far higher than the rate in written language. This higher frequency in the conversations over the written texts may be due to the genre effects, the content of the texts and discourse, and individual speakers. The frequent use of *cái* (inanimate) as emphatics with non-classified nouns may also lead to the higher classifier frequency in spoken Vietnamese than in written language. Since non-classified nouns in Vietnamese do not require a classifier, this use of *cái* (inanimate) as emphatics is not obligatory, totally depending on the speaker's intention of attracting the listener's attention or focus on certain things. As it has been observed that *cái* (inanimate) in this case usually has a phonological stress (H. T. Nguyen 2004, 2013), it appears that Vietnamese speakers take the advantage of using this property as a paralinguistic device of emphasis in their speech, not in writing.

It is important to note that double classifiers are of a wide variety and appear often in the spoken corpus while they are rare and of a limited number in the written corpora. This may be because Vietnamese speakers prefer to use *cái* (inanimate) constructed with another classifier required for the noun for the purpose of emphasizing the noun in their speech. The use of *cái* (inanimate) functioning as emphatics is not effective in writing. That is why double classifiers rarely appear in written language. Almost all the double classifiers identified in this study follow

the pattern of *cái* (inanimate) preceding a specific classifier and a classified noun/nominal, except *cái con* (inani., animate). In this construction, the specific classifier performs its own function of classification, individualization, and/or nominalization, while *cái* (inanimate) is used as an emphatic. Thus, the other classifier cannot be omitted, while *cái* (inanimate) can be omitted without changing the meaning of the noun. This means, its appearance is optional, and it is called an extra classifier by previous researchers (D. H. Nguyen 1957; P. P. Nguyen 2002; H. T. Nguyen 2004, 2013).

Another substantial difference in classifier use in spoken and written discourse is that *cái* (inanimate) is very frequent in spoken Vietnamese, but not so often in written language, especially in the newspapers. The uses of this classifier appear to have many things uncovered, which are investigated and discussed in the next section.

5.3 The use of *cái* (inanimate) in the three corpora

The classifier *cái* (inanimate) is widely recognized in Vietnamese by researchers including Emeneau (1951), D. H. Nguyen (1957), Thompson (1965), and P. P. Nguyen (2002). The data of this study shows that *cái* (inanimate) is one of the most frequent inanimate classifiers across the three genres although it appears far more often in the spoken than in the two written corpora. This section discusses the uses of *cái* (inanimate) in the Narrative, Online Newspaper, and Spoken Corpora in sections 5.3.1, 5.3.2, and 5.3.3 respectively. The use of *cái* (inanimate) with non-classified nouns is examined and discussed in section 5.3.4. Section 5.3.5 summarizes the usage of *cái* (inanimate) in this study.

5.3.1 The use of *cái* (inanimate) in the Narrative Corpus

The classifier *cái* (inanimate) classifies, individuates and/or nominalizes 184 different inanimate nouns/nominals indicating large objects to very small ones in the Narrative Corpus. This classifier can go with either concrete nouns such as *cái cầu* (CL bridge), *cái tăm* (CL toothpick) as in (84a-b) or abstract nouns such as *cái mẹo* (CL trick) as in (84c). It not only classifies but also nominalizes verbs or adjectival verbs it precedes, such as the verb ‘ăn’ (eat) in (84d). Also, it classifies and nominalizes the adjectives or adjectival verbs ‘yên’ (peaceful), ‘vui’ (happy), ‘no’ (well-fed), and ‘ấm’ (warm) as in (84e), then turns them into the nouns ‘peacefulness, happiness, wellfedness’ and ‘warmth’ respectively.

- (84) a. *Cái cầu ấy thật cao*
 CL(inani.) bridge that really high
 ‘That bridge is really high’ (N1.179)
- b. *tôi đeo cái tăm xỉa răng.*
 I make CL (inani.) toothpick pick tooth
 ‘I am making toothpicks.’ (N1.102)
- c. *anh ta mới lập ra cái mẹo (rằng)*
 he then create CL(inani.) trick that
 ‘he then created the trick (that)’ (N.1.128)
- d. *vì đã có cái ăn.*
 because already have CL(inani.) eat
 ‘because he already has something to eat’ (N2.104)
- e. *Đánh tan giặc, giành lại cái yên, cái vui, cái no, cái ấm cho buôn làng,*
 defeat enemy get back CL(inani.) peaceful CL happy CL well-fed CL warm for village
 ‘Defeated the enemy, regained the peacefulness, happiness, wellfedness, warmth for the village’ (N2.282).

Additionally, *cái* (inanimate) appears with demonstratives to indicate the thing referred to by the noun that has previously been mentioned, such as *cái này* (this one), *cái ấy* (that one), *cái kia* (that one), and *cái nấy* (that one) as in (85). This supports Emeneau (1951:84)’s claim that in a “numerated substantive phrase” in Vietnamese, a classified noun may be omitted but the classifier must remain when the preceding context has identified the head noun. This means that the classifier in Vietnamese has anaphoric function. It can go with the interrogative ‘gì’ to make a question word ‘*cái gì*’ (what) as in (85a). The noun *cái váy của tôi* (my dress) in (85a) is mentioned in the preceding context, so the combination of *cái* (inanimate) and the demonstrative *này* (this): *cái này* (this one) can be traced back by referring to the antecedent noun indicating ‘this dress’. In this case, *cái* (inanimate) plays the function of a pronoun or pro-form for an antecedent. Similarly, in (85b), in order to figure out what ‘those ones’ refers to, we can trace back for their antecedents. However, in (85c) there are no antecedents for *cái này* (CL this) or *cái kia* (CL that) mentioned in the preceding context as ‘this’ or ‘that’ just means something that was guessed for. In this case, *cái* (inanimate) combining with the demonstrative indicates something generic, not specific.

(85) a. - Hai cô kiểm cái gì đó? - Lét, Le hỏi.

Two you look for CL(inani.) what so Let Le ask

‘What are both of you looking for? - Let and Le asked.’

- Kiểm cái váy của tôi. - Hai chị em cùng nói.

Look for CL(inani.) dress of me two sisters together say

‘Looking for my dress. - Both sisters said.’

- Có phải cái này không?

Have right CL(inani.) this question-word?

‘Is that this *one*?’ (N2.284)

b. - Những cái ấy đều là những cái làm hại mình hết cả,

PL CL(inani.) that all be PL CL(inani.) harm us all

‘Those are all the ones that harm us’ (N1.188)

c. hết đoán cái này, lại đoán là cái kia, nhưng không sao đoán đúng.

Just guess CL(inani.) this again guess be CL(inani.) that but no way guess correct

‘(they) guessed this, then guessed that, but were not able to guess correctly’ (N2.337).

The classifier *cái* (inanimate) also combines with numerals only with the omission of the classified noun when it is identified in the context as in (86). In this example, *cái* (inanimate) appears with the numeral ‘ba’ (three) with the omission of the classified nouns. The head nouns ‘house’, ‘kitchen’ and ‘bed’ can be traced back as they are anaphorically referred to.

(86) - *Cái* nhà che nắng, che mưa, ở được yên lành; *cái* bếp nấu nướng đồ ăn,

CL(inani.) house protect sun protect rain live get peaceful CL(inani.) kitchen cook bake food

ăn được no nề; *cái* giường nằm ngồi nghỉ ngơi, sức được khỏe mạnh.

eat get well-fed CL(inani.) bed lie sit relax health get healthy

Ba *cái* cùng có ơn to cả.

Three CL(inani.) together have gratitude big all

‘The house protects us from the sun and rain, helping us to live peacefully; the kitchen is for cooking food, helping us to eat well; the bed is for relaxing, helping to improve our health.

All the three are very helpful.’ (N1.39)

This evidence supports Bisang (1999)’s claim that in most languages the classifier can go with numerals and/or demonstratives. The data shows that the classifier has anaphoric function but does

not appear alone in its anaphoric function. That means, in Vietnamese a classifier can combine with a demonstrative and/or a numeral with the omission of the head noun when the noun is identified in the context, but never appears alone by itself. This evidence supports Bisang (1999:148)’s suggestion that the “classifier does not occur alone in its anaphoric function” in most languages. It also supports Emeneau (1951)’s and Thompson (1965)’s claim that classifiers in Vietnamese can appear with demonstratives and/or numerals with the omission of the noun identified in the preceding context.

Furthermore, *cái* (inanimate) also occurs with wh-word to create pronouns such as *cái gì* (what), *cái nào* (which/anything) as in (87). It combines with the morpheme ‘gì’ (what) to create the pronoun ‘what’, which can be used in either objective case as in (87a) and as a relative pronoun as in (87b) or in subjective case as in (87c). It appears with the morpheme ‘gì’ (what) to make an indefinite pronoun used in affirmative sentences like ‘something’ in English as in (87d) or in negative sentences like ‘anything’ as in (87e). Moreover, *cái* (inanimate) combines with a numeral and the morpheme ‘nào’ (which/anything) in negative sentences to emphatically refer to ‘any one single (item)’ that has been anaphorically mentioned with the omission of the head noun as in (87f).

(87) a. *Bác làm cái gì đấy?*

You do CL(inani.) what that

‘What are you doing?’ (N1.102)

b. *Hai chị em không hiểu là cái gì,*

two sisters not know be CL(inani.) what

‘Both sisters do not know what they are,’ (N2.337)

c. *Như trong nhà thì cái gì công to hơn cả?*

as in home then CL(inani.) what gratitude big than all

‘As in home, what is of the biggest gratitude?’ (N1.38)

d. *thế là trong mâm có cái gì vỡ đổ loảng xoảng cả.*

then be in table have CL(inani.) what break fall noisily all

‘then in the dining table there is something breaking and falling off noisily’ (N1.135)

e. *không rèn cái gì cả.*

not forge CL(inani.) what all

‘(he) does not forge anything at all’ (N1.187)

- f. Nhưng qua bao nhiêu ngày tháng, chẳng thấy ai đến mua cho một *cái* nào.
 But pass how many day month not see who come buy for one CL(inani.) which
 ‘Many days and months passed though; nobody came to buy one.’ (N1.101)

In sum, the classifier *cái* (inanimate) appears not only with a wide variety of inanimate nouns but also with a number of verbs and adjectival verbs in the Narrative Corpus. This classifier also occurs with numerals and/or demonstratives with an anaphoric function. Additionally, it combines with some other morphemes such as ‘gì’ (what) to make up the pronoun ‘*cái gì*’ (what) used in either subjective or objective case as well as a relative pronoun similar to ‘what’ in English. This combination ‘*cái gì*’ (CL what) can be used as ‘something’ in affirmatives or as ‘anything’ in negatives as in English. Furthermore, *cái* (inanimate) can go with numerals and/or the morpheme ‘nào’ (which) in negatives with the omission of the classified noun to indicate ‘anything’ or ‘nothing’, as the noun has been previously mentioned in the preceding context.

5.3.2 The use of *cái* (inanimate) in the Online Newspaper Corpus

Unlike being used as the most frequent classifier in the Narrative Corpus, *cái* (inanimate) is the third most frequent classifier in the Online Newspaper Corpus. It appears with 66 inanimate nouns/nominals. This classifier not only goes with a great variety of nouns but also with verbs and adjectival verbs. In the Online Newspaper corpus, it combines with a larger number of verbs and adjectival verbs than in the Narrative Corpus. This means, *cái* (inanimate) classifies, nominalizes and/or individuates these verbs and adjectival verbs as in (88). The classifier *cái* (inanimate) nominalizes the adjectival verbs *đẹp* (beautiful) in (88a) and *sai* (wrong) in (88b), and turns them into the nouns *cái đẹp* (CL beauty) and *cái sai* (CL wrong) respectively. Similarly, *cái* (inanimate) nominalizes the verb *suy nghĩ* (think) in (88c) and turns it into the noun/nominal *cái suy nghĩ* (thinking/thought).

- (88) a. *cái* đẹp không nhất thiết đi kèm với các tiêu chí to
 CL(inani.) beautiful not necessary go with PL criteria big
 ‘the beauty unnecessarily goes with the criteria of being big’ (O37.3109)
- b. phải chịu trách nhiệm về *cái* sai của cấp trên,
 must bear responsibility for CL(inani.) wrong of level superior
 ‘(he) must be responsible for the wrongdoing of his superior’ (O49.4663)

c. tôi thấy sợ *cái* suy nghĩ về việc *con* người như *món hàng*
 I find fear CL(inani.) think about issue CL(ani.) human like CL item
 ‘I fear of the thought of humans as items’ (O57.5644).

It is worth noting that *cái* (inanimate) can go with verb phrases in passive form as in (89). In this example, it appears with the passive morpheme ‘được’ (get positively) followed by the noun ‘dân biểu’ (people’s representatives) and the verb ‘bàn’ (discuss). It means this classifier can be used in the structure ‘*cái* (inanimate) + ‘được’ (PASS) + Subject + Verb’, which is likely to be similar as the structure ‘what is/has been done by somebody’ in English.

(89) *cái* được dân biểu bàn nhiều nhất
 CL(inani.) PASS people representative discuss much most
 ‘what has been discussed most by people’s representatives’ (O48.4582).

Additionally, the data shows that *cái* (inanimate) combines with demonstratives such as *cái này* (CL this), *cái đó* (CL that), and *cái kia* (CL that) with the omission of the classified noun when the noun is identified in the preceding context as it does in the Narrative Corpus. However, this classifier also appears with other morphemes to create pronouns such as ‘*cái gì*’ (what) as in (90a-b), ‘*cái gì đó*’ (something) as in (90c-d), or ‘*cái gì*’ (everything) as in (90e). In this case, the noun is not identified, and *cái* (inanimate) appears to have another function other than a classifier.

(90) a. thì không ai biết bình luận *cái* gì.
 then no who know comment CL(inani.) what
 ‘then no-one knows what to comment on.’ (O55.5439)
 b. phải biết nhìn *cái* gì đẹp mà đẩy lên,
 must know see CL(inani.) what beautiful so push up
 ‘must know what is beautiful to push up,’ (O125.9609)
 c. họ phải "trả giá" vì *cái* gì đó,
 they must pay price because CL(inani.) what that
 ‘they must “pay the price” for something,’ (O57.5699)
 d. Tôi thấy mình cần làm *cái* gì đó để cứu ba.
 I find I need do CL(inani.) what that to save father
 ‘I find that I need to do something to save my father.’ (O124.9557)

- e. người thứ ba với "cái gì cũng mới lạ"
 human third with CL(inani.) what also new strange
 'the third person with "everything" new and strange' (O92.7941).

Interestingly, *cái* (inanimate) appears with clauses as in (91a-b) and before a prepositional phrase as in (91c) with the omission of the noun which is previously identified in the context. In (91a), *cái* (inanimate) is used as a pro-form for the noun phrase mentioned in the subject preceding a clause functioning as a relative clause 'cái ông muốn nói đến' (what he wants to say). In (91b), it precedes a clause functioning as a *what* clause or a noun clause which serves as the subject of the sentence 'cái ta đạt được' (what we obtained). In (91c), *cái* (inanimate), preceding a prepositional phrase with 'mà' (for) and the gerund form of the verb 'spend', is used to refer to the noun 'salary'.

- (91) a. "chi phí giao dịch không chính thức" là cái ông muốn nói đến như một điểm trừ
 expense transaction not official be CL(inani.) he want say to like one point minus
 'unofficial "transaction expenses" are what he wants to say as a minus point' (O44.4047)
- b. Thế mà cái ta đạt được vẫn không làm lòng ta thỏa mãn và thanh thản.
 Though CL(inani.) we obtain get still not make soul we satisfied and relaxed
 'What we obtained still does not make our soul satisfied and relaxed though.' (O35.2864)
- c. Giáo viên nhận lương tháng một và hai trước Tết để "có cái mà chi tiêu".
 Teacher get salary January and February before Tet to have CL(inani.) for spend
 'Before Tet, the teachers get salaries for January and February which they have for spending.'
 (O50.4727).

In brief, *cái* (inanimate) is used for classifying, individuating nouns and/or nominalizing a number of verbs and adjectival verbs. It can go with verb phrases or clauses in the passive form in the construction of 'cái (inanimate) + 'được' (PASS) + Subject + Verb', which is likely to be similar as the structure 'what is/has been done by somebody' in English, and a *what* clause or a noun clause. It can appear with numerals and/or demonstratives with the omission of the classified noun which is identified in the context. Also, it combines with morphemes to create pronouns such as 'cái gì' (what), 'cái gì đó' (something), 'cái gì' (everything). It can also function as a pro-form for a noun phrase, as has been previously mentioned in the context.

5.3.3 The use of *cái* (inanimate) in the Spoken Corpus

As a very frequent classifier in the Spoken Corpus, *cái* (inanimate) classifies, individuates, and/or nominalizes over a thousand different nouns, verbs and adjectival verbs. It appears with verbs and nominalizes them such as *cái* hiểu biết (CL knowledge) and *cái* hy vọng (CL hope) as in (92a-b). It combines with adjectival verbs or phrases such as *cái* rất là thiết thực (very practical) as in (92c), and even with adjectives in comparatives or superlatives as *cái* nhỏ nhất (CL smallest) as in (92d). It can appear with numerals such as the plural morpheme ‘những’ as in (92a) and (92c) or the numeral một (one) as in (92d) or without a numeral as in (92b).

(92) a. dùng những *cái* hiểu biết của mình chia sẻ lại với con,

Use PL CL(inani.) understand know of oneself share again with children
‘use our knowledge to share with our children,’ (S21.6205)

b. Anh vẫn là *cái* hi vọng cuối cùng

English language be CL(inani.) hope last
‘English language is the last hope’ (S10.1373)

c. Thì họ đi những *cái* rất là thiết thực,

then they go PL CL(inani.) very be practical
‘then they brought the very practical things,’ (S16.225)

d. làm từ một *cái* nhỏ nhất

do from one CL(inani.) smallest
‘do from the one smallest’ (S8.8270).

Furthermore, the data shows that *cái* (inanimate) goes with verb phrases and adjective phrases with the omission of the classified noun as in (93a-b). In (93a), the combination of the numeral ‘one’ and the first occurrence of *cái* (inanimate) preceding the verb phrase of purpose ‘to help love more passionate and warmer’ refers to the clause following the VP. Thus, *cái* (inanimate) is a pro-form for the clause that is mentioned later in the sentence. This means *cái* (inanimate) functions as cataphoric reference in this case. This finding is very interesting and stands out as unexpected since this has never been shown in previous studies to date. In (93b), the combination of the numeral ‘one’ and the second occurrence of *cái* (inanimate) preceding the adjective phrase ‘worth to be proud of’ refers to the NP mentioned earlier in the subject of the sentence. In this case, *cái* (inanimate) functions as anaphoric reference.

