

A corpus-based analysis of Vietnamese ‘classifiers’ *con* and *cái**

Giang PHAM
University of Minnesota

Kathryn KOHNERT
University of Minnesota

Abstract

The purpose of this study is to examine the use of two common Vietnamese ‘classifiers,’ *con* (animacy) and *cái* (inanimacy) using language corpora data of over one million words. This information may contribute to an ongoing debate of whether Vietnamese ‘classifiers’ are a distinct word class or a subclass of nouns. Frequency and distributions were calculated using computer software. Lexical-semantic functions were manually analyzed for each occurrence. Findings indicated that *con* and *cái* were highly frequent and distributed across text genres. However, neither form consistently demonstrated a classifying function: *con* indicated animacy less than 24% of the time, and *cái* indicated inanimacy less than 65% of the time. Corpus-based analysis is a useful tool to make comparisons between prototypical and ‘real-life’ language use. If Vietnamese ‘classifiers’ are not consistently used as such, considering this group of words a subclass of nouns rather than a distinct word class may be more parsimonious.

Classifiers are words used to categorize word classes based on an attribute such as shape, function, or animacy. They are found in numerous languages and often reflect how specific cultures perceive the world (Aikhenvald, 2000). Historically, classifiers were thought of as function words or ‘empty words’ (Daley, 1998) that were ‘restricted’ or ‘not independent’ due to their inability to stand alone as a noun phrase (Thompson, 1965). Currently, linguists argue that this group of words should be considered content words since they contain varying degrees of lexical-semantic meaning (Luu, 2000). In addition, Cao (1998; 2000) argued that this group of words is no more ‘restricted’ than other words that are considered to be count nouns such as *lần* ‘time’ and *noi* ‘place’.

Vietnamese has been reported to have as few as three words that arguably could be considered ‘classifiers’ (Cao, 1998) to as many as 200 classifiers (e.g. D. H. Nguyen, 1957); the exact number and word list are

*Funding for this project was provided by the Graduate Research Partnership Program at the University of Minnesota, awarded to the first author under the faculty mentorship of the second author. We thank the Center for Cognitive and Social Processes in Language for use of equipment. Please address correspondence to Giang Pham in the Department of Speech-Language-Hearing Sciences, 115 Shevlin Hall, 164 Pillsbury Drive S.E., University of Minnesota, Minneapolis, MN 55455 or via email at tangx098@umn.edu or to Kathryn Kohnert at kohne005@umn.edu.

debated (Adams, 1989; Aikhenvald, 2000). Linguists with the largest estimate of classifiers such as D. H. Nguyen (1957) considered all words that occurred in the slot adjacent to the number in a noun phrase as classifiers. Linguists with low to no estimates do not consider ‘classifiers’ a distinct word class (Cao, 1998).

There is much debate on whether Vietnamese ‘classifiers’ should be considered a separate word class (e.g., Adams, 1989; Aikhenvald, 2000; Luu, 2000) or a subclass of nouns (Cao, 1998; 2000; C.T. Nguyen, 1999; D. H. Nguyen, 1957). In addition, it is debated whether a ‘classifier’ is a subconstituent in a noun phrase (Thompson, 1965) or the nucleus (e.g., Cao, 1998; 2000; C. T. Nguyen, 1999). Early work on Vietnamese classifiers (e.g. Emeneau, 1951) considered this group of words a separate word class with the meaning of ‘one unit quantity’. Thompson (1965) argued that classifiers are distinct from nouns since they do not occur in single-word descriptive complements. In addition, Luu (2000) argued that the functions of indicating kind and quantity distinguish classifiers from nouns.

Cao (1988; 1998; 2000) argues that ‘classifiers’ should not be considered a separate word class from nouns but rather a subset of count nouns that have an additional semantic function of classification. Cao (2000) suggested that if the function of ‘classifiers’ is to categorize or classify words (in most cases nouns) its function must be to classify nouns into smaller subclasses that indicate different kinds of objects. Based on this definition, Cao (1998; 2000) admitted to only three Vietnamese count nouns that may have this function of indicating kind: *cái* (inanimate), *con* (animate), and *người* (human).