- (93) a. anh nghĩ là thực ra một *cái* để giúp cho tình yêu nó nồng cháy, nó ấm áp hơn
 I think that actually one CL(inani.) to help for love it passionate it warm more
 chúng ta phải tìm cách tạo ra cho nhau những *cái* phút ấm áp, những phút lãng mạn như thế.
 we must find way make for each other PL CL(inani.) minute warm PL minute romantic so
 ‘I think that actually one thing to help love more passionate and warmer is that we have to
 find the way to make such romantic and warm minutes.’ (S15.4293)
- b. nhưng mà *cái* niềm đam mê của bạn ý đối với riêng môn nhạc thôi ạ
 but that CL(inani.) CL passion of him for with alone subject music then ah
 đã là một *cái* đáng để tự hào rồi ạ.
 PAST be one CL(inani.) worth to proud then ah
 ‘but his passion for music alone is the one thing worth being proud of.’ (S3.6541)

Additionally, *cái* (inanimate) combines with a large number of nouns or verbs in the construction ‘gọi là’ (called as) plus nouns or verbs as in (94). This classifier appears with the verb gọi là (called) plus the verb phát minh (invent) as in (94a) or plus the noun vinh quang (glory) as in (94b). This classifier can follow a plural morpheme such as ‘những’ as in (94a-b). This construction of *cái* (inanimate) + ‘gọi là’ (called as) + noun/verb is likely to be similar as the structure of ‘what is/are called + noun’ in English.

- (94) a. Tất cả những *cái* gọi là phát minh
 all PL CL(inani.) call be invent
 ‘All the called inventions or all what are called inventions’ (S9.8521)
- b. bên cạnh những *cái* tạm gọi là vinh quang
 beside PL CL(inani.) temporary call be glory
 ‘besides the temporarily called glories or besides what are temporarily called glories’
 (S20.5920).

The study finds that *cái* (inanimate) appears with numerals and demonstratives with the omission of the classified noun which has been identified in the preceding context as it does in the Online Newspaper Corpus. It is interesting to find that *cái* (inanimate) also goes with numbers as in (95a) and modal words ‘nên’ (should) and ‘không nên’ (should not) as in (95b). To be specific, it appears with the number indicating an amount of money ‘extra 50 million VND’ in (95a).

(95) a. và nói với P. S. là đưa *cái* 50 triệu dư cho anh ta.

and say with P S that give CL(inani.) 50 million extra for him

‘and asked to P. S. to give the extra 50 million VND for him.’ (S16.2)

b. Thì mình cứ lo lên thôi để cho mọi người thấy rằng à *cái* nên và

then we just speak out only for every human see that ah CL(inani.) should and

cái không nên, chứ còn tôi chả hy vọng gì *cái* chuyện là anh em mình

CL(inani.) not should but then I not hope what CL(inani.) matter that brothers us

có thể làm thay đổi thế giới được đâu.

can make change world get at all

‘then we just speak out for everyone to see that ah the ‘should’ and ‘should not’ things, but

I do not hope for the thing that we brothers can change the world at all.’ (S16.181)

More interestingly, *cái* (inanimate) appears with clauses with or without conjunction words ‘mà’ (which) or ‘là’ (that) in the Spoken Corpus. These clauses can be noun clauses or wh-clause in which the classifier functions as the head without any conjunction as in (96a-b). These nominal clauses modify the head in the subject of the sentences. This classifier also goes with nominal clauses with the conjunction word ‘mà’ (which) preceding the clause as in (96c-e) or ‘là’ (that) as in (96f). In this case, a numeral may or may not precede *cái* (inanimate). A numeral might be absent as in (96a-c), while the plural morpheme ‘những’ is present as in (96d-e), and the numeral một (one) in (96f). In this combination, *cái* (inanimate) functions as ‘what’ in the noun clause or what-clause in English. The classifier in these examples functions as a pro-form which is post-modified by the nominal clause, so they are definite.

(96) a. *cái* em lo sợ là khi chị tập trung vào thời gian đi ca hát thì chị sẽ không

CL(inani.) I anxious be when you concentrate on time go sing then you will not

có thời gian lo cho gia đình.

have time spend for family

‘what I am anxious for is that when you concentrate on going for singing, then you will not

have time for family.’ (S18.4704)

b. Bởi vì *cái* em viết

Because CL(inani.) I write

‘Because what I wrote’ (S2.5068)

c. tìm được *cái* mà mình yêu thích.

find get CL(inani.) which we love

‘find what we love.’ (S17.772)

d. thì dựa trên những *cái* mà con chia sẻ như vậy

then base on PL CL(inani.) which they share so

‘then based on the things that they shared/ then based on what they shared’ (S21.6252)

e. mình làm những *cái* mà không nhất thiết là lựa chọn của mình,

we do PL CL(inani.) which not necessary be choice of us

‘we do the things which are not necessarily our choice,’ (S22.1070)

f. Thảo Nguyên thì được một *cái* là em rất là chịu khó.

Thao Nguyen then get one CL(inani.) be she very be diligent

‘Thao Nguyen has one thing that she is very diligent.’ (S14.3837)

In addition, *cái* (inanimate) occurs with idioms such as *lời ra tiếng vào* (words out words in) as in (97a). It even appears with a number of English words such as ‘mini concert’ or ‘livestream’ as in (97b-c).

(97) a. mấy *cái* lời ra tiếng vào nó nhiều lắm.

some CL(inani.) word out word in it much so

‘rumours are a lot.’ (S13.3142)

b. sau đó lại quay trở thành một *cái* mini concert

after that come film become one CL(inani.) mini concert

‘after that we filmed it into the one mini concert’ (S13.3162)

c. nhiều người đang xem *cái* livestream của chúng ta nè.

many people PROG watch CL(inani.) livestream of us yeh

‘many people are watching our livestream yeh.’ (S20.5731).

In sum, *cái* (inanimate) can be used in many different structures, especially in the Spoken Corpus, since it can go with a great variety of different parts of speech including nouns (even English nouns), verbs or verb phrases, adjectives or adjective phrases, numbers, modal words, idioms, and noun clauses with or without conjunction words. This classifier can occur with numerals and/or demonstratives with the omission of the classified noun when the noun is identified in the preceding context. Furthermore, *cái* (inanimate) can be constructed with a specific

classifier in the classifier doubling construction to emphasize the noun and the definiteness of the noun as H. T. Nguyen (2004, 2013) and Simpson and Ngo (2018) discuss. The rich combination possibilities of *cái* (inanimate) in different constructions with various parts of speech may explain its very frequent use in Vietnamese conversations. However, I argue that the frequent use of *cái* (inanimate) may be due to its use for emphasis in spoken discourse since it can appear with non-classified nouns or optional-classifier nouns which do not require a classifier in Vietnamese. This is discussed in the next section.

5.3.4 *The use of cái (inanimate) with non-classified nouns*

In this section, I will discuss a number of cases in which *cái* (inani.) appears with non-classified nouns or optional-classifier nouns. As reviewed in 2.5.3, *cái* (inani.) can go with non-classified nouns, optional-classifier nouns or co-occur with a specific classifier plus classified nouns in the classifier doubling construction (Emeneau 1951; D. H. Nguyen 1957; H. T. Nguyen 2013; and Simpson and Ngo 2018). Since the presence of *cái* (inani.) in these cases is optional, it is called an extra *cái* (inanimate), or even the “definite extra *cái* (inani.)” by Simpson and Ngo (2018:224). According to D. H. Nguyen (1957) and other researchers, nouns indicating ‘districts, cities’, or time such as ‘days, weeks, months’ in Vietnamese are non-classified nouns, which do not take classifiers. Nevertheless, these nouns appear with *cái* (inani.) in the Spoken Corpus. It goes with the non-classified noun ‘huyện’ (district) as in (98a). The noun is definite and emphasized in the presence of the name of the district ‘Kim Boi’. Similarly, *cái* (inani.) is found with the noun ‘khu vực’ (area) as in (98b). The possessive ‘its’ makes the noun definite, while the noun is emphasized by the appearance of *cái* (inani.).

- (98) a. được phân công về một *cái* huyện miền núi Kim Bôi giảng dạy
 PASS assign to one CL(inani.) district region mountain Kim Boi teach
 ‘(I) was assigned to teach at the one mountainous district of Kim Boi’ (S7.7978)
- b. *cái* khu vực của nó
 CL(inani.) area of it
 ‘the area of it’ (S10.1407).

The non-classified nouns indicating time such as ‘day, hour, minute, year’ are found with *cái* (inani.) as in (99a-k). The noun ngày (day) appears with *cái* (inani.) as in (99a-b). A numeral

may be present such as the plural morpheme ‘những’ as in (99a) or may not as in (99b). This noun is definite in the presence of the post-modifier ‘initial’ in (99a) and the defining clause in (99b). Similarly, *cái* (inani.) appears with ‘giây phút’ (second minute) as in (99c), and ‘năm’ (year) as in (99d-f). It precedes the noun ‘year’ with numerals such as the plural morpheme ‘những’ as in (99d) or ‘một’ (one) as in (99e), or without a numeral as in (99f). The combination of *cái* (inani.) and non-classified nouns may precede a demonstrative such as ‘đó’ (that) as in (99d) and (99f) or a defining clause as in (99e). In short, non-classified nouns indicating ‘district, area’ or ‘day, year, hour, minute’ can go with *cái* (inani.). The examples given provide evidence to support previous researchers’ argument that *cái* (inani.) appears with non-classified nouns for emphasis (D. H. Nguyen 1957; Diep 2005; H. T. Nguyen 2013).

- (99) a. Những *cái* ngày đầu, ấn tượng của mọi người về em
 PL CL(inani.) day initial impression of every human about me
 ‘For the initial days, everyone’s impression about me’ (S20.5800)
- b. Như vậy *cái* ngày mà phát hành *cái* thẻ,
 So CL(inani.) day which issue CL(inani.) card
 ‘So the day when the card was issued,’ (S9.8674)
- c. cảm ơn những *cái* giây phút ấy.
 thank PL CL(inani.) second minute that
 ‘thanks for those minutes.’ (S1.1150)
- d. Tôi nhớ là những *cái* năm 70 đó
 I remember that PL CL(inani.) year 70 that
 ‘I remember that those years of 70s’ (S16.147)
- e. và đây là một *cái* năm mà nó chuyển giao
 and here be one CL(inani.) year which it transfer
 ‘and this is the one year that it has transferred’ (S13.3190)
- f. Em thì thật ra em nhớ *cái* năm đó
 I then actually I miss CL(inani.) year that
 ‘I, then actually I miss that year’ (S20.5787).

Similarly, other non-classified nouns appear with or without *cái* (inani.) in the Spoken Corpus. These nouns can be concrete nouns such as hình ảnh (image) or abstract nouns such as tên

(name), *cách sống* (way of life, lifestyle), and *bí quyết* (secret). The noun ‘*hình ảnh*’ (image) goes with *cái* (inani.) as in (100a) or without *cái* (inanimate) as in (100b). This noun appears with *cái* (inani.) 10 times and without it 41 times in the Spoken Corpus, while it appears without a classifier once in the narratives and 35 times in the newspapers as in (100c). This evidence reveals that the noun *hình ảnh* (image) is a non-classified noun and does not appear with a classifier in the written corpora, while it is used with or without *cái* (inani.) in the spoken corpus. Thus, the use of *cái* (inani.) is not obligatory in this case, depending on the speaker’s intention and choice.

(100) a. *mình nên giữ cái hình ảnh của mình.*

we should keep CL(inani.) image of self

‘we should keep our image.’ (S4.6905)

b. *Chị vẫn rất nhớ hình ảnh Thùy Dung ngồi đàn piano*

I still very miss image Thuy Dung sit play piano

‘I still miss the image of Thuy Dung playing piano very much’ (S12.2413)

c. *Hình ảnh của anh làm tôi ngỡ ngàng.*

image of him make me surprised

‘His image made me surprised.’ (O26.1620).

It is similar for other non-classified nouns or optional-classifier nouns including *tên* (name), *cách sống* (way of life, lifestyle), and *bí quyết* (secret). These nouns appear with or without *cái* (inanimate) in the Spoken Corpus. The noun *tên* (name) appears without a classifier as in (101a) and with *cái* (inanimate) as in (101b). The nouns *cách sống* (way of life, lifestyle) and *bí quyết* (secret) appear without a classifier as in (102a) and (103a), and with *cái* (inanimate) as in (102b) and (103b) respectively. The data shows that an extra *cái* (inanimate) in this case performs the function of emphasizing the noun rather than its main function of classification and individualization as it appears with non-classified nouns which do not require a classifier.

(101) a. *Rồi giới thiệu tên đi.*

Then introduce name go

‘Then introduce (your) name.’ (S18.4610)

b. *có thể biết nhiều hơn là à tại sao lại có cái tên như vậy.*

can know much more be ah why again have CL(inani.) name such

‘can know more than ah why you have such a name.’ (S13.3320)

- (102) a. Thực ra đây là một cách sống thông minh
 Actually that be one way live clever
 ‘Actually, that is a clever lifestyle’ (S12.2269)
- b. Nó trở thành một *cái* cách sống rất là tuyệt vời
 It become one CL(inani.) way live very be wonderful
 ‘It becomes the very wonderful lifestyle’ (S8.8162)
- (103) a. ông đang nói ra một bí quyết
 you CONT speak out one secret
 ‘you are speaking out a secret’ (S9.8387)
- b. sẽ cùng nhớ lại những *cái* bí quyết
 will together remind PL CL(inani.) secret
 ‘will together remind of the secrets’ (S10.1171).

It is interesting that a number of English count nouns appear with *cái* (inani.) in the Spoken Corpus including ‘clip, format, mini show, resort’. These nouns are definite due to the presence of the demonstrative *này* (this) in (104a) or previous mention in (104b-d). Clearly, they do not require a classifier, but do appear with *cái* (inani.) which is phonologically stressed. Thus, *cái* (inani.) is used to emphasize with these ‘borrowed’ nouns. It may appear without any numerals as in (104a-b) and (104d) or with a numeral such as a plural morpheme *nhều* *những* (many PL) as in (104c).

- (104) a. và em xem lại *cái* clip này,
 and I watch again CL(inani.) clip this
 ‘and I watched this clip again,’ (S19.4791)
- b. thông báo *cái* format của chương trình
 inform CL(inani.) format of program
 ‘informed the format of the program’ (S19.4853)
- c. dù là Tường có rất là *nhều* *những cái* mini show
 although be Tuong have very be many PL CL(inani.) mini show
 ‘although Tuong/I have many mini shows’ (S13.3163)
- d. có độ khoảng hai chục người vào *cái* resort ở Mũi Né.
 have about two ten human come CL(inani.) resort in Mui Ne
 ‘there are about twenty people came in the resort in Mui Ne.’ (S16.204).

In sum, a number of non-classified nouns in Vietnamese that appear with or without *cái* (inanimate) in the Spoken Corpus have been analysed. Table 12 shows the non-classified nouns that have been analysed above with the number of tokens and percentage of their appearance with or without *cái* (inanimate) in the Spoken Corpus for comparison. Due to the time limit, I could not explore all the non-classified nouns that appear with an extra *cái* (inanimate) in the corpus.

Table 12: Comparison of non-classified nouns with or without cái (inanimate)

Nouns	Appearing with an extra <i>cái</i> (inani.)		Appearing without an extra <i>cái</i> (inani.)	
	No. of tokens	%	No. of tokens	%
điều (thing)	96	2.22	130	3.01
ngày (day)	18	0.42	24	0.55
hình ảnh (image)	10	0.23	41	0.95
tên (name)	10	0.23	11	0.25
năm (year)	8	0.18	66	1.53
giây phút (moment)	5	0.12	5	0.12
phút (minute)	5	0.12	26	0.6
cách sống (lifestyle)	4	0.09	3	0.07
bí quyết (secret)	2	0.05	2	0.05
khu vực (region)	2	0.05	0	0
resort (resort)	2	0.05	0	0
clip (clip)	1	0.02	3	0.07
format (format)	1	0.02	0	0
huyện (district)	1	0.02	0	0
mini show (mini show)	1	0.02	3	0.07

As analysed in section 5.1.1, *cái* (inanimate) used with non-classified nouns functions as emphatics. While these nouns are used quite often with the extra *cái* (inanimate) in the Spoken Corpus, they do not appear with it in the narrative and online newspaper corpora. This leads to the assumption that the use of *cái* (inanimate) with non-classified nouns results in the higher classifier frequency in the Spoken Corpus as discussed in section 5.1.1.1. Table 12 shows that non-classified

nouns going with or without an extra *cái* (inanimate) include nouns indicating ‘thing, image, minute, day, year’, which appear without *cái* (inanimate) more often than with *cái* (inanimate). Nouns indicating ‘district, region’ and English loaned nouns including ‘format, resort’ just appear with *cái* (inanimate) in the spoken corpus. For other nouns such as ‘name, lifestyle’, the number of their occurrences with *cái* (inanimate) is almost the same as the number of tokens without *cái* (inanimate). This evidence reveals that the use of *cái* (inanimate) with non-classified nouns in Vietnamese depends on the speaker’s attention. The data shows that the appearance of *cái* (inanimate) with non-classified nouns is optional and is not dependent on any linguistic or discourse factor. All these nouns appearing with *cái* (inanimate) in the spoken corpus are definite in the presence of a demonstrative or a possessive or previous mention in the context. In the absence of *cái* (inanimate), these nouns may be indefinite or definite, depending on the presence of a demonstrative, a possessive, a determiner or previous mention in the context. Thus, *cái* (inanimate) in this case performs the function of emphasizing the nouns. However, there is not enough evidence to claim that the extra *cái* (inanimate) makes the noun definite as claimed by researchers including H. T. Nguyen (2013), and Simpson and Ngo (2018). As discussed in section 5.1.1, Vietnamese speakers use *cái* (inanimate) with non-classified nouns for emphasis in their speech, but not in writing.

5.3.5 Summary

In sum, *cái* (inanimate) can go with different parts of speech including nouns, verbs, adjectival verbs, verb phrases or clauses in Vietnamese. Specifically, it appears with non-classified nouns which do not require a classifier. It is also constructed with a specific classifier before classified nouns in the classifier doubling construction. It is worth noting that *cái* (inanimate) is used with non-classified nouns and in the double classifier construction mostly in the spoken corpus, but hardly ever in the written corpora. In this case, *cái* (inanimate) is argued to emphasize the noun and making the noun definite in previous research (H. T. Nguyen 2004, 2013; and Simpson and Ngo 2018). This study provides evidence that *cái* (inanimate) in combination with non-classified nouns or constructed with another classifier functions as an emphatic. However, there is not enough evidence to ascertain that it forces the noun definite as some researchers argued. As analysed in sections 5.3.1, 5.3.2, and 5.3.3, *cái* (inanimate) not only functions as a classifier in Vietnamese but also performs other functions when combining with other words in various

constructions. It appears to undergo the process of grammaticalization as it is generalized in function and meaning, according to Hopper and Traugott (2003). The grammaticalization of *cái* (inanimate) is an interesting issue, but this is beyond the focus of this study and would be left for future research. Table 13 shows different constructions of *cái* (inanimate) in the three corpora of this study and its functions.

Table 13: Constructions of cái (inanimate) in the study

<i>cái</i> (inani.) constructions	Functions
<i>cái</i> (inani.) + classified nouns	classification and individualization
<i>cái</i> (inani.) + non-classified nouns	emphatic
<i>cái</i> (inani.) + CL + classified nouns	emphatic
<i>cái</i> (inani.) + verbs	class., individualization, and nominalization
<i>cái</i> (inani.) + adjectival verbs	class., individualization, and nominalization
<i>cái</i> (inani.) + demonstrative	classification and individualization
numeral + <i>cái</i> (inani.) + (demonstrative)	classification and individualization
<i>cái</i> (inani.) + gì (interrogative)	what
<i>cái</i> (inani.) + nào (which)	which
<i>cái</i> (inani.) + được (passive) + verb	classification and individualization
<i>cái</i> (inani.) + gì (negative)	anything
<i>cái</i> (inani.) + gì đó (interrogative)	something
<i>cái</i> (inani.) + clause (nominal/wh-clause)	classification and individualization
<i>cái</i> (inani.) + to-infinitive VP	class., individualization, and nominalization
<i>cái</i> (inani.) + AdjP (comparative/superlative)	class., individualization, and nominalization
<i>cái</i> (inani.) + gọi là (called as) + N/nominal	class., individualization, and/or nominalization
<i>cái</i> (inani.) + number (amount of money)	classification and individualization
<i>cái</i> (inani.) + mà/là (which/that) + clause	class., individualization, and nominalization
<i>cái</i> (inani.) + idiom	class., individualization, and nominalization
<i>cái</i> (inani.) + nên (should)	class., individualization, and nominalization
<i>cái</i> (inani.) + không nên (should not)	class., individualization, and nominalization
<i>cái</i> (inani.) + English count nouns	classification and emphatic

In addition to the two major functions of Vietnamese classifiers, classification and individualization, claimed by Bisang (1999), it is interesting that this study found new evidence that as an inanimate classifier, *cái* (inanimate) functions as cataphoric reference in Vietnamese. This function has not yet been mentioned by researchers to date although Bisang (1999) discuss that Vietnamese classifiers function as anaphoric reference.

Furthermore, the analysis of *cái* (inanimate) used in the three corpora in the previous sections reveals that *cái* (inanimate) appears in various constructions, performing the functions of a classifier as well as other functions. It combines with several other words to create wh-words or pronouns such as ‘what, which, something, anything’. Specifically, in the Spoken Corpus, it goes with different classes of words including numbers, modal words, idioms, and noun clauses with or without conjunction words, performing the function of classification, individualization, and/or nominalization. Surprisingly, a number of English nouns are also found with *cái* (inanimate) in the Spoken Corpus while these nouns are count nouns and definitely do not require a classifier to be individuated and counted. The appearance of *cái* (inanimate) with non-classified nouns emphasizes these English nouns to attract the listener’s attention to their speech. It appears that language contact influences the use of classifiers with ‘borrowed’ English nouns. In this case, Vietnamese speakers treat these English nouns as Vietnamese nouns.

Table 13 lists 22 constructions of *cái* (inanimate) used in the three corpora altogether. It is interesting that the study found a variety of inanimate classifiers combine with verbs and adjectival verbs in the corpora. They classify, nominalize and individuate these verbs and adjectival verbs. These classifiers will be discussed in detail in the next section.

5.4 Analysis of classifiers with the nominalization function

The data shows that a number of classifiers function as nominalizers, which appear quite often in the Online Newspaper and Spoken Corpora although they are infrequent in the Narrative Corpus. However, as far as I know, not much attention has been paid to this group of classifiers although it is discussed in a few studies including H. T. Nguyen (2004, 2013), T. T. Hoang (1996), and T. B. N. Nguyen (2013). In this section, I will analyse and discuss a number of inanimate classifiers having this function in the corpora of this study. Classifiers with the function of nominalization usually classify and nominalize verbs indicating action or process or adjectival

verbs, turning them into nominals, and/or individuate them. As discussed in section 5.1.2, beyond *cái* (inanimate), the two most frequent classifiers in the online newspaper and spoken corpora are *sự* (event) and *cuộc* (life, strike, match) although they rarely appear in the narrative corpus. These two classifiers function as nominalizers in Vietnamese.

As the most frequent classifier in the Online Newspaper Corpus, *sự* (event) classifies, nominalizes and/or individualizes 192 different verbs and adjectival verbs, and 86 different verbs and adjectival verbs in the Spoken Corpus. However, it appears only 6 times with 6 different stative or adjectival verbs in the Narrative Corpus. In the Online Newspaper Corpus, *sự* (event) is the most frequent, so the number of ‘nominalized’ entities it goes with exceeds the number of inanimate nouns used with the general classifier *cái* (inanimate). Specifically, *cái* (inanimate) classifies 110 different nouns in this corpus. This means that with the main function of nominalization, *sự* (event) combines with a great variety of verbs and adjectival verbs in the online newspaper and spoken corpora. It appears with stative verbs including *thành công* (succeed) and *hợp tác* (cooperate) as in (105a-b). It nominalizes these verbs and turns them into the nouns *sự thành công* (success) and *sự hợp tác* (cooperation). Similarly, it classifies and nominalizes adjectival verbs such as *buồn bã* (sad), *chuyên nghiệp* (professional), *chu đáo* (thoughtful), and *im lặng* (silent) as in (105c-e), and turning them into the nouns ‘sorrow, professionalism, thoughtfulness, and silence’ respectively.

- (105) a. quyết định *sự* thành công của Việt Nam
 decide CL(event) succeed of Vietnam
 ‘decides the success of Vietnam’ (O26.1773)
- b. *sự* hợp tác quay quảng cáo lần đó không thành
 CL(event) cooperate video commercials time that not succeed
 ‘the cooperation for videoing commercials that time did not succeed’ (O80.7315)
- c. không nhận ra được *sự* buồn bã không có lý do,
 not realize get CL(event) sorrowful without reason
 ‘do not realize the sorrow without a reason,’ (O47.4426)
- d. *Sự* chuyên nghiệp và chu đáo của chính quyền Hàn Quốc
 CL(event) professional and thoughtful of government South Korea
 ‘The professionalism and thoughtfulness of the government of South Korea’ (O102.8502)

- e. khích lệ tôi bằng *sự* im lặng hàm nghĩa đồng tình.
 encourage me by CL(event) silent mean consent
 ‘encouraged me by the silence meaning consent.’ (O131.9978).

Like *sự* (event), *cuộc* (life, strike) also functions as a nominalizer in Vietnamese. It appears with 53 different verbs in the online newspaper corpus and with only 13 verbs in the spoken corpus although it is frequently used in both of these corpora. This is because the nominals it goes with in the spoken corpus is repeatedly used due to the topics of the talk shows, while the online newspaper corpus covers a wider variety of topics. Meanwhile, *cuộc* (life, strike) is not frequent and appears with 11 different nominals in the narrative corpus. Altogether, it classifies, nominalizes and individualizes 61 nouns or verbs in the three corpora. It appears with the verbs thi (compete), trò chuyện (talk), and điều tra (investigate) as in (106a-c), nominalizing and individualizing them. Furthermore, it combines with the noun đời (life), classifying and individualizing it as in (106d).

- (106) a. lúc mà kết thúc *cuộc* thi
 moment that end CL compete
 ‘the moment that ended the competition’ (S13.3334)
- b. giờ chúng ta sẽ bắt đầu *cuộc* trò chuyện.
 now we will begin CL talk
 ‘Now let us begin the talk’ (S15.4262)
- c. Các cuộc điều tra đang tiếp tục
 PL CL investigate PROG continue
 ‘The investigations are continuing to’ (O17.992)
- d. Một *cuộc* đời quá buồn tẻ.
 one CL life too boring
 ‘A very boring life’ (S2.5060).

Both *sự* (event) and *cuộc* (life, strike) normally classify, nominalize and/or individualize verbs or adjectival verbs. However, *sự* (event) mainly appears with stative verbs and adjectival verbs, while *cuộc* (life, strike) usually combines with action verbs. The nouns/nominals with *sự* (event) usually refer to concepts, definitions, perceptions, or state, while the nouns/nominals with *cuộc* (life, strike) often indicate a process, something that takes place in a period of time such as ‘competition, talk, investigation’ as in (106a-c), or something like ‘an outing’, or ‘a walk’. It is

interesting that functioning as nominalizers, they are used more frequently in the newspapers and spoken corpora compared to narratives. They appear to become an effective tool to nominalize verbs and adjectival verbs into nouns in Vietnamese, which contributes to the change and development of this language.

In addition to them, *việc* (activity) is also a frequent classifier functioning as a nominalizer in the Online Newspaper Corpus although it is rarely used in the Narrative and Spoken Corpora. It appears 116 times in the newspaper, while it is only twice in the narrative and six times in the spoken corpus. It is interesting to find that *việc* (activity) nominalizes 95 different verbs, mainly action or process verbs in the Online Newspaper Corpus. As in (107a), appearing with the VP ‘học ngoại ngữ’ (learn foreign languages), *việc* (activity) turns it into the NP *việc học ngoại ngữ* (the learning of foreign languages). Similarly, combining with the verbs *chăm sóc* (take care of), *giáo dục* (educate), *thay đổi* (change), *sử dụng* (use), and *khai thác* (exploit) as in (107b-e), it nominalizes and turns them into nouns or nominals. It is important to note that this classifier does not appear with a numeral in all cases in the corpus. It is likely to indicate the job of doing something rather than individualizing the nominals.

(107) a. *việc học ngoại ngữ sẽ là rất ưu việt đối với con.*

CL(activity) study foreign language will be very preeminent for with him

‘the learning of foreign languages will be very preeminent for him.’ (O67.6541)

b. *việc chăm sóc một gia đình lớn không dễ dàng.*

CL(activity) take care of one family big not easy

‘the taking care of a big family is not easy.’ (O4.208)

c. *trong việc giáo dục con cái*

in CL(activity) educate children

‘in the education of children’ (O6.274)

d. *và việc thay đổi này mở ra cho mọi người,*

and CL(activity) change this open out for every human

‘and this change is for everyone,’ (O46.4322)

e. *kết hợp với việc sử dụng và khai thác tài nguyên nước trong lưu vực*

coordinate with CL(activity) use and exploit resource water in basin

‘coordinating with the use and exploitation of water resources in the basin’ (O61.6132).

Furthermore, other classifiers also function as nominalizers including *nỗi* (feeling, worry, sad) and *niềm* (sentiment), which are usually used to describe human feelings. They are used less often than the three frequent classifiers analysed above. Specifically, with 35 tokens, *nỗi* (feeling, worry, sad) goes with 12 different adjectival verbs in the Online Newspaper Corpus, while it appears 12 times in the spoken, but only once in the narrative corpus. It usually nominalizes adjectives or adjectival verbs indicating emotional states with negative meanings, for example, *khiếp sợ* (terrified), *buồn* (sad), *lo lắng* (worried) as in (108a-b). It may appear without a numeral as in (108a) and with a numeral as in (108b-c). Furthermore, *nỗi* (feeling, worry, sad) appears with *cái* (inanimate), with 8 tokens, in the classifier doubling construction *cái nỗi* (inanimate, feeling, worry) as in (108c) in the Spoken Corpus.

- (108) a. là *nỗi* *khiếp sợ* của toàn nhân loại
 be CL(feeling) terrified of all humankind
 ‘is the terror of the humankind’ (O28.1949)
- b. anh không bao giờ để một *nỗi* *buồn* kéo dài quá 24 giờ đồng hồ,
 he never let one CL(feeling) sad last long over 24 hour clock
 ‘he never lets a sorrow last for over 24 hours,’ (S8.8151)
- c. thế nhưng mà nó vẫn cứ ám ảnh một *cái* *nỗi* lo
 then but which it still obsessed one CL(inani.) CL(feeling) worry
 ‘but then it is still obsessed with the one worry’ (S16.269).

In contrast, *niềm* (sentiment) usually nominalizes stative or adjectival verbs indicating emotional states with positive meanings. With 29 tokens, this classifier goes with 8 different adjectives/adjectival verbs in the Online Newspaper Corpus, while it appears 27 times in the Spoken Corpus, but none in the Narrative Corpus. It usually goes with adjectives/adjectival verbs such as *vui sướng* (joy), *tin* (trust), *tự hào* (proud), *đam mê* (passionate) as in (109a-d). Also, *niềm* (sentiment) combines with *cái* (inanimate) in the classifier doubling construction *cái niềm* (inani., sentiment) as in (109d). It may appear with a numeral such as *một* (one), *nhiều* (many) or *cả hai* (both two) as in (109a-b) and (109d) or without a numeral as in (109c).

- (109) a. ai cũng long lanh một *niềm* *vui sướng*,
 who also sparkling one CL(sentiment) happy
 ‘everyone is sparkling with happiness,’ (O122.9423)

- b. sẽ mang lại rất nhiều *niềm* tin
 will bring back very many CL(sentiment) trust
 ‘will bring a lot of trust’ (S12.2560)
- c. nó là *niềm* tự hào
 it be CL(sentiment) proud
 ‘it is the pride’ (S5.7528)
- d. bản thân em có thể theo đuổi cả hai *cái* *niềm* đam mê của mình.
 self I can pursue both CL(inani.) CL(sentiment) passionate of self
 ‘I myself can pursue both of my passions.’ (S17.531).

In sum, I have analysed and discussed some of the inanimate classifiers functioning as nominalizers in Vietnamese. It is interesting to find that most of them appear quite frequently in the online newspaper and spoken corpora, but not often in the narrative corpus. This difference might be due to the content of texts and discourse and different topics in the three genres as well as individual speakers. More importantly, the three classifiers, *sự* (event), *cuộc* (life, strike), and *việc* (activity) are more frequently used in the concurrent newspaper and spoken corpora compared to the narrative. This finding suggests that there may be some changes in language use over time. The study found that *sự* (event), *cuộc* (life, strike), and *việc* (activity) can combine with a wide variety of verbs and adjectival verbs in the online newspaper and spoken corpora. While *sự* (event) mainly goes with stative and adjective verbs to form nouns/nominals referring to concepts, definitions, or states, *cuộc* (life, strike) is usually used with action verbs to create nouns/nominals to indicate a process or something that prolongs for a period of time or for life. However, *việc* (activity) normally appears with action verbs to form nouns/nominals indicating the job of doing something. Interestingly, the study has identified a pair of two classifiers with the nominalization function often used for expressing feelings from different extremes. While *nỗi* (feeling, worry, sad) classifies and nominalizes emotional adjectival verbs with negative meanings, *niềm* (sentiment) appears with emotional adjectival verbs with positive meanings. In short, all the classifiers functioning as nominalizers become a useful tool for forming nouns/nominals in Vietnamese. They help Vietnamese speakers to create more nouns/nominals for easier communication and expressions of new ideas and feelings as a response to the changing world nowadays. The frequent use of *sự* (event), *cuộc* (life, strike), and *việc* (activity) in the newspaper and spoken corpora leads to the assumption that language change is in progress in Vietnamese.

5.5 Vietnamese classifier constructions and other issues

The above sections have discussed all the major findings of this study. This section discusses other findings in the research. Section 5.5.1 discusses classifier constructions in Vietnamese. The definiteness of the noun is discussed in section 5.5.2. Section 5.5.3 summarizes the main points.

5.5.1 Vietnamese classifier constructions

As reviewed in the literature, previous researchers claim that the most typical structure of Vietnamese classifier constructions is Numeral - Classifier - Noun (D. H. Nguyen 1957; Thompson 1965; Aikhenvald 2000). However, the data in this study shows that Classifier - Noun would be potentially the prototypical classifier pattern since over 63% of the classifier tokens found in this study follow this pattern. This evidence also supports Daley (1998)'s and Tran (2018)'s suggestions. Table 14 shows the major classifier constructions with numbers of tokens and percentage found in the study. As shown in the table, the number of tokens following the pattern without a numeral is double of the number of tokens with a numeral.

With the naturalistic data of this corpus-based study, it suggests that the Numeral - Classifier - Noun construction might not be the typical pattern of Vietnamese classifier phrases because it accounts for only 27% of all the tokens found in the corpora. The Vietnamese classifier construction claimed in previous research may be the general one which represents all the possible constituents, but not the prototypical pattern.

Table 14: Major classifier constructions in the study

Classifier constructions	Number of tokens	%
Num + CL + N + (Attri.)	2343	27.16
CL + N + (Attri.)	5499	63.75
Other constructions	784	9.09
Total	8626	100.00

In addition to these two major classifier constructions identified in this study as shown in Table 14, classifiers in Vietnamese appear in other constructions, which account for 9% of all the

tokens in the three corpora. As mentioned in the literature, the classifier in Vietnamese can occur with a numeral and/or a demonstrative with the omission of the classified noun when the noun is identified in the preceding context (Emeneau 1951; Thompson 1965). For one thing, it means that the classifier can act as a kind of “pro-form”. It stands in as a kind of nominals for the more specific noun, but it has more information than does a simple pronoun.

As analysed and discussed in section 5.3, Vietnamese classifiers, especially *cái* (inani.), can be used in a number of constructions. That means, *cái* (inani.) can combine with different parts of speech in various structural constructions. Different classifier constructions in each of the three corpora have been analysed and listed with the number of tokens that appear in each of the constructions. However, this is beyond the focus of this study and it is a long list, so it is placed in the Appendix E for reference. Despite different numbers of constructions, a larger number of classifier constructions are used in the Spoken Corpus than in the Narrative and Online Newspaper Corpora. The data reveals that in several classifier constructions, a possessive, a demonstrative, or an ordinal number can appear with a single classifier or a double classifier with the omission of the head noun when the noun is identified. In these cases, the entity is individuated due to the appearance of the classifier while the presence of a possessive, a demonstrative, or an ordinal number makes the noun definite. Furthermore, *cái* (inani.) can appear with a number of morphemes to form wh-words such as ‘what’, ‘which’, and pronouns such as ‘something’, ‘anything’, or ‘nothing’. In these cases, the noun is indefinite regardless of the presence of a numeral.

In this study, when the classified noun is identified in the context, the classifier appears mainly with numerals and/or demonstratives with the omission of the noun as shown in Table 15.

Table 15: Classifier constructions without the head noun in the three corpora

Corpus	Narrative		Online Newspaper		Spoken	
CL Constructions	No. of tokens	%	No. of tokens	%	No. of tokens	%
Num + CL	33	1.81	10	0.40	39	0.90
(Num) + CL + Dem	28	1.53	13	0.53	209	4.83

These classifier constructions appear most often in the Spoken Corpus with about 5% of all the classifier tokens in the corpus. In these constructions, due to the appearance of the classifier, the entity is individuated. The definiteness of the noun depends on the presence or absence of the demonstrative or other determining factors in the phrase or previous mention, not on the appearance of the numeral. That means, the presence or absence of the numeral does not influence the definiteness of the noun.