The present study focuses on two of the most common Vietnamese count noun constituents or ‘classifiers’ known to have the function of indicating kind: *con* for animate objects and *cái* for inanimate objects. Lexical-semantic function in actual usage of these word forms is examined using corpora data. In the field of corpus linguistics, a corpus (‘corpora’ in the plural form), may be generally defined as a collection of multiple samples of spoken or written language. A corpus may vary in size, composition (e.g., genre or text type), and format (printed or machine-readable). When large enough in number and adequate in variety of samples (according to one’s purpose), language corpora may reveal much information on the frequency and distribution of linguistic patterns that are exemplars of ‘real life’ language use (McEnery & Wilson, 1996). Large corpora have been collected on English (e.g., Kucera & Francis, 1967) and many other languages (see Wilson, 2006 for review).

In corpus-based analyses, frequency and distributional characteristics are often presented in small percentages. Readers unfamiliar with corpus data may have difficulty in determining whether the frequency of a lexical item is considered high or low. For example, based on a corpora of American English by Kucera and Francis (1967) the three most frequent words in American English were “the” with 6.90% frequency, “of” with 3.59% frequency, and the word “and” with 2.84% frequency. These percentages are the highest possible in this language corpora, and there is quite a range of percentages that may be

considered highly frequent. For example, the 100 most frequent words based on over one million words in Kucera and Francis (1967) ranged from 6.90% to 0.09% frequency.

In this study, we examine the frequency and distribution of *con* and *cái* and their lexical-semantic functions in daily language. This information may contribute to a clearer understanding of how *con* and *cái* are used in ‘real-life’ situations as well as further inform the ongoing linguistic debate on the categorization of Vietnamese ‘classifiers’. To begin our discussion, we first introduce the Vietnamese language corpus used for the analysis. We then describe lexical-semantic functions of *con* and *cái*.

Vietnamese language corpus

The Corpora of Vietnamese Texts (Pham, Kohnert, & Carney, 2008) is the first language corpus in Vietnamese available electronically at http://vnspeechtherapy.com/vi/CVT/3_CVT_The%20Basics.htm. The CVT consists of over one million Vietnamese words from newspaper articles and children’s literature. Since one of the original goals of the CVT was to investigate language use in Vietnamese Americans, texts were collected primarily from Vietnam and the United States. Newspaper articles included a variety of topics such as world news, news in Vietnam, health, economics, education, sports, editorials, science, and culture. Children’s literature consisted of over 350 children’s picture books. The vast majority of books were published in Vietnam due to the relatively limited availability of children’s books in Vietnamese from other countries. Picture books that were published outside of Vietnam were primarily from the United States and England with a few books published in Australia and New Zealand.

Lexical-semantic functions

Little is known about historical uses of *con* and *cái* within the Vietnamese language (Le, 2000; S.N. Nguyen, 2000). These two words have been reported to have different historical origins. According to G.T. Nguyen (2003), the word *con* has its roots in Mon-Khmer languages, while S.N. Nguyen (2000) reported that *cái* originated from Chinese. This study focuses on current uses of these forms. The following is an introduction to the lexical-semantic functions of word forms *con* and *cái*. Abbreviations “cl-animate” and “cl-inanimate” are used to refer to the classifying function of these words.

Con

The prototypical semantic function of the word form *con* as a classifier is to indicate animacy. It is primarily used with non-human animate things (D. H. Nguyen, 1957) and may occur before animal names to refer to each individual member of that group of animals such as *con vịt* ‘duck’, *con gà* ‘chicken’, and *con bò* ‘cow’. In addition to animals, *con* is used to refer to

other ‘non-human living things’ (D. H. Nguyen, 1957) such as ghosts (*con ma*) and monsters (*con quái*).