In sum, the findings of this study show that there are various classifier constructions in the three corpora, which makes the Vietnamese classifier system more complex. The classifier pattern of CL - N with over 63% of all the classifier tokens identified in the corpora of this study would potentially be the prototypical classifier construction in Vietnamese. This result is in line with the findings in Daley (1998)'s and Tran (2018)'s studies. The numeral classifier construction Numeral - Classifier - Noun claimed in prior research is more general with all three major possible constituents of the classifier phrase in Vietnamese, but it might not be the typical one. The classifiers in Vietnamese can appear with numerals and/or demonstratives, ordinal numbers or possessives with the omission of the head noun when the noun is identified in the context. However, not all inanimate classifiers in Vietnamese can appear in these constructions. Normally, *cái* (inani.) is used in these constructions as well as other constructions as it can combine with many different parts of speech and morphemes/words. The data of the study shows evidence that *cái* (inani.) is undergoing grammaticalization. However, this is not the focus of the current study and would be left for future research.

5.5.2 Definiteness of the noun

The data in this study supports Bisang (1999)'s discussion that the definiteness of the noun is not determined by the occurrence of the classifier since in the Classifier - Noun construction, the noun can be either indefinite or definite depending on the context. I should specify clearly that an inanimate classifier in Vietnamese does not function as a determiner for the noun it precedes. The data shows that in the Classifier - Noun construction, the definiteness of the noun depends on the context. That means, the previous mention of the noun in the context and/or the presence of a demonstrative, a possessive, an ordinal number, or a defining clause that follows the noun clearly decides the definiteness of the noun. When the noun is previously identified in the context, the noun phrase in which the classifier preceding either a demonstrative, a possessive, an ordinal

number, or a defining clause with the omission of the noun is definite. In the noun/nominal phrase with or without the omission of the noun when it is previously identified, the presence of a numeral does not influence the definiteness of the phrase. In short, the definiteness of the noun is determined by previous mention of the noun and/or by the presence of either a demonstrative, a possessive, an ordinal number, or a defining clause in the post-modification in the Vietnamese noun phrase. It is not decided by the presence of the classifier.

In the classifier doubling construction, the extra *cái* (inanimate) that appears with a specific classifier plus a classified noun is argued to force “interpretations of definiteness” by Simpson and Ngo (2018:224). However, the examples given in their analysis is believed to be elicited utterances which do not show the context. Thus, it is hard to ascertain whether the extra *cái* (inanimate) or its context determines the definiteness of the noun. However, the data of this corpus-based study reveals that even in the classifier doubling construction, it is not the extra *cái* (inanimate) that determines the definiteness of the noun, but the context does.

As this issue is beyond the focus of the study, I just have some comments on this but do not discuss it in detail. In brief, the findings show that a classifier in Vietnamese does not determine the definiteness of the noun regardless of constructions it appears in. The presence of one of these constituents including a demonstrative, a possessive, an ordinal number, or a defining clause and/or previous mention in the context determines the definiteness of the noun. The extra *cái* (inanimate) which appears with a non-classified noun or with a specific classifier plus a classified noun in the classifier doubling construction is assumed to function as an emphatic.

5.5.3 Summary

What I have discussed in this section is the typical classifier construction in Vietnamese and the factors that determine the definiteness of the noun. The results of this research show that there is variation in classifier constructions in Vietnamese as a number of different classifier constructions are used in each of the corpora. The data in this study suggests that the typical classifier construction would probably be Classifier - Noun, not the Numeral - Classifier - Noun as claimed by researchers including Emeneau (1951), Thompson (1965), and D. H. Nguyen (1957). This result also supports the suggestions made by Daley (1996) and Tran (2018).

The data also shows that the definiteness of the noun is determined either by the presence of a demonstrative, a possessive, an ordinal number, or a defining clause or by previous mention in the context. The definiteness of the noun in the Classifier - Noun construction is not an exception. It means that the presence of the classifier in this construction or any other constructions does not play any role in determining the definiteness of the noun. This supports Bisang (1999)'s discussions that the occurrence of the classifier in this construction does not determine the definiteness of the noun. The data also reveals that the extra *cái* (inanimate) in the classifier doubling construction performs the function of an emphatic, not a determiner as Simpson and Ngo (2018) argue.

Chapter 6

Summary, conclusions, and implications

The purpose of this chapter is to summarize the dissertation. The goals and methodology are summarized in section 6.1, while the major findings and conclusions are in section 6.2. Section 6.3 considers implications for teaching classifiers and makes recommendations for future research.

6.1 Summary of the goals and methodology of the study

6.1.1 Summary of the goals of the dissertation

Variation is an inherent part of language (Labov 1969), and classifiers are an important category of Vietnamese which has not been investigated extensively on a corpus-based study to date. As the system of language that we speak and write is changing (Tagliamonte 2012), it is hypothesized that variation in classifier use exists across genres. With investigation of inanimate classifiers in folktales, online newspapers, and talk shows, the dissertation attempts to better understand how classifiers are used in spoken and written Vietnamese nowadays, lending to new insights about synchronic variation of Vietnamese classifiers. This study analyses and compares the use of classifiers with respect to frequency and distribution among the three corpora with the focus on the uses of *cái* (inanimate) and double classifiers because *cái* (inanimate) is hypothesized to appear very frequently in conversations compared to written texts, while double classifiers are observed to be used often in spoken Vietnamese (compared to concurrent written Vietnamese). The current research further explores the functions of each classifier in the doubling construction. It also looks at the frequency of classifiers among different age groups in the spoken corpus since this comparison is expected to reveal some differences in language use among them. The discrepancy in the classifier frequency among different age groups may suggest language change in Vietnamese in apparent time.

6.1.2 Summary of the methodology of the study

The data that is used for this dissertation comes from three corpora. The Narrative Corpus consists of 141 Vietnamese folktales, with the word count of about 115,000 words. The Online Newspaper Corpus contains 140 contemporary e-articles with the word count of 135,900 words,

while the Spoken Corpus comprises 22 talk show episodes with the duration of 14 hours and the word count of 151,000 words. The study has been done within the framework of corpus linguistics and the variationist framework. All the inanimate classifier tokens in the three corpora have been investigated and analysed under a number of linguistic variables such as classifier types (single or double), lexical semantic functions of classifiers, presence or absence of numerals, definiteness of nouns, noun referents, previous mention in discourse or not, and kinds of nouns (concrete or abstract). However, social variables, age factor to be specific, have been examined in the spoken corpus only because the age of the speakers in the talk shows can be looked up due to the speakers' being well-known in Vietnam, while the age of the writers in the newspaper and narrative corpora could not be identified. The criteria for identifying an inanimate classifier in Vietnamese regarding position and lexical semantic functions are set out and shown in Table 1 in section 3.2.2. However, we cannot rely on word class to identify the head noun because there are no markers or form of words to indicate parts of speech in Vietnamese (H. T. Nguyen 2013).

6.2 Major findings and conclusions of the dissertation

6.2.1 Frequency of classifier use

Based on the examination of 8626 inanimate classifier tokens found in the three corpora, it can be concluded that there is variation in classifier use regarding frequency and distribution across genres in Vietnamese. The first interesting finding is that the classifier frequency in spoken Vietnamese is higher than in written language. Vietnamese classified nouns require a classifier to be individualized and counted, while non-classified nouns do not need a classifier (Emeneau 1951, D. H. Nguyen 1957, Thompson 1965). Although many nouns can appear with several different classifiers, Vietnamese classifiers categorize the head noun based on the inherent feature or characteristic of the noun's referent such as animacy, shape, size, length, dimension, function, or material. This means, classified nouns have to go with a certain classifier, depending on the property of the thing that the noun refers to, and/or the speaker's focus. The higher classifier frequency in spoken Vietnamese over written language may be due to the genre effects, the content of texts or discourse, and/or individual speakers. Since *cái* (inanimate) has been used as an emphatic frequently with non-classified nouns in the spoken corpus, but not in the written corpora, I argue that the frequent use of *cái* (inanimate) with non-classified nouns as emphatics results in the higher classifier frequency in spoken Vietnamese than in written discourse. With a

phonological stress on *cái* (inanimate) as an effective paralinguistic device for emphasis (Biber and Conrad 2009), it is the speaker's intention to attract the listener's attention and/or focus on certain nouns in their speech. The use of *cái* (inanimate) as an emphatic can be seen as some kind of language change in progress in spoken Vietnamese, but not in written language.

Furthermore, classifier use in the spoken corpus follows the pattern that the older the speakers are, the more classifiers appear in their speech. This means there is a decline in classifier frequency by younger age group despite a significant increase overall in classifier use in the spoken corpus compared to the other two written corpora. This decrease may be due to the content of the discourse and topics of the talk in each of the talk show episodes, and individual effects. It is interesting to find that the use of *cái* (inanimate) with non-classified nouns as emphatics does not affect the classifier use pattern among the three age groups. In contrast, the finding that younger speakers sometimes omit the required classifier before classified nouns with specific referents in their speech leads to the assumption that language simplification may be taking place for gains of speed among younger speakers when the omission of the classifier does not impact the coherence of the discourse (Chandrasekar et. al. 1998). This may be seen as language change in apparent time in progress in Vietnamese, but this needs further investigation. In short, Vietnamese speakers may use more *cái* (inanimate) with non-classified nouns as emphatics in spoken Vietnamese, not in written language, which results in the higher classifier frequency in spoken language than in written language, although language simplification might be taking place in the speech of younger generations.

6.2.2 Distribution of classifiers in the three corpora

Due to the genre effects, the content of texts or discourse, and individual speakers/writers, the overall distribution of classifiers differs within and across the three genres. The findings of the study show some diachronic and synchronic variations in the overall distribution of classifiers across the genres. The finding that more frequent classifiers in the narrative corpus overlap with those in Löbel (2000)'s ten core classifiers compared to the other two concurrent corpora reveals that the distribution of classifiers in the narrative is more relevant with previous studies. However, the online newspaper and spoken corpora have more frequent classifiers in common including *sự* (event) and *cuộc* (life, strike), which are infrequent in the narrative. The frequent use of these two

classifiers in the concurrent corpora suggests that there may be some change in classifier choice in Vietnamese over time, which can be seen as diachronic variation.

In addition, comparing the distribution of classifiers in the two concurrent corpora, newspaper and spoken, the most noticeable difference is that *cái* (inanimate) is very frequent in the spoken corpus, while *sự* (event) is most frequent in the newspaper. It is interesting to find that *sự* (event) appears with a number of verbs and adjectival verbs in the newspaper, while *cái* (inanimate) is used with these verbs in the spoken corpus. This finding reveals that the choice of classifiers for the same noun/nominal is different in different genres. The formality of the newspaper appears to be higher with more frequent appearance of *sự* (event) in nouns/nominals in the non-deictic category and less *cái* (inanimate) as deictic category. However, the formality of the talk shows appears to be lower than the newspaper with more frequent occurrence of *cái* (inanimate), especially those functioning as pronouns or deictic category. This means different classifiers are used for the same noun depending on the different formality level in various genres. I argue that the choice of classifiers is dependent on the formality of the genres. This is a synchronic variation in language use across genres in Vietnamese. In short, inanimate classifiers in Vietnamese are distributed differently across genres. The use of different classifiers diachronically reveals language change over time, and the choice of classifiers is influenced by the formality of the genre as well as the speaker's intention.

Furthermore, *cái* (inanimate) is especially frequent in the spoken corpus compared to the two written corpora. As discussed in 5.3, it has a great capability of combining with a variety of different parts of speech including concrete and abstract nouns (even English nouns), verbs/verb phrases, adjectives/ adjective phrases, wh-words, modal words, idioms, numbers, and noun clauses with or without conjunction words. Specifically, *cái* (inanimate) appears with non-classified nouns as emphatics quite often in spoken Vietnamese. Also, it is constructed with different inanimate classifiers before classified nouns in the double classifier construction. The finding of this study suggests that *cái* (inanimate) as an inanimate classifier is being grammaticalized into a grammatical morpheme that co-occurs with different grammatical types and appears to expand in function in Vietnamese discourse. In short, *cái* (inanimate) is described as a classifier that is being generalized in function and use in Vietnamese grammar. However, this is not the focus of the current study, and would be left for future research.

6.2.3 Functions of classifiers in Vietnamese

In addition to the two primary functions: classification and individualization (including identification), as well as anaphoric reference (Bisang 1999), the results of this study show that a number of Vietnamese classifiers function as nominalizers as discussed by H. T. Nguyen (2004). These classifiers combine with various verbs and adjectival verbs to form nouns with different semantic types in Vietnamese. The two most frequent classifiers in the two concurrent newspaper and spoken corpora are *sự* (event) and *cuộc* (life, strike), functioning as nominalizers. While *sự* (event) usually goes with stative verbs or adjectival verbs to form nouns mainly indicating concepts or perceptions, such as *sự thành công* (CL succeed) and *sự chuyên nghiệp* (CL professional), the classifier *cuộc* (life, strike) normally appears with action verbs to make nouns indicating processes or something that is taking place for a period of time such as *cuộc sống* (CL live), *cuộc phỏng vấn* (CL interview), and *cuộc nói chuyện* (CL talk). The other two less frequent classifiers are *nỗi* (feeling, worry, sad) and *niềm* (sentiment). While *nỗi* (feeling, worry, sad) usually nominalizes adjectival verbs indicating emotional states with negative meanings such as *nỗi lo lắng* (CL worry) and *nỗi buồn* (CL sad), the classifier *niềm* (sentiment) nominalizes adjectival verbs indicating emotional states with positive meanings such as *niềm vui* (CL happy) and *niềm tự hào* (CL proud). It appears that this group of classifiers functioning as nominalizers play a significant role in forming nouns in Vietnamese. This interesting finding reveals that Vietnamese classifiers not only perform the primary functions of numeral classifiers as in other languages, but also appear to add more semantics to the nouns/nominals. This property makes classifiers in Vietnamese a more important class of words in the language.

Another significant finding of the study is that *cái* (inanimate) functions as cataphoric reference. It means that the noun/nominal does not appear with the classifier but occurs in the next upcoming clause in the context. If we want to find out what the classifier indicates in the absence of the noun, we have to refer to the next clause in the discourse. This function has never been mentioned in prior research to date. Furthermore, the results show that *cái* (inanimate) appears quite often with non-classified nouns or is constructed with a classifier before classified nouns in the double classifier construction in spoken Vietnamese, but not in written language. In these cases, *cái* (inanimate) functions as an emphatic. This use of *cái* (inanimate) can be a characteristic to differentiate spoken language from written language because it can be used as a paralinguistic

device for emphasis in their speech, not in writing (Biber and Conrad 2009). This property appears to be a language variation in spoken and written Vietnamese. In the double classifier construction, the specific classifier performs its own function of classification, individualization, and/or nominalization, while *cái* (inanimate) emphasizes the nouns as claimed in previous research (D. H. Nguyen 1957; H. T. Nguyen 2004, 2013; Simpson and Ngo 2018). However, the data of this study does not have evidence to support their argument that *cái* (inanimate) in this case makes the nouns definite. The nouns in the double classifier construction in the study are definite in the presence of a demonstrative, a possessive, or previously mentioned in the context. In the double classifier construction found in the corpus, *cái* (inanimate) can be omitted as it is added to the noun phrase functioning as an emphatic, but the other classifier cannot be removed since it is required to make the classified noun individualized and/or nominalized. This evidence shows that the Vietnamese classifier system is highly complex with differences in their uses in spoken and written language, especially the use of *cái* (inanimate) with non-classified nouns and double classifiers.

In short, the uses and functions of *cái* (inanimate) with non-classified nouns or constructed with another classifier in Vietnamese reveal language variation in spoken and written language. With the function of nominalization, these classifiers can form a wide variety of nouns when combining with different verbs and adjectival verbs, which contributes considerably to the development of Vietnamese. This means that the Vietnamese classifier system appears to be developing in use and function, and of higher importance in the language.

6.2.4 Other findings

As an evidence that corpus-based studies reveal more findings on actual language use, the current research attests the real set of classifiers used in naturalistic data of the three Vietnamese corpora. Apart from the major findings mentioned in the previous sections, the dissertation has identified other findings. Firstly, the study has identified 248 inanimate classifiers excluding 49 double classifiers in the three corpora, in which 110 mensural classifiers altogether are used. This number is much higher than those claimed in prior research because the highest number of classifiers including human and animate (non-human) types claimed by P. P. Nguyen (2002) is 195. It appears that Vietnamese has a very high number of classifier types as it contains a large variety of mensural classifiers in the corpora (Grinevald 2000). This result also appears to be a significant discrepancy between the inventories of classifiers found in previous studies, which

mainly work on constructed or elicited utterances, and the real set of classifiers attested in actual speech (Craig 1986). Furthermore, the findings of this study ascertain that the Vietnamese classifier system does not consist of only three general classifiers as recognized by Cao (1998). The finding shows that Vietnamese has a highly complex classifier system with a rich inventory of classifiers (Emeneau 1951; P. P. Nguyen 2002) including a wide variety of proper classifiers and mensural classifiers.

Secondly, the data of the study shows that the Classifier - Noun would be the typical classifier construction in Vietnamese since over 63% of all the tokens in the corpora follow this pattern. This evidence supports the suggestions made by Daley (1998) and Tran (2018). This would challenge the claim made by prior researchers (Emeneau 1951; D. H. Nguyen 1957; Thompson 1965) that the Numeral - Classifier - Noun is the prototypical classifier pattern of Vietnamese because 27% of all the tokens in the corpora have this construction. In Vietnamese, a classifier, either single or double, can combine with a demonstrative and/or a numeral with the omission of the head noun when the noun is identified in the context, but a classifier never occurs by itself alone. This evidence supports Bisang (1999:148)'s suggestion that the "classifier does not occur alone in its anaphoric function" in most languages.

Thirdly, the study has found that in any classifier construction, the noun is definite in the presence of a demonstrative, a possessive, an ordinal number, and a defining clause, or the previous mention of the noun in the context. In the Classifier - Noun construction without the presence of any other linguistic components in the NPs, the noun might be definite due to the previous mention of the noun in the context. This means, in any case the presence of the classifier does not determine the definiteness of the noun. This supports Bisang (1999)'s discussions that the classifier does not determine the definiteness of the noun in the Classifier - Noun construction. In another construction in Vietnamese, a classifier can combine with wh-words with the omission of the head noun when the noun has been identified in the context. In this case, the entity is individuated due to the appearance of the classifier, but the noun is indefinite despite the absence of a numeral.

Furthermore, another important finding of the study is that although the classifier *chiếc* (individual) can combine with various nouns indicating small objects to big things, either common or proper nouns. The analysis of 268 tokens of *chiếc* (individual) in the three corpora shows that this classifier always appears with concrete nouns in Vietnamese, but never occurs with an abstract

noun. This evidence supports Tran's (2018) argument that *chiếc* (individual) can combine with concrete nouns, but not with abstract nouns, while *cái* (inanimate) can go with both concrete and abstract nouns. Also, *cái* (inanimate) is constructed with *chiếc* (individual) in the double classifier construction, in which *chiếc* (individual) definitely individuates the classified nouns, while *cái* (inanimate) emphasizes the nouns. In this case, either of the classifiers can be omitted. However, when *chiếc* (individual) is removed, *cái* (inanimate) will perform the function of individualization, not the function of emphasis as in the double classifier construction in which it is constructed with *chiếc* (individual).

6.2.5 Concluding summary

In summary, with the descriptive examination of inanimate classifiers in the three Vietnamese corpora, this study reveals that the traditional description of classifier use in Vietnamese does not match to what was observed in naturalistic data of Vietnamese corpora, especially spoken Vietnamese. Vietnamese speakers use *cái* (inanimate) more widely in spoken Vietnamese, and it is constructed with other inanimate classifiers, including an animate classifier *con* (animate). The use of double classifiers, in which *cái* (inanimate) functions as an emphatic, while the other classifier performs its own function of classification, individualization, and/or nominalization, is quite frequent in spoken Vietnamese, but infrequent in written language. Despite the wide use of *cái* (inanimate) in Vietnamese, it is less preferred than *sự* (event) in newspapers. The finding that different classifiers are used with the same nouns in the two concurrent corpora suggests a synchronic variation. This reveals that the choice of classifiers is dependent on the formality of the genre. However, the finding that *sự* (event) and *cuộc* (life, strike) appear frequently in these concurrent corpora, but rarely in the narrative, leads to the assumption that language change may be in progress in Vietnamese over time. This would be considered as a diachronic variation in classifier use in Vietnamese.

Moreover, the finding shows the pattern that classifiers are used more frequently in spoken Vietnamese than in written language. As classified nouns require a certain classifier to be classified and individuated, while non-classified nouns do not need a classifier, the higher classifier frequency in spoken Vietnamese over written language may be due to the genre effects, and the content of the texts or discourse. However, the frequent use of *cái* (inanimate) with non-classified nouns as emphatics in spoken Vietnamese may lead to the higher frequency in spoken language

than in written language. This use is dependent on the speaker's choice with the purpose of attracting the listener's attention to certain points in their speech. With a phonological stress, *cái* (inanimate) is a useful paralinguistic device for emphasis in speech, but not in writing (Biber and Conrad 2009). Despite a higher frequency of classifier use overall in spoken Vietnamese, the study found a decline in classifier use among younger age group. Interestingly, the data reveals that the use of *cái* (inanimate) with non-classified nouns as emphatics does not influence the classifier frequency pattern among the age cohorts. In contrast, the finding that younger speakers sometimes omit the required classifier for classified nouns with specific referents in their speech leads to the hypothesis that language simplification without impacting the coherence of the discourse for gains of speed might affect the frequency of classifier use among younger speakers (Chandrasekar et. al. 1998). This needs further investigation as the data of this study is not enough for testing the hypothesis.

In short, the uses of Vietnamese classifiers might be undergoing some changes in apparent time and across genres. The differences in classifier use patterns across the genres as well as in spoken and written language reveal the complexity of the Vietnamese classifier system. Especially, the findings that *cái* (inanimate) can combine with different parts of speech in various constructions reveal that it is being generalized in function and use in Vietnamese grammar, which merits future investigation.

6.3 Implications and recommendations for future research

This section considers implications of the research findings for teaching and learning Vietnamese for native and second language learners in 6.3.1. Recommendations for future research are made in 6.3.2. Finally, section 6.3.3 concludes the dissertation.

6.3.1 Implications for teaching Vietnamese classifiers

This corpus-based study provides a comprehensive and descriptive picture of how inanimate classifiers are actually used in the three genres. The findings of this study provide teachers and learners of Vietnamese language with new insights to approach classifiers from a different perspective. This helps them realize the importance of the Vietnamese classifier system and highly attend to classifiers in language teaching as this class of words has not been paid much attention to (Diep 2005). Although the current pedagogy is not known, the findings of this study

can be applied in the teaching and learning for native and second language learners. Since most nouns in Vietnamese require a classifier to be individuated, it is necessary for the speaker or writer to choose an appropriate classifier for the noun. Therefore, understanding how to use classifiers in the noun phrase in Vietnamese is important for second language learners. As the Vietnamese classifier system is highly complex, it is not easy for them to master all the classifiers in Vietnamese, but learning the frequent classifiers which can appear with different nouns is helpful for them to use the language properly. For instance, *cái* (inanimate), which is a general inanimate classifier and can combine with different parts of speech, is the first classifier that Vietnamese learners should learn. Performing various functions and being used in different constructions, especially in spoken language, it is a very useful word in Vietnamese that learners of the language want to master. Additionally, *cái* (inanimate) is widely used with non-classified nouns functioning as an emphatic in spoken Vietnamese. It is also constructed with another classifier before classified nouns to perform the function of emphasis in spoken Vietnamese, but rarely used in written language. Furthermore, a number of frequent classifiers functioning as nominalizers that can combine with a wide variety of verbs and adjectival verbs to form nouns with different semantic types including *sự* (event) and *cúộc* (life, strike, match) should be taught for second language learners of Vietnamese. They are helpful classifiers which appear often in the concurrent newspaper and spoken discourse of Vietnamese nowadays. Due to the possibility of combining with many other verbs to create various nouns in Vietnamese, these classifiers are of importance to second language learners.

Since certain classifiers are required for different classified nouns in Vietnamese, they are an essential part of the noun phrase. The differences in the uses of classifiers in different genres make it more difficult to use classifiers properly in Vietnamese. Also, the choice of classifiers is dependent on the formality of the genre. Therefore, using the appropriate classifier in the right context and genre in Vietnamese appears to be complicated. This study provides detailed description of which frequent classifiers can combine with which nouns in Vietnamese. This can be sources for a reference grammar of Vietnamese for non-linguists and Vietnamese language resource materials for second language learners. In addition, learning that *cái* (inanimate) can be used as cataphoric reference and anaphoric reference would be beneficial for second language learners of Vietnamese as well. In short, this corpus study has made substantial contributions to the knowledge of the Vietnamese classifier system, with a comparative picture of classifier use in

written and spoken Vietnamese. The findings of this study on the uses of inanimate classifiers across different genres and among different age groups provide useful resources for language teaching for native and second language learners of Vietnamese.

6.3.2 Recommendations for future research

As this study investigates inanimate classifiers in three Vietnamese corpora to better understand the classifier use patterns and variation across genres, it would be better for future studies to examine classifiers in more genres. The investigation of inanimate classifiers in spoken Vietnamese has brought about many interesting findings, especially about the use of *cái* (inanimate), double classifiers, and classifiers functioning as nominalizers. The findings from the spoken corpus shows a decline in classifier use among younger speakers compared to older speakers. This pattern of classifier use among different age groups suggests language simplification may be taking place in Vietnamese in apparent time. However, this would need further investigation on a larger scale, so my future research would focus on spoken Vietnamese with more speakers of different ages. Additionally, this research looks at the category of inanimate classifiers for an intensive and in-depth investigation. To better describe an overall picture of Vietnamese classifiers, further studies on the other two categories of classifiers, human and animate non-human classifiers, are needed. Especially, the investigation of the human classifier type could introduce interesting results because the choice of a human classifier in Vietnamese may be influenced by social status and age factors. Finally, the analysis of the use of *cái* (inanimate) reveals that *cái* (inanimate) is being grammaticalized in Vietnamese. Thus, the grammaticalization of *cái* (inanimate) would be investigated in the future.

6.3.3 Conclusion

In conclusion, this dissertation is the first known corpus-based discourse analysis of inanimate classifiers in spoken and written Vietnamese as well as among different age groups. As a large-scale study of frequency, distribution and function of attested inanimate classifiers in the three Vietnamese corpora, this study attends closely to the use of *cái* (inanimate) as a free-standing classifier and as a part of double classifiers. With a detailed and comparative description of how inanimate classifiers are used in the three corpora, this study has brought new insights about the uses of Vietnamese classifiers across the three genres as well as in spoken and written Vietnamese.

The findings of this study show evidence that corpus-based studies reveal more findings about actual language use and contribute substantially to the knowledge treasure of the world's languages generally and classifiers specifically. Even though this study identified differences from previous research, the study supports the claims made by previous researchers about the functions of Vietnamese classifiers as a numeral classifier system (Bisang 1999; Löbel 2000; H. T. Nguyen 2004, 2013). It also supports the claims that Vietnamese has a large number of classifiers including proper classifiers and mensural classifiers (Emeneau 1951; D. H. Nguyen 1957; Thompson 1965; P. P. Nguyen 2002). The results of the study show that as a highly complex classifier system, Vietnamese classifiers, which perform different functions, play an important role in the language. The findings about the uses of classifiers in the three corpora of this study reveal that the choice of classifiers is dependent on the formality of the genre, and language change may be in progress in Vietnamese. With substantial contributions to the knowledge of the Vietnamese classifier system, this study underscores the importance of recognizing variation of classifier use in spoken and written Vietnamese and across genres as well as among different age groups of speakers.

References

- Adams, Karen Lee (1986). Numeral classifiers in Austroasiatic. In Colette Craig (ed.), *Noun classes and categorization*. Typological Studies in Language 7. John Benjamins Publishing Company. Amsterdam/Philadelphia. pp. 241-262.
- Adams, Karen Lee (1989). *Systems of numeral classification in the Mon-Khmer, Nicobarese and Asian subfamilies of Austroasiatic*. Research School of Pacific Studies, The Australian National University Canberra, Australia.
- Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford University Press.
- Allan, Keith (1977). *Classifiers*. Language, Vol. 53, No. 2, pp. 285-311.
- Atkinson, Mark; Smith, Kenny and Kirby, Simon (2018). Adult Learning and Language Simplification. Cognitive Science 42, pp. 2818-2854. Wiley Periodicals Inc.
- Bale, Alan and Coon, Jessica (2014). *Classifiers are for Numerals, Not for Nouns: Consequences for the Mass/Count Distinction*. Linguistic Inquiry 45: 695–707.
- Behrens, Leila (2003). Classifiers, metonymies, and genericity: A study of Vietnamese. In Zelinsky-Wibbelt, Cornelia. *Text, Context, Concepts*, pp. 65-125. De Gruyter Mouton.
- Bybee, Joan (2006). Frequency of Use and the Organization of Language. Oxford University Press.
- Biber, Douglas (1986). Spoken and Written Textual Dimensions in English: Resolving in the Contradictory Findings. Language, Vol. 62, No. 2, pp. 384-414. Linguistics Society of America.
- Biber, Douglas and Conrad, Susan (2009). *Register, Genre, and Style*. Cambridge University Press.
- Bisang, Walter (1993). *Classifiers, Quantifiers and Class Nouns in Hmong*. Studies in Language, pp. 1-51.
- Bisang, Walter (1999). Classifiers in East and Southeast Asian languages: counting and beyond. In Gvozdanovic, J. (Ed.) *Numeral Types and Changes Worldwide*, pp. 113-185. Mouton de Gruyter, Berlin.
- Cao, Xuan Hao (1988). The Count/Mass Distinction in Vietnamese and the Concept of ‘Classifier’. Z. Phon. Sprachwiss. Kommunik.forsch. (ZPSK), Berlin 41, pp. 38-47.
- Cao, Xuan Hao (1998). Tiếng Việt: Mấy vấn đề về Ngữ Âm, Ngữ Pháp, Ngữ Nghĩa

- (Vietnamese: Issues in Phonetics, Syntax, and Semantics). Ho Chi Minh City: Nhà Xuất bản Giáo dục (Education Publisher).
- Chandrasekar, Raman; Doran, Christine and Srinivas, Bangalore (1998). Motivations and Methods of Text Simplification. DOI: 10.3115/993268.993361
- Chierchia, Gennaro (1998). *Reference to Kinds Across Language*. Natural Language Semantics Vol. 6, No. 4, pp. 339-405. Springer. <https://www.jstor.org/stable/23748806>
- Craig, Colette. (1986). *Noun classes and categorization*. Typological Studies in Language 7. John Benjamins Publishing Company. Amsterdam/Philadelphia.
- Craig, Colette (1992). Classifiers in a functional perspective. In Fortescue, M. et al. (Ed.), *Layered Structure and reference in Functional Perspective*, pp. 277-301. John Benjamins Publishing Company. Amsterdam.
- Daley, Karen Ann. (1998). *Vietnamese classifiers in narrative texts*. A Publication of the Summer Institute of Linguistics and the University of Texas at Arlington.
- Denny, Peter. (1976). "What are Noun Classifiers Good For?" Papers from the 12th Regional Meeting of the Chicago Linguistic Society. 122-132.
- Diep, Quang Ban. (2005). *Ngữ pháp tiếng Việt* (Vietnamese Grammar). Thua Thien - Hue: Nhà Xuất bản Giáo dục (Education Publisher).
- Dixon, R. M. W. (1986). Noun Classes and Noun Classification in typological perspective. In Craig, Colette (Ed.), *Noun classes and categorization*, pp. 105-112. John Benjamins Publishing Company. Amsterdam/Philadelphia.
- Emeneau, M. B. (1951). *Studies in Vietnamese (Annamese) Grammar*. University of California Press. Berkeley and Los Angeles.
- Erbaugh, Mary S. (1986). Taking stock: The development of Chinese noun classifiers historically and in young children. In Craig, Colette (Ed.), *Noun classes and categorization*, pp. 399-436. John Benjamins Publishing Company. Amsterdam/Philadelphia.
- Greenberg, Joseph H. (1972). *Numeral classifiers and substantial number: Problems in the genesis of a linguistic type*. Working Papers on Language Universals, No. 9, pp. 1-39. Stanford University, California Committee on Linguistics.
- Greenberg, Joseph H. (1974). *Language Typology: A Historical and Analytic Overview*. De Gruyter Mouton. The Netherlands.

- Grinevald, Colette (2000). *A morphosyntactic typology of classifiers*. In: Senft, J. (Ed.), *Nominal Classification*. Cambridge University Press, Cambridge, UK, pp. 50–92.
- Grinevald, Colette (2015). Classifiers, Linguistics of. In Wright, James D (Ed.), *International Encyclopedia of the Social and Behavioral Sciences Reference Work*, 2nd edition, pp. 811-818.
- Heylighen, Francis and Dewaele, Jean-Marc (1999). Formality of Language: definition, measurement and behavioral determinants. Internal Report, Center “Leo Apostel”, Free University of Brussels.
- Honeyfield, John (1977). Simplification. *TESOL Quarterly*, Vol. 11, No. 4, pp.431-440. Teachers of English to Speakers of Other Languages, Inc.
- Hoang, Tat Thang (1996). *Hoạt động của loại từ tiếng Việt trong các phong cách ngôn ngữ* (Operations of Vietnamese classifiers in linguistic styles). PhD dissertation. National University of Social Sciences and Humanities – Hanoi.
- Hopper, P. J., and Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge; New York: Cambridge University Press.
- Labov, William (1969). *Contraction, Deletion, and Inherent Variability of the English Copula*. *Language*, Vol. 45 No. 4; pp. 715-762. Linguistic Society of America.
<https://www.jstor.org/stable/412333>.
- Li, Charles N. and Thompson, Sandra A. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
- Löbel, Elisabeth (2000). Classifiers versus genders and noun classes: A case study in Vietnamese. In Barbara Unterbeck et. al. (Ed.) *Gender in Grammar and Cognition*, pp. 259-319. Berlin: Mouton de Gruyter.
- Luu, Van Lang. (2000). *Một số vấn đề về loại từ trong tiếng Việt* [Some issues about Vietnamese classifiers]. Trung tâm Khoa học Xã hội và Nhân văn: Viện Ngôn ngữ học Quốc Gia [Centre of Natural and Social Sciences: National Institute of Linguistics]. *Loại từ trong các ngôn ngữ ở Việt Nam* [Classifiers of the languages in Vietnam, pp.9-31]. Hà Nội: Nhà Xuất bản Khoa học Xã hội (Publisher of of Social Sciences).
- Ly, Toan Thang (1998). *Representation of space in Vietnamese classifiers*. *Mon-Khmer Studies* 29:71-80.
- McEnery, Tony and Wilson, Andrew (1996). *Corpus Linguistics: An Introduction*, 2nd edition.