Adams (1989) and D. H. Nguyen (1957) reported some exceptional uses of *con* as a ‘classifier’ of animacy. For example, *con* is used with some inanimate objects such as postage stamps (*con tem*) and seals (*con dấu*). Adams (1989) and H.V. Nguyen (2003) cite the use of *con* for *inanimate* objects that could have been originally viewed as being animate or in motion such as river (*con sông*), street (*con đường*), spindle (*con thoi*) and body parts such as heart (*con tim*) and eye (*con mắt*). Although there are a few cases in which *con* is used to refer to persons considered ‘less than human’ (Adams, 1989) such as a prostitute (*con điểm*) or gambler (*con bạc*), D. H. Nguyen (1957) denied that the use of *con* necessarily has a derogatory connotation and gave the neutral example of *con người* ‘person’.

Apart from its classifier meaning of animacy, Tan (1994) reported numerous other meanings for the word *con*. Perhaps the most familiar meaning of the word *con* for native speakers is ‘child’. *Con* may signify a human child (e.g., *con của mẹ* [child of mother] ‘mother’s child’), the offspring of an animal (e.g., *gấu con* [bear offspring] ‘cub’), and more metaphorically the fruit of a plant (e.g., *Cây chuối này được năm con* [Tree banana this to-be-able five *con*] ‘This banana tree has five fruits’). *Con* may be used as a kinship term for ‘child’ in a parent-child relationship as well as non-familial relationships with comparable age differences between the speaker and listener. For instance, *con* may be used to refer to oneself when speaking with listeners approximately the age of one’s parents or to address a listener who may be approximately the age of one’s child (Luong, 1990). *Con* is also found in compound nouns such as *bà con* [grandmother child] ‘relatives’.

The prototypical syntactic order for ‘classifiers’ in Vietnamese consists of a numeral/quantifier + ‘classifier’ + noun (Aikhenvald, 2000). H.V. Nguyen (2003) suggested that *con* may occur without a numeral or quantifier preceding it when it is followed by a distinctive modifier (e.g., *Con chó đang ngủ rất dữ* [Cl-animate dog present-aspect sleep very fierce] ‘The dog that is sleeping is very fierce’) or a demonstrative (e.g., *con mèo đó* [cl-animate cat that] ‘that cat’). Cao (1998) stated that when the referent is known in conversation, the referent noun may be omitted, and *con* may stand alone as a pro-form. For example, in a market a customer may request *Bán cho tôi bốn con gà* [Sell for me four cl-animate chicken] ‘Sell me four chickens’. The vendor may ask *Bao nhiêu?* ‘How many?’ and the customer may respond *Bốn con* [Four cl-animate] ‘Four [of them]’. In these cases, the semantic function of *con* is individuation or representing individual units of the referent.

Cái

The ‘classifier’ *cái* is commonly known to indicate inanimacy (D. H. Nguyen, 1957). According to H.V. Nguyen (2003), the word *cái* may be used as a classifying noun that ‘unitizes’ the noun that follows it. Similar to *con*, the ‘classifier’ *cái* may also stand alone if the referent noun is defined in previous

contexts such as *Bán cho tôi ba cái* [Sell for me three ones] “Sell me three [of them]”. *Cái* may also be used to nominalize a verb as in *cái chết* [cl-inanimate to die] ‘death’ or an adjective as in *cái đẹp* [cl-inanimate beautiful] ‘beauty’ (H. V. Nguyen, 2003). In addition, Adams (1989) noted that *cái* may indicate the female gender as in *con mèo cái* [cl-animate cat cl-female] ‘a female cat’.