- Edinburgh, UK: Edinburgh University Press.
- McEnery, Tony and Hardie, Andrew (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- Meder, Theo (2010). From a Dutch Folktale Database towards an International Folktale Database. *Fabula*, Vol. 51 (1/2), p. 6. De Gruyter Online Journals.
- Ngo, Binh (2012). *Vietnamese Classifiers: (Non-)specificity*. MA thesis, Northeastern Illinois University.
- Nguyen, Dinh Hoa (1957). *Classifiers in Vietnamese*. *Word*, 13:1, 124-152, DOI: 10.1080/00437956.1957.11659631.
- Nguyen, Dinh Hoa (1997). *Vietnamese*. John Benjamins Publishing Company. Amsterdam/Philadelphia.
- Nguyen, Hung Tuong (2004). *The structure of the Vietnamese Noun Phrase*. Unpublished doctorate dissertation, Boston University, USA.
- Nguyen, Hung Tuong (2013). The Vietnamese noun phrases. In Hole, Daniel; Löbel, Elisabeth (Eds.), *Linguistics of Vietnamese – An International Survey*, (pp. 57-87). De Gruyter Mouton.
- Nguyen, Van Ngoc On Nhu (2016). *Truyện cổ nước Nam (Vietnamese folktales)*. Nhà Xuất bản Kim Đồng (Kim Dong Publisher). Part: People, Book 1.
- Nguyen, Phu Phong (2002). *Những vấn đề ngữ pháp tiếng Việt (Vietnamese grammar issues)*. Nhà Xuất bản Đại học Quốc gia Hà Nội (National University Publisher – Hanoi).
- Nguyen, Tai Can (1975). *Từ loại danh từ trong tiếng Việt hiện đại 1960* (Parts of speech: noun in contemporary Vietnamese 1960). Nhà xuất bản Đại học và Trung học chuyên nghiệp Hà Nội (Hanoi University and Professional College Publisher).
- Nguyen, Thi Bich Ngoan (2013). So sánh đối chiếu hiện tượng danh hóa động từ trong tiếng Việt và tiếng Anh (Contrastive analysis of nominalization of verbs in Vietnamese and English). *Journal of Science*, Ho Chi Minh City University for Teachers' Training, No. 46, pp. 13-22.
- Pham, Giang and Kohnert, Kathryn (2008). *A corpus-based analysis of Vietnamese con and cái*. *Mon-Khmer Studies* 38: 161-171.
- Pham, Thi Thuy Hong (2012). *Khảo sát loại từ tiếng Việt và các phương thức chuyển dịch sang tiếng Indonesia* (Survey of Vietnamese classifiers and methods to translate into

- Indonesian). PhD dissertation. National University of Social Sciences and Humanities - Hanoi.
- Poplack, Shana (1993). Variation Theory and Language Contact. In D. Preston (ed.), *American dialect research: An anthology celebrating the 100th anniversary of the American Dialect Society*, 251-263. Amsterdam: Benjamins.
- Seiler, Hansjakob (1986). *Apprehension. Language, object and order III: the universal dimension of apprehension*. Tübingen: Gunter Narr.
- Simpson, A.; Soh, H. L. and Nomoto, H. (2011). Bare classifiers and definiteness, A cross-linguistic investigation. *Studies in Language* 35:1, 168-193. DOI 10.1075/sl.35.1.10sim. John Benjamins Publishing Company.
- Simpson, Andrew (2008). Classifiers and DP Structure in Southeast Asia. In Guglielmo Cinque and Richard Kayne (Ed.). *The Oxford Handbook of Comparative Syntax*, pp. 806–838. Oxford: Oxford University Press.
- Simpson, Andrew and Ngo, Binh (2018). Classifier syntax in Vietnamese. *J East Asian Linguist* 27:211-246. Springer Science + Business Media B.V., part of Springer Nature.
<https://doi.org/10.1007/s10831-018-9181-5>
- Tagliamonte, Sali A. (2012). *Variationist Sociolinguistics*. Wiley-Blackwell.
- Thompson, L. C. (1965). *A Vietnamese Grammar*. Seattle: University of Washington Press.
- Tran, Dai Nghia (1996). Sự tổ hợp loại từ với danh từ trong tiếng Việt hiện đại (Combination of classifiers and noun in contemporary Vietnamese). PhD dissertation summary. Institute of Linguistics. National Centre of Social Sciences and Humanities. Hanoi.
- Tran, Jennie (2011). *The Acquisition of Vietnamese Classifiers*. Unpublished Doctorate Dissertation. University of Hawai'i, USA.
- Tran, Jennie (2013). Vietnamese classifier phrases from the perspective of how children acquire them. In Hole, Daniel; Löbel, Elisabeth (Eds.), *Linguistics of Vietnamese – An International Survey*, (pp. 87-123). De Gruyter Mouton.
- Tran, T. K.; Bui, K. and Pham, D. K. (1960). *Việt Nam văn phạm* [Vietnamese Grammar] Hà Nội: Nhà Xuất bản Tân Việt.
- Tran, T. T. Hai (2018). *A corpus-based study of the functions and structure of classifier constructions in Vietnamese*. Manuscript. General's paper. University of Manitoba.
- Truong, Van Chinh (1970). *Structure de la Langue Vietnamienne*. (Khảo luận về Ngữ pháp Việt

Nam). Paris.

The General Statistics Office of Vietnam (2019).

Microsoft Excel 2016.

Appendix A

Distribution of inanimate CLs in the Vietnamese Narrative Corpus

Inanimate CLs	No. of occurrences	%
cái (inanimate)	404	22.10
cây (tree, long object)	180	9.85
quả (fruit, round object)	67	3.67
chiếc (individual)	59	3.23
bờ (bank, shore, fence)	46	2.52
gốc (root)	45	2.46
thứ (type, kind)	40	2.19
hòn (round)	38	2.08
con (animate)	36	1.97
dòng (flow, river, line)	22	1.20
ngọn (peak-shaped object)	22	1.20
đám (procession, patch, mass)	20	1.09
hạt (seed, small round object)	20	1.09
bát (bowlful)	18	0.98
cành (branch of tree)	18	0.98
gùi (quiver)	18	0.98
chum (big jar)	17	0.93
củ (bulb)	17	0.93
mũi (point, top part)	17	0.93
quan (money)	16	0.88
đồng (money)	15	0.82
bài (unit of song, lesson)	14	0.77
bên (side)	14	0.77
cuộc (life, strike, match)	14	0.77
tòa (building, palace)	14	0.77
đĩa (plate)	13	0.71

đôi (pair)	13	0.71
lá (leaf)	13	0.71
thanh (long, thin object)	13	0.71
miếng (slice, piece)	12	0.66
bãi (stretch, beach)	11	0.60
bộ (set)	11	0.60
bữa (meal, party)	11	0.60
món (dish)	10	0.55
ống (tube)	10	0.55
bụi (bush)	9	0.49
cánh (wing, door)	9	0.49
con (sudden onset, hunger, rain)	9	0.49
dãy (array)	9	0.49
đồng (load)	9	0.49
lóng (part of bamboo tree)	9	0.49
nắm (closed handful)	9	0.49
trận (fight, rain, wind)	9	0.49
đáy (bottom)	8	0.44
đầu (tip, front)	8	0.44
tảng (big piece)	8	0.44
vũng (puddle)	8	0.44
bắp (banana)	7	0.38
bông (flower)	7	0.38
gói (package)	7	0.38
lòng (trust, grateful)	7	0.38
luồng (current)	7	0.38
mẫu (piece)	7	0.38
ngón (finger)	7	0.38
thúng (basket)	7	0.38
vỏ (cover)	7	0.38

vườn (garden)	7	0.38
bầu (gourdful)	7	0.38
ánh (glow)	6	0.33
bó (bunch, bundle)	6	0.33
cục (piece)	6	0.33
đỉnh (top, summit)	6	0.33
khu (area of forest)	6	0.33
lưỡi (sharp long object)	6	0.33
mảnh (piece)	6	0.33
ngôi (unit of house)	6	0.33
sợi (thread)	6	0.33
sự (event)	6	0.33
túp (tent)	6	0.33
bồ (basket)	5	0.27
chén (cupful)	5	0.27
đoạn (section, part)	5	0.27
hóc (corner)	5	0.27
mái (roof, unit of house)	5	0.27
nhát (action of chopping)	5	0.27
phía (direction)	5	0.27
rễ (root)	5	0.27
trái (fruit, round object)	5	0.27
vết (mark)	5	0.27
bàn (hand)	4	0.22
bộ (bunch)	4	0.22
bung (basket)	4	0.22
căn (unit of house)	4	0.22
đầu (basketful of rice)	4	0.22
dây (string)	4	0.22
gian (section of house)	4	0.22

khúc (section)	4	0.22
loại (kind)	4	0.22
niêu (potful)	4	0.22
nồi (potful)	4	0.22
tính (quality)	4	0.22
túi (bag)	4	0.22
bức (CL picture, wall)	3	0.16
cái đám (patch of grass, procession)	3	0.16
chảo (pan)	3	0.16
ché (jar)	3	0.16
đỉnh (potful)	3	0.16
đường (path)	3	0.16
góc (corner)	3	0.16
hũ (jarful)	3	0.16
khoảng (area)	3	0.16
mặt (item)	3	0.16
mẻ (turn)	3	0.16
mối (hatress)	3	0.16
nùi (hank)	3	0.16
nuơng (field)	3	0.16
ổ (net)	3	0.16
phiến (flat stone)	3	0.16
rừng (forest)	3	0.16
thân (tree-trunk)	3	0.16
thằng (human, male, young)	3	0.16
thôi (bar)	3	0.16
vị (taste, kind of medicine)	3	0.16
vùng (basketful)	3	0.16
xâu (string)	3	0.16
bao (bag)	2	0.11

bè (raft)	2	0.11
buồng (bunch)	2	0.11
cái vị (medicine)	2	0.11
cán (handle)	2	0.11
cặp (pair)	2	0.11
chùm (bunch)	2	0.11
đàng (side)	2	0.11
điều (thing)	2	0.11
giấc (CL dream, sleep)	2	0.11
giỏ (basket)	2	0.11
hàng (row of fence)	2	0.11
hồ (lake)	2	0.11
làn (wave)	2	0.11
lỗ (hole)	2	0.11
lớp (layer)	2	0.11
manh (piece)	2	0.11
môn (subject)	2	0.11
quyển (volume)	2	0.11
số (amount)	2	0.11
tấm (thin)	2	0.11
tầng (layer)	2	0.11
thửa (area of field)	2	0.11
việc (activity)	2	0.11
vụ (case)	2	0.11
ang (big jar)	1	0.05
bản (copy, version)	1	0.05
bị (basketful)	1	0.05
bụng (bellyful)	1	0.05
búp (bobbin, CL thread)	1	0.05
bước (step)	1	0.05

cái con (inanimate, animate)	1	0.05
chai (bottle)	1	0.05
chòm (bunch of leaves)	1	0.05
chóm (CL mountain)	1	0.05
cỗ (set)	1	0.05
coi (unit of betel)	1	0.05
dải (range of clouds)	1	0.05
đòn (action of decision)	1	0.05
dúm (handful)	1	0.05
đuôi (tail of loin-cloth)	1	0.05
gánh (loadful)	1	0.05
gáo (ladleful)	1	0.05
giá (basket)	1	0.05
giáp (big bowlful)	1	0.05
giống (type, kind)	1	0.05
hòm (boxful)	1	0.05
hột (seed)	1	0.05
khe (chink)	1	0.05
khóm (cluster)	1	0.05
lộc (bud)	1	0.05
mâm (table of food)	1	0.05
mầm (bamboo shoot)	1	0.05
mảnh (piece)	1	0.05
mớ (load, bunch)	1	0.05
nấm (CL graveyard)	1	0.05
ngòi (CL pen)	1	0.05
nguồn (source)	1	0.05
nỗi (CL sorrow, pain)	1	0.05
nụ (bud)	1	0.05
nuộc (tight)	1	0.05

quãng (part)	1	0.05
que (stick)	1	0.05
rãnh (small stream)	1	0.05
rẫy (mountain field)	1	0.05
ruộng (field)	1	0.05
sét (set)	1	0.05
sọt (crateful)	1	0.05
suối (stream)	1	0.05
thuyền (boatful)	1	0.05
tình (relationship)	1	0.05
tờ (sheet)	1	0.05
trã (trayful)	1	0.05
túm (handful)	1	0.05
vác (bunch)	1	0.05
viên (small, round object)	1	0.05
xanh (panful)	1	0.05
Total	1828	100.00

Appendix B

Distribution of inanimate CLs in the Online Newspaper Corpus

Inanimate CLs	No. of occurrences	%
sự (event)	277	11.21
cuộc (life, strike, match)	187	7.56
cái (inanimate)	180	7.28
chiếc (individual)	144	5.83
việc (activity)	116	4.69
bộ (set)	68	2.75
vụ (catastrophe)	61	2.47
con (animate)	58	2.35
dòng (river, line)	57	2.31
căn (unit of house)	55	2.22
loại (kind, sort)	48	1.94
chuyến (trip)	46	1.86
cây (tree)	45	1.82
bức (picture, wall)	43	1.74
ngôi (unit of house)	41	1.66
số (amount)	39	1.58
nỗi (worry, sad, scare)	35	1.42
bài song, lesson, writing)	30	1.21
nền (institution)	30	1.21
phần (section, part)	30	1.21
niềm (sentiment)	29	1.17
tình (relationship)	29	1.17
trận (match, fight)	28	1.13
bản (script, report)	26	1.05
mối (care, relationship)	24	0.97
tính (quality)	21	0.85

bãi (stretch, beach)	20	0.81
món (dish)	21	0.85
khoảng (unit of time, area)	19	0.77
cơn (anger, wind)	18	0.73
đám (mass, patch, procession)	18	0.73
trái (fruit, round object)	18	0.73
lá (leaf, thin)	17	0.69
ngọn (peak)	17	0.69
khoản (amount)	17	0.69
mảnh (piece)	16	0.65
quả (fruit, round object)	16	0.65
lòng (trust, quality)	16	0.65
ánh (light, look)	15	0.61
bàn (table, hand)	14	0.57
cánh (wing, field)	14	0.57
bên (side)	13	0.53
điếu (cigarette)	13	0.53
nụ (bud)	12	0.49
lớp (layer)	12	0.49
tấm (thin object)	12	0.49
đôi (pair)	11	0.44
tờ (sheet)	13	0.53
vết (mark)	11	0.44
bữa (meal, party)	10	0.40
cành (branch)	10	0.40
chậu (pot)	10	0.40
cuốn (colume)	10	0.40
mái (roof, unit of house)	10	0.40
tòa (building)	10	0.40
bầu (gourdful)	9	0.36

đồng (money)	9	0.36
loài (kind)	9	0.36
thứ (kind, sort)	9	0.36
cú (blow)	8	0.32
đoạn (section)	8	0.32
góc (corner)	8	0.32
giấc (sleep, dream)	7	0.28
gốc (root)	7	0.28
làn (wave)	7	0.28
chai (bottleful)	6	0.24
chặng (section)	6	0.24
điều (action, article)	6	0.24
gian (room)	6	0.24
hòn (round)	6	0.24
môn (subject)	6	0.24
thảm (carpet)	6	0.24
bờ (bank)	5	0.20
dãy (array)	5	0.20
giọt (drop)	5	0.20
ngón (finger)	5	0.20
quãng (section)	5	0.20
sườn (side of hill)	5	0.20
vẻ (beauty)	5	0.20
bông (flower)	5	0.20
bát (bowlful)	4	0.16
đầu (tip)	4	0.16
hàng (row)	4	0.16
hộp (box)	4	0.16
kiểu (type)	4	0.16
sợi (thread)	4	0.16

thước (measuremet)	4	0.16
cốc (cupful)	3	0.12
cung (road)	3	0.12
cuộn (roll)	3	0.12
đòn (blow)	3	0.12
đụn (dune)	3	0.12
hạt (small round, seed)	3	0.12
liều (dose)	3	0.12
lô (load)	3	0.12
ly (cupful)	3	0.12
miếng (piece)	3	0.12
nạn (corruption)	3	0.12
rừng (forest)	3	0.12
túi (bag)	3	0.12
thùng (box)	3	0.12
viên (round, pill)	3	0.12
vỏ (cover)	3	0.12
vườn (garden)	3	0.12
vựa (granary)	3	0.12
cái phần (part)	2	0.08
chĩnh (jar)	2	0.08
chùm (bunch)	2	0.08
củ (bulb)	2	0.08
đỉnh (top)	2	0.08
đống (heap)	2	0.08
giống (kind)	2	0.08
gói (packet)	2	0.08
lọ (bottle)	2	0.08
lỗ (hole)	2	0.08
mớ (load, bunch)	2	0.08

mũi (point, top part)	2	0.08
nén (bar, stick)	2	0.08
nồi (pan)	2	0.08
quyển (volume)	2	0.08
ruộng (field)	2	0.08
tập (set, volume, episode)	2	0.08
thanh (bar, long object)	2	0.08
tia (glow)	2	0.08
ao (pond)	1	0.04
bao (bag)	1	0.04
bình (pot)	1	0.04
cái bộ (set)	1	0.04
cặp (pair)	1	0.04
chén (bowlful, cupful)	1	0.04
chóp (top, peak)	1	0.04
cụm (cluster, bunch)	1	0.04
dải (range, band)	1	0.04
dàn (set)	1	0.04
đóa (flower)	1	0.04
đồi (hill)	1	0.04
gánh (loads)	1	0.04
giàn (framework)	1	0.04
khẩu (gun)	1	0.04
khúc (section)	1	0.04
lát (slice)	1	0.04
lưỡi (sharp part)	1	0.04
màn (scene)	1	0.04
mảng (patch)	1	0.04
manh (piece)	1	0.04
ngậm (gulp)	1	0.04

nút (tight)	1	0.04
que (stick)	1	0.04
rạn (reef)	1	0.04
tán (cluster)	1	0.04
tảng (big piece)	1	0.04
thìa (spoonful)	1	0.04
xe tải (truck)	1	0.04
Total	2472	100.00

Appendix C

Distribution of inanimate CLs in the Vietnamese Spoken Corpus

Inanimate Classifiers	No. of occurrences	%
cái (inanimate)	2657	61.42
bài (unit of song, lesson)	204	4.72
cuộc (strike, life)	201	4.65
cái sự (inanimate, event)	144	3.33
sự (event)	129	2.98
chiếc (individual)	65	1.50
trình (relationship)	61	1.41
con (animate)	58	1.34
đám (procession, patch, mass)	47	1.09
cái bài (song, lesson, text)	41	0.95
phần (section, part)	29	0.67
niềm (sentiment)	27	0.62
cái cuộc (inanimate, strike, life)	26	0.60
bộ (set)	22	0.51
chuyến (trip)	22	0.51
món (dish)	21	0.49
cây (tree, long object)	19	0.44
vở (play)	19	0.44
bức (picture)	15	0.35
cái con (inanimate, animate)	16	0.37
cái phần (inanimate, part)	15	0.35
số (amount)	14	0.32
cái đám (inanimate, procession)	14	0.32
khoảng (period)	13	0.30
bên (side)	12	0.28
cái niềm (inanimate, sentiment)	12	0.28

nỗi (feeling, worry)	12	0.28
cái bộ (inanimate, set)	13	0.30
cái việc (inani., nom.)	11	0.25
điều (affair)	10	0.23
cái khoảng (inanimate, period)	10	0.23
mối (relation)	10	0.23
đoạn (section)	9	0.21
ngôi (unit of house)	9	0.21
cái nỗi (inanimate, nom.)	8	0.18
chặng (part)	9	0.21
cơn (anger, wind)	8	0.18
tấm (thin object)	8	0.18
cái chuyến (inanimate, trip)	8	0.18
bản (script)	7	0.16
bước (stage)	7	0.16
cái món (inanimate, dish)	7	0.16
dòng (flow, line)	7	0.16
lòng (trust, quality)	7	0.16
cái ánh (inani., glow)	6	0.14
cái mối (inanimate, relationship)	6	0.14
lá (leaf)	6	0.14
quả (fruit, round object)	6	0.14
trái (fruit, round object)	6	0.14
tính (quality)	6	0.14
việc (activity)	6	0.14
tràng (applause)	5	0.12
thứ (kind, sort)	5	0.12
cú (blow)	5	0.12
cái buổi (inanimate, session)	5	0.12
bó (bunch)	4	0.09

bữa (meal)	4	0.09
cái bước (inanimate, stage)	4	0.09
cái chiếc (inanimate, individual)	4	0.09
cái dòng (inanimate, flow, line)	4	0.09
cái tính (inanimate, quality)	4	0.09
đôi (pair)	4	0.09
cuốn (volume)	4	0.09
giấc (sleep)	4	0.09
nụ (smile)	4	0.09
cái cơn (inanimate, anger, wind)	4	0.09
ánh (glow)	3	0.07
bàn (hand)	3	0.07
cái cú (inanimate, blow)	3	0.07
cái cuốn (inanimate, volume)	3	0.07
cái ngôi (inanimate, unit of house)	3	0.07
cái quả (inanimate, round object)	3	0.07
cái sợi (inanimate, thread)	3	0.07
cột (pole)	3	0.07
mái (house)	3	0.07
màn (scene)	3	0.07
quyển (volume)	3	0.07
tờ (sheet)	3	0.07
vết (mark)	3	0.07
viên (small round)	3	0.07
bầu (atmosphere)	2	0.05
bông (flower)	2	0.05
cái căn (inanimate, unit of house)	2	0.05
cái giấc (inanimate, sleep)	2	0.05
cái quãng (inanimate, section)	2	0.05
cái quyển (inanimate, volume)	2	0.05

cánh (door)	2	0.05
củ (root)	2	0.05
đồng (field)	2	0.05
giọt (drop)	2	0.05
loại (kind)	2	0.05
luồng (flow)	2	0.05
miếng (piece)	2	0.05
ngọn (top part, mountain)	2	0.05
quãng (section)	2	0.05
nền (institution)	2	0.05
vẻ (beauty)	2	0.05
cái bản (inanimate, script)	1	0.02
bát (bowlful)	1	0.02
cái bức (inanimate, picture)	1	0.02
cái cánh (inanimate, door)	1	0.02
cái cây (inanimate, tree, long object)	1	0.02
cái chặng (inanimate, section)	1	0.02
cái điều (inanimate, affair)	1	0.02
cái đoạn (inanimate, section)	1	0.02
cái đồng (inanimate, money)	1	0.02
cái đốt (inanimate, knot)	1	0.02
cái khúc ((inanimate, part)	1	0.02
cái làn (inanimate, wave)	1	0.02
cái mảnh (inanimate, piece)	1	0.02
cái ngọn (inanimate, top part)	1	0.02
cái nụ (inanimate, smile)	1	0.02
cái set (inanimate, set)	1	0.02
cái tấm (inanimate, picture, degree)	1	0.02
cái tờ (inanimate, sheet)	1	0.02
cái túi (inanimate, bag)	1	0.02

cặp (pair)	1	0.02
cọng (grass)	1	0.02
cục (small piece)	1	0.02
cung (section)	1	0.02
điếu (cigarette)	1	0.02
điệu (dance)	1	0.02
đồng (load)	1	0.02
đốt (section, knot)	1	0.02
hàng (row)	1	0.02
liều (dose)	1	0.02
lọ (bottle)	1	0.02
manh (thin piece)	1	0.02
nấm (grave)	1	0.02
ngày (day)	1	0.02
sàng (wise)	1	0.02
tập (episode)	1	0.02
túi (bag)	1	0.02
vũng (puddle)	1	0.02
cái cái (cái cái cái cái/viên viên)	44	1.02
Total	4326	100.00