C.T. Nguyen (1999) found another lexical-semantic function for *cái* as an emphasizer used to point out an item from the remaining ones or to point out something in disgust. For example, *cái* may be used freely in front of a noun phrase that contains a restrictive/determinative attribute such as *cái quyển sách ở trên bàn ấy* [cl-emphasizer cl-volume book at top table that] ‘that book on top of the table’ (H.V. Nguyen, 2003). To express disgust, *cái* may appear with nouns that refer to people or animals such as *cái thằng ngu ấy* [cl-emphasizer pron-young male stupid that] ‘that stupid boy’.

Methods

The purpose of this study was to analyze the functions of Vietnamese ‘classifiers’ *con* and *cái* based on corpora data from the Corpora of Vietnamese Texts (CVT: Pham, Kohnert, & Carney, 2008). There were two research objectives: 1) to investigate the frequency and distributional characteristics of Vietnamese ‘classifiers’ *con* and *cái* and 2) to examine the lexical-semantic functions commonly reflected in these word forms.

To address the first objective, a computer software concordance program, MonoConc Pro 2.2 (Barlow, 2003), was used to calculate frequencies and distributions of the word forms *con* and *cái* found in the CVT. Calculations were based on the entire corpus for a sample size of over one million word forms. To address the second objective, two separate concordance lists were generated using MonoConc Pro 2.2 with *con* and *cái* as the target words, respectively, which provided the sentential context for each occurrence of *con* and *cái* in the corpus. These concordance lists were manually analyzed and categorized into lexical-semantic functions by the first author who is bilingual/biliterate in Vietnamese and English. The word form *con* was separated into six working categories based on its possible meanings: 1) meaning of child, 2) indicate animacy, 3) compound noun, 4) kinship term, 5) meaning small, and 6) an “other” category for the remaining items. The word form *cái* was separated into six functional categories: 1) indicating inanimacy, 2) nominalization, 3) indicating individual unit, 4) meaning female, 5) addition of emphasis, and 6) “other”. Percentages of lexical-semantic functions were calculated based on this categorization system.

Results

This study used corpora data to empirically investigate linguistic properties of the two most common Vietnamese ‘classifiers’ *con* and *cái*. First, frequency and distributional characteristics were calculated based on the entire

corpora data. Second, word forms were categorized into lexical-semantic functions.

Frequency and distributional characteristics

The lexical items *con* and *cái* are among the 150 most frequent words in the CVT (Pham et al, 2008). Based on a total of 1,063,912 lexical items in the CVT, *con* was ranked 17th in overall frequency and *cái* was ranked 118th. The form *con* occurred a total of 4,857 times and *cái* occurred a total of 1,627 times in the entire sample. Table 1 displays the distribution of *con* and *cái* in the entire corpus. Distributions are listed as percentages of occurrences ($[\text{\#occurrences}/1,063,912] \times 100$) in increments of ten percent. For example, based on the entire corpora the target word *con* occurred with 0.129% frequency in the first ten percent (0-10%) and with 0.151% in the second ten percent (10-20%) and so forth. This method of dividing the corpora by 10 percent increments is consistent with previous large language corpora analyses (cf. Kucera & Francis, 1967) as a gross measure of lexical distribution across a wide variety of text genres. Although *con* occurred more frequently than *cái*, both forms were shown to be distributed throughout the corpora. *Con* and *cái* were not restricted to any particular type of text, although their distributions did vary depending on text type. For example, based on the entire corpora, *cái* occurred with 0.041% frequency in the second 10 percent (10-20%) and with 0.004% in the fourth 10 percent (30-40%). These findings are consistent with corpora data on English by Kucera and Francis (1967) who found that highly frequent words occurred throughout the corpora with a certain amount of variability across text genres.

Table 1. Distribution of *con* and *cái*

% corpus	% occurrences/corpus total	
	Con	Cái
0-10	0.129	0.031
10-20	0.151	0.041
20-30	0.041	0.013
30-40	0.016	0.004
40-50	0.023	0.013
50-60	0.021	0.008
60-70	0.025	0.012
70-80	0.010	0.010
80-90	0.028	0.014
90-100	0.013	0.008
Total	0.457	0.153

Note. Based on the entire CVT (Pham et al, 2008).