Appendix D

List of actual classifiers in the three corpora

Narrative	Online Newspaper	Spoken
ánh (glow)	ánh (light, look)	ánh (glow)
bài (song, lesson, text)	bài (song, lesson, text)	bài (song, lesson, text)
bản (script, version)	bản (script, version, report)	bản (script, version)
bàn (hand)	bàn (table, hand)	bàn (hand)
bát (bowlful)	bát (bowlful)	bát (bowlful)
bầu (gourdful)	bầu (gourdful)	bầu (atmosphere)
bên (side)	bên (side)	bên (side)
bộ (set)	bộ (set)	bộ (set)
bông (flower)	bông (flower)	bông (flower)
bữa (meal, party)	bữa (meal, party)	bữa (meal, party)
bức (picture, wall)	bức (picture, wall)	bức (picture, wall)
cái (inanimate)	cái (inanimate)	cái (inanimate)
cánh (wing, door)	cánh (wing, field)	cánh (door, field)
cặp (pair)	cặp (pair)	cặp (pair)
cây (tree, long object)	cây (tree, long object)	cây (tree, long object)
chiếc (individual)	chiếc (individual)	chiếc (individual)
con (hunger, anger, wind,)	cơn (anger, wind, rain)	cơn (anger, wind, rain)
con (animate)	con (animate)	con (animate)
củ (bulb, root)	củ (bulb, root)	củ (bulb, root)
cuộc (life, strike, match)	cuộc (life, strike, match)	cuộc (life, strike, match)
đám (procession, patch)	đám (procession, patch)	đám (procession, patch)
điều (article, action)	điều (action, article, action)	điều (article, affair)
đoạn (section, part)	đoạn (section, part)	đoạn (section, part)
đôi (pair)	đôi (pair)	đôi (pair)
dòng (flow, river, line)	dòng (flow, river, line)	dòng (flow, river, line)
đồng (load)	đồng (load)	đồng (load)

đồng (money, field)	đồng (money, field)	đồng (field)
giấc (CL dream, sleep)	giấc (dream, sleep)	giấc (dream, sleep)
hàng (row)	hàng (row)	hàng (row)
khoảng (period, area)	khoảng (period, area)	khoảng (period, area)
lá (leaf, thin object)	lá (leaf, thin object)	lá (leaf, thin object)
loại (kind, sort)	loại (kind, sort)	loại (kind, sort)
lòng (trust, grateful, quality)	lòng (trust, quality)	lòng (trust, quality)
mái (roof, unit of house)	mái (roof, unit of house)	mái (house)
manh (piece)	manh (piece)	manh (thin piece)
miếng (slice, piece)	miếng (piece)	miếng (piece)
mối (relationship)	mối (care, relationship)	mối (relation)
món (dish)	món (dish)	món (dish)
ngôi (unit of house)	ngôi (unit of house)	ngôi (unit of house)
ngọn (peak-shaped object)	ngọn (peak)	ngọn (top part, mountain)
nỗi (worry, sorrow, pain)	nỗi (worry, sad, scared)	nỗi (worry, sad, scared)
nụ (smile, bud)	nụ (smile, bud)	nụ (smile, bud)
quả (fruit, round object)	quả (fruit, round object)	quả (fruit, round object)
quãng (section)	quãng (section)	quãng (section)
quyển (volume)	quyển (volume)	quyển (volume)
số (amount)	số (amount)	số (amount)
sự (event)	sự (event)	sự (event)
tấm (thin object)	tấm (thin object)	tấm (thin object)
thứ (kind, sort)	thứ (kind, sort)	thứ (kind, sort)
tình (love, emotion)	tình (love, emotion)	tình (love, emotion)
tính (quality)	tính (quality)	tính (quality)
tờ (sheet)	tờ (sheet)	tờ (sheet)
trái (fruit, round object)	trái (fruit, round object)	trái (fruit, round object)
túi (bag)	túi (bag)	túi (bag)
vết (mark)	vết (mark)	vết (mark)
việc (activity)	việc (activity)	việc (activity)

viên (round, pill)	viên (round, pill)	viên (round, pill)
vỏ (cover)	vỏ (cover)	vở (play)
vụ (case, catastrophe)	vụ (catastrophe)	vũng (puddle)
đòn (blow)	đòn (blow)	niềm (sentiment)
mảnh (patch)	mảnh (patch)	ngày (day)
mảnh (piece)	mảnh (piece)	sàng (wise)
ngón (finger)	ngón (finger)	tràng (applause)
sợi (thread)	sợi (thread)	bó (bunch)
tảng (big piece)	tảng (big piece)	phần (section, part)
tòa (building)	tòa (building)	tập (episode)
bao (bag)	bao (bag)	chặng (part)
bãi (stretch, beach)	bãi (stretch, beach)	chuyến (trip)
bờ (bank, shore, fence)	bờ (bank)	cọng (grass)
căn (unit of house)	căn (unit of house)	bước (stage)
cái con (inani., ani.)	cái bộ (set)	cái con (inani., ani.)
cái đám (inani., procession)	cái phần (inani., part)	cái đám (inani., procession)
ang (big jar)	ao (pond)	cái bộ (inani., set)
bắp (banana)	bình (pot)	cái phần (inani., part)
bè (raft)	cành (branch)	cái ánh (inani., glow)
bị (basket)	chai (bottleful)	cái bài (song, lesson, text)
bồ (basket)	chặng (section)	cái bản (inani., script)
bó (bunch, bundle)	chậu (pot)	cái bức (inani., picture)
bộ (bunch)	chén (bowlful, cupful)	cái bước (inani., stage)
bụi (bush)	chĩnh (jar)	cái buổi (inani., session)
bung (basket)	chóp (top, peak)	cái căn (inani., house)
bụng (bellyful)	chùm (bunch)	cái cánh (inani., door)
bước (step)	chuyến (trip)	cái cây (inani., tree, long)
buồng (bunch)	cốc (cupful)	cái chặng (inani., section)
búp (bobbin, CL thread)	cú (blow)	cái chiếc (inani., individual)
cái vị (medicine)	cụm (cluster, bunch)	cái chuyến (inani., trip)

cán (handle)	cung (road)	cái cơn (inani., anger, wind)
cành (branch of tree)	cuốn (colume)	cái cú (inani., blow)
chai (bottle)	cuộn (roll)	cái cuộc (inani., strike, life)
chảo (pan)	dải (range, band)	cái cuốn (inani., volume)
ché (jar)	dàn (set)	cái điều (inani., affair)
chén (cupful)	đầu (tip)	cái đoạn (inani., section)
chòm (bunch of leaves)	dãy (array)	cái dòng (inani., flow, line)
chỏm (CL mountain)	điếu (cigarette)	cái đồng (inani., money)
chum (big jar)	đỉnh (top)	cái đốt (inani., knot)
chùm (bunch)	đóa (flower)	cái giấc (inani., sleep)
cỗ (set)	đồi (hill)	cái khoảng (inani., period)
coi (unit of betel)	đụn (dune)	cái khúc ((inani., part)
cục (piece)	gánh (loads)	cái làn (inani., wave)
dải (range of clouds)	giàn (framework)	cái mảnh (inani., piece)
đàng (side)	gian (room)	cái mối (inani., relation)
đầu (basketful)	giống (kind)	cái món (inani., dish)
đầu (tip, front)	giọt (drop)	cái ngôi (inani., house)
dãy (array)	góc (corner)	cái ngọn (inani., top part)
đáy (bottom)	gốc (root)	cái niềm (inani., sentiment)
dây (string)	gói (packet)	cái nỗi (inani., worry)
đĩa (plate)	hạt (small round, seed)	cái nụ (inani., smile)
đỉnh (pot)	hòn (round)	cái quả (inani., round object)
đỉnh (top, summit)	hộp (box)	cái quãng (inani., section)
dúm (handful)	khẩu (gun)	cái quyển (inani., volume)
đuôi (tail of loincloth)	khoản (amount)	cái set (inani., set)
đường (path)	khúc (section)	cái sợi (inani., thread)
gánh (loadful)	kiểu (type)	cái sự (inani., nom.)
gáo (ladleful)	làn (wave)	cái tấm (inani., picture, degree)
giá (basket)	lát (slice)	cái tính (inani., quality)
gian (section of house)	liều (dose)	cái tờ (inani., sheet)

giáp (big bowlful)	lọ (bottle)	cái túi (inani., bag)
giỏ (basket)	lỗ (hole)	cái việc (inani., activity)
giống (type, kind)	lô (load)	cái cái (cái cá cái/viên viên)
góc (corner)	loài (kind)	cú (blow)
gốc (root)	lớp (layer)	cục (small piece)
gói (package)	lưỡi (sharp part)	cung (section)
gùi (quiver)	ly (cupful)	cuốn (volume)
hạt (seed, small round)	màn (scene)	điếu (cigarette)
hồ (lake)	mớ (load, bunch)	điệu (dance)
hốc (corner)	môn (subject)	đốt (section, knot)
hòm (boxful)	mũi (point, top part)	giọt (drop)
hòn (round)	nạn (corruption)	liều (dose)
hột (seed)	nén (bar, stick)	lọ (bottle)
hũ (jarful)	nền (institution)	luồng (flow)
khe (chink)	ngậm (gulp)	màn (scene)
khóm (cluster)	niềm (sentiment)	nấm (grave)
khu (area of forest)	nồi (pan)	nền (institution)
khúc (section)	nút (tight)	vẻ (beauty)
làn (wave)	que (stick)	
lỗ (hole)	rạn (reef)	
lộc (bud)	rừng (forest)	
lóng (part of bamboo tree)	ruộng (field)	
lớp (layer)	sườn (side of hill)	
lưỡi (sharp long object)	tán (cluster)	
luồng (current)	thảm (carpet)	
mầm (bamboo shoot)	thanh (bar, long object)	
mâm (table of food)	thìa (spoonful)	
mặt (item)	thùng (box)	
mẫu (piece)	thước (measuremet)	
mẻ (turn)	tia (glow)	

mớ (load, bunch)	trận (match, fight)	
môn (subject)	vựa (granary)	
mũi (point, top part)	vườn (garden)	
nấm (CL graveyard)	xe tải (truck)	
nắm (closed handful)	vẻ (beauty)	
ngòi (CL pen)	phần (section, part)	
nguồn (source)	tập (set, volume, episode)	
nhát (slice)		
niêu (pot)		
nồi (pot)		
núi (hank)		
nuộc (tight)		
nuơng (field)		
ổ (net)		
ống (tube)		
phía (direction)		
phiến (flat stone)		
quan (money)		
que (stick)		
rãnh (small stream)		
rẫy (mountain field)		
rễ (root)		
rừng (forest)		
ruộng (field)		
sét (set)		
sọt (crateful)		
suối (stream)		
tầng (layer)		
thân (tree-trunk)		
thằng (human, low s. s.)		

thanh (long, thin object)		
thỏi (bar)		
thửa (area of field)		
thúng (basket)		
thuyền (boatful)		
trã (trayful)		
trận (fight, rain, wind)		
túm (handful)		
túp (tent)		
vác (bunch)		
vị (taste, kind of medicine)		
vùng (basketful)		
vũng (puddle)		
vườn (garden)		
xanh (pan)		
xâu (string)		

Appendix E

Classifier constructions in each of the corpora

Narrative Corpus		Online Newspaper Corpus		Spoken Corpus	
Constructions	No. of tokens	Constructions	No. of tokens	Constructions	No. of tokens
Num + CL + N + (Attri.)	526	Num + CL + N + (Attri.)	591	Num + CL + N + (Attri.)	1226
CL + N + (Attri.)	1194	CL + N + (Attri.)	1825	CL + N+ (Attri.)	2480
Num + CL	33	Num + CL	10	Num + CL	39
(Num) + CL + Dem	28	(Num) + CL + Dem	13	(Num) + CL + Dem	209
CL + Wh-word (gì/nào)	40	CL + Wh-word (gì/gì đố/nào)	14	CL + Wh-word (gì/gì đố/nào)	241
CL + Name of ref. N	1	CL + Name of ref. N	7	CL + Name of ref. N	16
CL + Ordinal No.	1	Num + CL + called N	1	CL + Ordinal No.	8
How many + CL + (Attri)	1	CL + Wh-clause	2	How many + CL	1
Num + CL + Wh-word (gì/nào)	4	CL + called N	3	Num + CL + Clause	7
Total	1828	CL + (Clause/Attri.)	6	CL + CL + (Clause/Attri.)	4
		Total	2472	CL + CL + Poss/Dem	2
				CL + Wh-clause	15
				CL + (Clause/Attri.)	8
				(Num) + CL + called N	25
				Num + CL + CL	4
				(Num) + CL + be (là) + N/Wh-clause	3

(Num) + CL + mà (which) + Clause	35
CL + idiom	1
(Num) + CL + Poss	2
Total	4326

Appendix F

List of stories used for data in the Vietnamese Narrative Corpus

Book 1: Nguyễn, Văn Ngọc Ôn Như (2016). *Truyện cổ nước Nam (Vietnamese folktales)*. Nhà Xuất bản Kim Đồng (Kim Dong Publisher). Part: People, Volume 1.

1. Chum vàng bắt được
2. Kéo cày giả nợ
3. Cái cân thủy ngân
4. Cây tre trăm mắt
5. Cá rô rạch ngược
6. Cà cuống với người tị mủi
7. Giả chết bắt quạ
8. Sinh con rồi mới sinh cha
9. Ăn mây đánh đổ cầu ao
10. Cái gì to hơn
11. Mười voi
12. Nem công, chả phượng, râu rồng
13. Trạng Ếch
14. Có ai làm chứng
15. Câu đố nên vợ, nên chồng
16. Có vú, không đầu
17. Quýt làm, cam chịu
18. Chú lính ăn khoai
19. Bắt tép nuôi cò
20. Chừa đánh, đánh được
21. Bà chủ và người đi cày
22. Văn Mai và Thị Mật
23. Một hạt giời cho
24. Thịt bò, lộc sấn
25. Chưa đỗ ông nghè
26. Anh câm bật nói

27. Tay què, mặc tay
28. Chú Chích, cô Chòe
29. Có nợ thì có kia
30. Phụng hoàng đậu cây khế
31. Chuột ong đi trước
32. Vua Thế Tổ và ông lão nuôi ong
33. Vị thuốc quý hóa
34. Âm đước
35. Làm lành
36. Mài dao dạy vợ
37. Giết chó khuyên chồng
38. Kêu một việc được ba việc
39. Ông Tú và người buôn mèo
40. Bát canh hẹ
41. Bát canh hương án
42. Cây gì cửa chẳng được
43. Đậu đen chườm đầu
44. Nồi kê ông Thổ
45. Vùng khoai lang
46. Làm rể Chương Đài
47. Đẽo cày giữa đường (I)
48. Đẽo cày giữa đường (II)
49. Giời tóc, gió rung
50. Nịnh đời
51. Con khá hơn thầy
52. Người học trò muốn đậu
53. Học văn hay học võ
54. Tam đại con gà
55. Thầy đồ ăn bánh rán
56. Cờ gian bạc lận
57. Lộc giời hơn lộc nước

58. Nghĩa cũ, tình nay
59. Không giết gián
60. Trọng nghề
61. Người học trò và con chó đá
62. Chó đá đổ máu
63. Dốt học cũng thông
64. Sáng mắt ra
65. Tài với học
66. Không ăn bí
67. Lươn ngắn, trạch dài
68. Con đẻ, con nuôi
69. Mất giỗ, đổ cau
70. Đi lừa tiền cơm
71. Vạc, cò
72. Bát Vạn là quả ớt
73. Mèo lại hoàn mèo
74. Lá húng! Lá húng!
75. Hai vợ chồng anh thầy bói
76. Thằng bọm có con ngựa
77. Đổi lòng lành
78. Hai anh em và con chó đá
79. Thi vẽ nhanh
80. Van như vạc
81. Tri âm với khướu
82. Hai thằng trộm và con ngựa
83. Quân tử ruồi
84. Người ăn mía và người chủ vườn
85. Hai thứ mọt khác nhau
86. Cô lô gốc mít
87. Thèm
88. Sợ sét bà

89. Anh chăn dê và anh xách ngỗng
90. Chiêm bao thịt chó
91. Nụ cà, hoa mướp
92. Không hoa, không chồng
93. Người học trò với con rùa
94. Cây đa biết nói
95. Ba con trâu đực thành chín con
96. Củ khoai và cái cầu
97. Kéo cây lúa lên
98. Thầy dạy học trò
99. Thanh yên so với phật thủ
100. Namô chuỳnh
101. Anh thợ rèn bừa
102. Cuốc, cày, bừa tranh công
103. Cơm với cà
104. Thịt ngóe, canh cà
105. Mẹ hiền, con thảo
106. Cây táo và nhà láng giềng
107. Buôn vịt giò
108. Bữa rượu cháy nhà
109. Chiêm bao thấy lợn kêu
110. Sợ ma bao giờ
111. Làm giường cho vợ đẻ
112. Tù lì tám tiền.