Lexical-semantic functions of *con* and *cái*

We now turn our attention to the lexical-semantic functions of *con* and *cái*. Table 2 displays the percent of *con* occurrence according to its lexical-semantic functions in the order of most to least frequent. The most frequent lexical-semantic function of *con* was to indicate animacy, which occurred 23.97% of the entire corpora. It is important to note that although in the descriptive linguistics literature, animacy is the primary lexical-semantic function given to *con*, this function occurred in less than 24% of the entire sample. The second most frequent function was to serve as part of a compound noun 19.39% of sample such as in the word *bà con* ‘relatives’. The third most frequent lexical-semantic function of *con* was to signify ‘small’ or ‘offspring’, which occurred in over 18.49% of the sample. *Con* was used as a kinship term to represent the speaker (first person) or the listener (second person) in over 16.20% of the sample. The “Other” category consisted of lexical items traditionally referred to as exceptions to the classifying function of inanimacy such as *con đường* ‘street’ and comprised 12.56% of the data. Finally, *con* signified ‘child’ or an individualized unit in 9.39% of the sample.

Table 2. Lexical-semantic categories of the word form “con”

Function	Example	% (#Occurrence/Total)
Indicates animacy	<i>Con gấu</i> [cl-animate bear] ‘bear’	23.97% (1164/4857)
Compound noun	<i>Bà con</i> [grandmother child] ‘relatives’	19.39% (942/4857)
Small or offspring	<i>Gấu con</i> [bear small] ‘cub’	18.49% (898/4857)
Kinship term	<i>Con chưa buồn ngủ.</i> [Child-Speaker not-yet to-want sleep] ‘I am not sleepy yet’	16.20% (787/4857)
Other	<i>Con đường</i> ‘street’	12.56% (610/4857)
Child or Individual unit	<i>Hai con của anh</i> [two child of slightly older man-3 rd person] ‘His 2 children’	9.39% (456/4857)
Total		100% (4857/4857)

Note. The percent of *con* occurrence is presented according to its lexical-semantic functions in the order of most to least frequent.

Table 3 displays the percent of *cái* occurrence according to its lexical-semantic functions in the order of most to least frequent. The most frequent lexical-semantic function of *cái* was to indicate inanimacy (> 65%). It was noted that *cái* as an indicator of inanimacy was observed to precede concrete nouns with 70.86% (754/1064) frequency and abstract nouns such as *cái quyền* ‘privilege’ with 29.14% (310/1064) frequency. The second most frequent

lexical-semantic function of *cái* was nominalization (19.48%) in which an adjective or verb becomes a noun when preceded with *cái* as in *cái chết* [cái to-die] ‘death’. The third most frequent lexical-semantic function was to stand alone to indicate an individual unit (11.37%) such as *bị đá một cái* [passive-marker to-kick one unit] ‘to get kicked one [kick]’. Lexical-semantic functions that comprised less than 4% of the data included an “Other” category, the function of emphasis (deictic), and the indication of a female animal such as *hươu cái* [caribou female] ‘doe’. The ‘Other’ category comprised words that were difficult to classify such as lexical items *con cái* ‘[one’s] children’ or onomatopoeias such as [tế] *cái rầm* [[to-fall] cái thump] ‘to fall –[a] thump!’

Table 3. Lexical-semantic categories of the word form “cái”

Function	Example	% (#Occurrence/Total)
Indicates inanimacy	<i>Cái ghế</i> ‘chair’	65.40% (1064/1627)
Nominalization	<i>Cái chết</i> [Cái to-die] ‘death’	19.48% (317/1627)
Individual unit	<i>Một cái này</i> [one Cái this] ‘This one’	11.37% (185/1627)
Other	<i>Con cái</i> ‘[one’s] children’	2.09% (34/1627)
Emphasis	<i>Cái ông bất lịch sự này!</i> [Cái man without respect this] ‘This irresponsible man!’	1.04% (17/1627)
Female	<i>Chuột cái</i> [mouse female] ‘female mouse’	0.61% (10/1627)
Total		100% (1627/1627)

Note. The percent of *cái* occurrence is presented according to its lexical-semantic functions in the order of most to least frequent.

In summary, *con* and *cái* have been shown to be highly frequent lexical items that occur across text genres. As shown in tables 2 and 3, there are multiple lexical-semantic functions associated with *con* and *cái* in addition to their well-known classifying functions of indicating animacy and inanimacy, respectively.

Discussion

This study examined the usage of two common Vietnamese ‘classifiers’ *con* and *cái* in daily language based on written corpora data. Findings indicated that the prototypical meanings of *con* and *cái* were to indicate animacy and inanimacy, respectively. However, the proportion of use differed across forms. *Con* demonstrated the semantic function of indicating animacy in less than one fourth of occurrences. *Cái* indicated inanimacy in the majority of

cases (> 65%) but still less than the proportion one would predict for a prototypical word. Corpus-based analysis is a useful tool to make comparisons between how linguistic forms are prototypically described and how forms are actually used in daily language. Awareness of these differences contributes to a more comprehensive description and understanding of common linguistic structures such as Vietnamese ‘classifiers’ *con* and *cái*.

One implication of this study is related to the ongoing debate in Vietnamese linguistics about the classification of ‘classifiers’. Certain linguists consider Vietnamese classifiers a separate word class (e.g., Thompson, 1965, Luu, 2000), while others argue that ‘classifiers’ are a subclass of nouns (e.g., D.H. Nguyen, 1957; Cao, 1988). As shown in Table 2, there were a small but substantial number of lexical items with *con* that were categorized as “Other”. Traditionally, this group of words would be considered exceptions to the rule of animacy. However, many exceptions to the rule may call into question the rule itself. As shown in Table 1, *con* indicated animacy less than a quarter of the time. In addition, although *cái* was relatively more consistent than *con*, it still had more than one lexical-semantic function. Given that neither form consistently demonstrated a classifying function (e.g., consistency may be operationally defined as > 80%), considering ‘classifiers’ as a separate word class based on their classifying function may not be supported. Considering linguistic forms such as *con* and *cái* as a subclass of nouns may be more parsimonious (Cao 1988; 1998).

Information on the actual uses of *con* and *cái* as well as other highly frequent words in Vietnamese may also be applied to the teaching of Vietnamese as a foreign language. Explicit instruction on the classifying function of words such as *con* and *cái* may be beneficial, particularly for speakers whose native language does not have the concept of classifiers. This study found that multiple lexical-semantic functions mapped onto these two forms in daily language use. Therefore, it may be helpful to have explicit practice with identifying the function or meaning of words in sentential or discourse contexts rather than studying single vocabulary items. Particularly because Vietnamese is an isolating language that does not use bounded morphemes to indicate time, possession, or person (Tang, 2006), a functional approach to learning grammar may prove more useful (Cao, 2000).

REFERENCES

- Adams, Karen L. 1989. *Systems of Numeral Classification in the Mon-Khmer, Nicobarese and Asian Subfamilies of Austroasiatic*. Canberra, Australia: Pacific Linguistics Series B-101.
- Aikhenvald, Aleksandra Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Barlow, Michael. 2003. *MonoConc Pro 2.2: A Professional Concordance Program* [Computer software]. Houston, TX: Athelstan.
- Cao, Hao. X. 1988. The count/mass distinction in Vietnamese and the concept of ‘classifier’. *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung* 41: 38-47.