Book 2: Lữ, Huy Nguyên and Đặng, Văn Lung (Ed.) (2013). *100 truyện cổ tích Việt Nam (100 Vietnamese folktales)*. Nhà Xuất bản văn học Đông Á (East Asia Literature Publisher).

113. Bánh chưng, bánh giầy
114. Hổ và thỏ
115. Anh Bọt Thây
116. K'Chơi và Ma Lêng

117. Vua Lợn
118. Thỏ và Ốc
119. Khi và Rùa
120. Tấm Cám
121. Cha con Đăm Bông Pha
122. Quả cà thần
123. Chàng ngọc sẵn hươu
124. Trồng tre lên trăng
125. Hơ Mênh chém rắn thần
126. Mưa, gió, mặt trời và mặt trăng
127. Sự tích củ mài và cây cơm xôi
128. Thạch Sanh
129. Lươn thần và cậu bé nghèo khổ
130. Lết và Le
131. Hai chú Cuội
132. Bắn tài
133. Nàng Lòa, con ngựa mù và chàng Thông minh
134. Chàng đánh cá Y Ang
135. Cái ang vàng
136. Cô bé chăn vịt
137. Mồ Côi xử kiện
138. Nghề đặc biệt
139. Nàng Cu Pên
140. Hai anh em
141. Chiếc thoi vàng

Appendix G

List of e-articles used for data in the Vietnamese Online Newspaper Corpus

<https://vnexpress.net>

No.	Title of the article	Publishment (D/M/Y)
1	'Còn mảnh nhôm ghim ở tay nạn nhân điều thuốc lá phát nổ'	25/12/2011
2	Cây mai vàng giá 2,4 tỷ đồng	18/01/2020
3	Bộ ria mép đại sứ khiến người Hàn nổi giận	18/01/2020
4	Bản di chúc tạo cơn sốt sinh con	17/01/2020
5	Cha Meghan nói con gái 'làm mất giá' hoàng gia Anh	19/01/2020
6	Điều gì giúp Albert Einstein trở thành thiên tài?	17/01/2020
7	Năm phương pháp nuôi dạy trẻ song ngữ	19/01/2020
8	Con chó nhồi bông giúp kết tội kẻ giết người	12/01/2020
9	Ký ức về hộp mứt Tết giúp đòi lại công lý	10/01/2020
10	Cuộc hẹn của người đàn bà ngoại tình	11/01/2020
11	Vật bất ly thân của ngôi sao golf Mỹ	12/01/2020
12	Cựu kỷ lục gia marathon dính bê bối doping	11/01/2020
13	Vận rủi của cựu vô địch PGA Tour	12/01/2020
14	Trump muốn tái đắc cử mới ký nốt thỏa thuận với Trung Quốc	10/01/2020
15	Đường cất cánh gian nan của máy bay Trung Quốc	10/01/2020
16	Chính phủ Nhật Bản khổ vì giá cả quá ổn định	13/01/2020
17	Iran thừa nhận bắn nhầm máy bay Ukraine	11/01/2020
18	Cô gái Việt 'rớt tim' trên chuyến bay tới Iran	11/01/2020
19	Iran 'thấy xương gầy' dưới đòn trừng phạt của Mỹ	13/01/2020
20	Thanh tra CP: Kết luận của HN về đất đai ở Đồng Tâm là 'chính xác'	25/04/2019
21	Đào Nhật Tân khoe sắc	13/01/2020
22	Việt Nam - Jordan: Cuộc chiến chuyển trạng thái	13/01/2020
23	Uber hợp tác với Hyundai phát triển taxi bay	08/01/2020
24	Người đi xe đạp có nồng độ cồn sẽ bị phạt 600.000 đồng	31/12/2019
25	Xe bồn đi lùi trên cao tốc Hà Nội - Hải Phòng	01/01/2020

26	Giấc mơ công nghiệp	15/01/2020
27	Giá của khẩu trang	05/02/2020
28	Tin vào Y học	03/02/2020
29	Cuộc chiến chống dịch	01/02/2020
30	Lương lãnh đạo	31/01/2020
31	Viêm phổi Vũ Hán	30/01/2020
32	Tết của trẻ con	29/01/2020
33	Khởi đầu mới	25/01/2020
34	Tết của dân quê	22/01/2020
35	Ngày sám hối	23/01/2020
36	Sống dễ lắm	24/01/2020
37	Những trào lưu sắm Tết	21/01/2020
38	Đi trong lòng mùa xuân	18/01/2020
39	Tết trong căn nhà rộng	15/01/2020
40	Nước mắt người trồng hoa	10/01/2020
41	Cơn khát điện	07/01/2020
42	'Cảm ơn' dịp Tết	08/01/2020
43	Tiền dưới lòng thành phố	02/01/2020
44	Ưu tiên của quốc gia	04/01/2020
45	Hai mặt của kiêu hời	29/12/2019
46	Cơ hội cuối cùng	28/12/2019
47	Bóng ma trầm cảm	27/12/2019
48	Những đột phá đi đâu?	26/12/2019
49	'Tội đâu anh chịu'	25/12/2019
50	Thương Tết	01/01/2020
51	Cái chết của đồng bằng	01/01/2019
52	Tầm nhìn bãi rác	21/08/2019
53	Những dòng sông chết	20/06/2019
54	Vòng tròn của niềm tin	11/02/2019
55	Tham nhũng tàng hình	17/12/2019
56	Cho tiền ăn xin	26/08/2019

57	Chuyện tay buôn người	02/11/2019
58	Lịch sử cho ai	17/07/2019
	www.dantri.com.vn	
59	Ngôi trường tiểu học "5 trong 1" ở quần đảo Trường Sa	22/02/2020
60	Chất lượng không khí diễn biến xấu: Hà Nội “đứng đầu”	23/02/2020
61	Điều tra án tham nhũng: Đến cơ quan giám định cũng “né” trách nhiệm	21/02/2020
62	Gần một tháng không mưa, ĐBSCL quay cuồng đối phó hạn-mặn	23/02/2020
63	Sóng lớn đánh chìm tàu, 5 thuyền viên may mắn được cứu sống	22/02/2020
64	Ông Trump nổi giận vì quan chức Mỹ để 14 người nhiễm virus corona về nước	16/02/2020
65	Real Madrid đón tin dữ sau trận thua sốc	23/02/2020
66	Man Utd - Watford: Thời cơ vàng cho “Quý đỏ”	23/02/2020
67	Bộ GD&ĐT đề nghị các tỉnh cho học sinh đi học từ ngày 2/3	23/02/2020
68	Đề xuất thành lập lại Bộ G.đục: Hãy duy trì nguyên trạng hệ thống ĐH	22/02/2020
69	Mẹ nam sinh Việt nói 8 thứ tiếng: “Dạy con thành nhân trước khi thành tài”	22/02/2020
70	Xúc động những bức thư từ tâm dịch Sơn Lôi hỏi âm học sinh Hà Nội	21/02/2020
71	Hài nhi ngừng tim 10 phút, bạn đọc giúp đỡ 1,2 tỉ đồng cứu sản phụ	26/02/2020
72	Buốt lòng 3 trẻ mồ côi, bé 12 tuổi: Con cố học hết lớp 9 rồi đi làm thuê	23/02/2020
73	Nỗi sợ dịch Covid-19 đẩy giá vàng tăng mạnh lên mức cao mới	22/02/2020
74	Xe nhập về nhỏ giọt, nguy cơ “sốt ảo” vì dịch covid-19	23/02/2020
75	Kinh hoàng: Phát hiện xưởng làm khẩu trang kháng khuẩn bằng giấy vệ sinh	13/02/2020
76	Loại quả giới nhà giàu Dubai ăn mỗi ngày từng được trồng ở Việt Nam?	16/02/2020
77	Vì sao xe hơi châu Âu về VN giảm mạnh trước khi EVFTA được ký kết?	17/02/2020
78	4 ngôi nhà "ngập cây xanh và ánh nắng" của VN được báo Mỹ khen nức nở	23/02/2020
79	Khám phá ngôi làng thơ mộng bên bờ biển, đẹp như bước ra từ cổ tích	23/02/2020
80	Doanh nghiệp bất động sản đang gặp khó khăn với những dự án nào?	22/02/2020
81	Hà Nội: Khách hàng lại căng băng rôn đòi nhà 8B Lê Trực	09/02/2020
82	Sừng sốt trước những bức tranh tỉ mỉ tới từng... chấm màu	09/02/2020
83	Yêu cầu phục hồi di tích cầu Ngõi chợ Thượng nguyên trạng	22/02/2020
84	Bức họa trị giá 615 tỷ đồng của Picasso bị phá hoại khi đang trưng bày	23/02/2020
85	Cuộc sống của NSUT Ng. Huyền như thế nào sau khi chia tay C. Trung?	02/01/2020

86	Hình tượng chuột trong đời sống văn hóa Á Đông	10/01/2020
87	Những bộ phim hay nhưng “lỡ duyên” nổi tiếng	25/01/2020
88	Tháng ngày lập nghiệp đầy khó khăn của đạo diễn “Ký sinh trùng”	22/02/2020
89	Hình ảnh phun trào tuyệt đẹp của núi lửa băng	11/02/2020
90	Hoàng gia Thụy Điển và “bộ sưu tập” cung điện ấn tượng	21/02/2020
91	Mãn nhãn với mùa “hoa tuyết” trên đại ngàn Tây Nguyên	22/02/2020
92	TS. Nguyễn P.H. - "Người truyền lửa" cho phong trào DHS Việt tại Đức	22/02/2020
93	Hồ Hoài Anh tặng Lưu Hương Giang nhẫn kim cương sau ồn ào ly hôn	28/10/2015
94	Ca sĩ Bằng Kiều tiết lộ ảnh hiêm thuở nhỏ	05/03/2020
95	Thuyền trường nhảy từ độ cao 12m nguy hiểm để cứu khách sắp đuối nước	03/03/2020
96	Điểm danh những bãi biển "quyến rũ" nhất thế giới	03/03/2020
97	Một phụ nữ nghi bị sát hại dã man tại chòi canh rẫy	04/03/2020
98	Phó Chủ tịch TP Nha Trang Lê Huy Toàn bị tuyên 9 tháng tù	05/03/2020
99	HN: Khởi tố vụ án, tạm giữ nghi can đốt pháo “đỏ đường” tại đám cưới	28/02/2020
100	Những lý do càng yêu lâu càng dễ chia tay	04/03/2020
101	Ngắm bộ ảnh cưới "tình bể bình" của cầu thủ Phan V Đ và hot girl NA	05/03/2020
102	Nhiều chất độc, chất gây ung thư có trong thuốc lá điện tử	14/01/2020
103	Những chiếc “túi hóa trị” đậm tình người của người phụ nữ mắc ung thư	05/03/2020
104	Những ai cần đề phòng ung thư thực quản “ghé thăm”	04/03/2020
105	Nguyên nhân máy giặt rung lắc, gây tiếng ồn lớn và cách khắc phục	03/03/2020
106	Mạng Tiktok lại rộ lên thử thách nguy hiểm, có thể gây chết người như chơi	05/03/2020
107	Những smartphone có camera tốt nhất trong năm 2019	04/03/2020
108	Điểm danh những smartphone có thiết kế đặc biệt nhất trong năm 2019	30/11/2019
109	Kì lạ hai xe cùng chèn ép, "bắt nạt" xe cứu thương	18/12/2019
110	Xe Mazda phóng nhanh, đánh võng như "say rượu" giữa phố đông	04/03/2020

www.vietnamnet.vn

111	Chồng bật cười gặp vợ dẫn hàng chục người đến nhà nghỉ đánh ghen	06/03/2020
-----	--	------------

www.tuoiitre.vn

112	Cuộc sống cách ly phòng dịch COVID-19 ở Canada, Ý, Hàn	09/03/2020
-----	--	------------

113	Bộ Công an vào cuộc vụ thang máy giả, 'không khởi tố được' ở BD	01/01/2020
114	Cái tát mạnh vào Boeing	09/03/2020
115	Bị cáo Trương D Nhất: 'Tôi phải làm theo sự phân công và qđịnh của tổng b. tập'	09/03/2020
116	Dòng tiền vào chứng khoán khá cao dù VN-Index mất gần 56 điểm	09/03/2020
117	Giá xe hơi có thể giảm mạnh nhờ 'cú đâm' chính sách mới	17/02/2020
118	Cuộc sống diệu kỳ ở nơi đá nở hoa Mèo Vạc	03/02/2020
119	Ngắm mình từ bên trong	08/03/2020
120	Paris có một... cà phê vỉa hè chất Việt	08/03/2020
121	‘Chuyên đi để đời’ sưởi ấm những trái tim	06/03/2020
122	Thắng trận derby nước Ý, Juventus bỏ Inter Milan lại trong cuộc đua vô địch	09/03/2020
123	Cần sự quyết liệt trong quản lý để giải quyết các vấn nạn môi trường	25/02/2020
124	Giấy vệ sinh có tác dụng gì mà dân tình đổ xô đi mua?	07/03/2020
125	Hoàng đầu ấn - loài hoa người xưa gặp là sợ, giờ thành điểm 'check-in'	09/03/2020
126	Nộp phạt qua mạng: Tiết kiệm và tiện cho dân	08/03/2020
127	Có một thời Hà Nội như thế, thời mẹ tôi	08/03/2020
128	Còn có thể ăn cái tết đậm ấm nào với bố mẹ thì ráng mà về!	28/01/2020
129	'Ngoại mừng quá, không ngủ được con à'	23/01/2020
130	Đầu năm, tôi cảm thấy rất may mắn!	28/01/2020
131	'Có ai biết đường ra Bắc không?'	18/01/2020
132	Khoảnh khắc thay đổi đời tôi: Cảm ơn bố mẹ đã buông tay con!	06/11/2019
133	Tôi hận cái nghèo	30/10/2019
134	Tám bằng đại học ngủ đông của tôi	10/10/2019
135	Tôi trả lại nụ cười cho chồng con	16/09/2019
136	Mẹ kể tôi 'xù lông nhím', trở thành siêu nhân bảo vệ tôi trước 'kẻ thù'	07/09/2019
137	‘Đừng để sự ra đi của ba là vô nghĩa’	31/08/2019
138	Tôi sốc nặng khi bác sĩ kết luận cặp thiên thần sinh đôi của tôi đều tự kỷ	24/08/2019
139	Chiếc bánh mì cha không kịp ăn!	08/08/2019
140	Ngọn đèn không tắt.	25/06/2019

Appendix H

List of talk show episodes used for data in the Vietnamese Spoken Corpus

1. TS 1 Talk Show Xin chào Việt Nam
<https://www.youtube.com/watch?v=pLPCLie9290>
2. TS 2 Chuyện đêm muộn: Cảm xúc của đàn bà.
<http://vtv.vn/video/chuyen-dem-muon-23-7-2017-241983.htm>
3. TS 3 Điều ước thứ 7- số 125- Bản hòa tấu cha và con.
<https://www.dailymotion.com/video/x627csh>
4. TS 4 Chuyện đêm muộn: Tuổi nghề
<https://vtv.vn/video/chuyen-dem-muon-02-10-2017-250720.htm>
5. TS 5 Điều ước thứ 7 số 122 - Chuyển xe bus tình yêu.
<https://www.youtube.com/watch?v=1PQqQfHudSY>
6. TS 6 Điều ước thứ 7 số 95 – Hãy cười như Thu
<https://www.facebook.com/BacGiangQueHuongToi.vn/videos/928145113977779/>
7. TS 7 Điều ước thứ 7 số 128 – Bài giảng cuối cùng của người lái đò
<https://vtv.vn/video/dieu-uoc-thu-7-21-10-2017-254859.htm>
8. TS 8 Vượt qua cảm xúc tiêu cực như thế nào?
<https://www.youtube.com/watch?v=wF0wNxwh5OE>
9. TS 9 Người Đương Thời - Đỗ Đức Cường - Cha đẻ máy ATM
<https://www.youtube.com/watch?v=ZFa5F-C371M>
10. TS10 Người Đương Thời - Diễn giả Trần Đăng Khoa
<https://www.youtube.com/watch?v=Q0dklxeU9Ck>
11. TS 11 Người Đương Thời - Nghệ sĩ ưu tú Thành Lộc
<https://www.youtube.com/watch?v=4mm1CCftfgo>
12. TS 12 Chuyện đêm muộn
<http://vtv.vn/video/chuyen-dem-muon-01-9-2017-243867.htm>
13. TS 13 Vũ Cát Tường chia sẻ về hình ảnh nữ tính trong MV Come Back Home vừa ra mắt.
<https://www.youtube.com/watch?v=sex7G1tb8yg>
14. TS 14 Vũ Cát Tường livestream on Yan

<https://www.youtube.com/watch?v=1BDJgzJZA6c&list=PLQPGKzLOcHEHFwd43a5lTwq1DBkKSvAgR&index=5>

15. TS 15 Chuyện đêm muộn

<http://vtv.vn/video/chuyen-dem-muon-29-9-2017-250069.htm>

16. TS 16 Chuyện đêm muộn – Đi mừng cưới

<https://www.youtube.com/watch?v=0vWBSKMHEY4>

17. TS 17 Chương trình Ghế không tựa

<https://www.youtube.com/watch?v=D2Fo6lRaYws>

18. TS 18 CeeMe Livestream

<https://www.youtube.com/watch?v=9pxc38jz1AU>

19. TS 19 Livestream Hoà Minzy

https://www.youtube.com/watch?v=_7BjNFCfrp8

20. TS 20 Trò chuyện cùng sao – Khách mời: Vũ Cát Tường

<https://www.facebook.com/baothethaovanhoa/videos/10155968754094885/>

21. TS 21 Chuyện đêm muộn: Tri thức 4.0

<https://vtv.vn/video/chuyen-dem-muon-10-4-2019-361170.htm>

22. TS 22 Talk show Chuyện đêm muộn – Hà Anh: Phụ nữ và Đức Hy sinh

<https://www.youtube.com/watch?v=LWIAPEf9LLs>