- Cao, Hao X. 1998. *Tiếng Việt; Máy Ván đề Ngữ Âm, Ngữ Pháp, Ngữ Nghĩa* [Vietnamese: Issues in phonetics, syntax, and semantics]. Ho Chi Minh City: Nhà Xuất Bản Giáo Dục.
- Cao, Hao. X. 2000. “Nghĩa của loại từ [The meaning of classifiers].” In Trung tâm khoa học xã hội và nhân văn: Quốc gia viện ngôn ngữ học [Center of natural and social sciences: National institute of linguistics], eds., *Loại Từ Trong Các Ngôn Ngữ ở Việt Nam* [Classifiers of the languages in Vietnam], pp. 32-87. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Cao, Hao. X. 2004. *Tiếng Việt: Sơ Thảo Ngữ Pháp Chức Năng* [Vietnamese: Initial draft of functional grammar]. Ho Chi Minh City: Nhà Xuất Bản Giáo Dục.
- Daley, Karen A. 1998. *Vietnamese Classifiers in Narrative Texts*. Dallas, TX: Summer Institute of Linguistics, Inc. (Library of Congress Catalog No: 97-61945)
- Emeneau, Murray B. 1951. *Studies in Vietnamese Grammar*. Berkley, CA: University of California Press.
- Kucera, Henry & Francis, Winthrop N. 1967. *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Le, Thai X. 2000. “Loại từ trong tiếng Việt và trong tiếng Hán - Đồng nhất và khác biệt [Classifiers in Vietnamese and in Chinese – Similarities and differences].” In Trung tâm khoa học xã hội và nhân văn: Quốc gia viện ngôn ngữ học [Center of Natural and Social Sciences: National Institute of Linguistics], eds., *Loại Từ Trong Các Ngôn Ngữ ở Việt Nam* [Classifiers of the languages in Vietnam], pp. 107-115. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Luong, Hy V. 1990. *Discursive Practices and Linguistic Meanings: The Vietnamese System of Person Reference*. Philadelphia, PA: John Benjamins Publishing Company.
- Luu, Lang V. 2000. “Một số vấn đề về loại từ trong tiếng Việt [Some issues about Vietnamese classifiers].” In Trung tâm khoa học xã hội và nhân văn: Quốc gia viện ngôn ngữ học [Center of Natural and Social Sciences: National Institute of Linguistics], eds., *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.
- McEnery, Tony & Wilson, Andrew. 1996. *Corpus Linguistics: An Introduction*, 2nd ed. Edinburgh, UK: Edinburgh University Press.
- Nguyen, Can T. 1999. *Ngữ Pháp Tiếng Việt, in Lần Thứ Sáu* [Vietnamese grammar, 6th edition]. Hà Nội: Nhà Xuất Bản Đại Học Quốc Gia.
- Nguyen, Dinh H. 1957. “Vietnamese classifiers.” *Word* 13:124-152.
- Nguyen, Giap T. 2003. *Từ Vựng Học Tiếng Việt, Tài Bản Lần Thứ Tư* [Vietnamese semantics, 4th edition]. Ho Chi Minh City: Nhà Xuất Bản Giáo Dục.
- Nguyen, Hue V. 2003. *Từ Điển Ngữ Pháp Tiếng Việt Cơ Bản – Dictionary of Basic Vietnamese Grammar: Bilingual Vietnamese-English*. Ho Chi Minh City: Nhà Xuất Bản Đại Học Quốc Gia.
- Nguyen, San N. 2000. *Tìm Hiểu Tiếng Việt Lịch Sử* [Investigating the history of Vietnamese]. Ho Chi Minh City: Nhà Xuất Bản Đại Học Sư Phạm.

- Pham, Giang, Kohnert, Kathryn, & Carney, Edward. 2008. Corpora of Vietnamese Texts: Lexical Effects of Intended Audience and Publication Place. *Behavior Research Methods* 40:154-163.
- Tan, Van. 1994. *Từ Điển Tiếng Việt* [Vietnamese dictionary]. Ha Noi: Nhà Xuất Bản Khoa Học Xã Hội.
- Tang, G. 2006. "Cross-linguistic analysis of Vietnamese and English with implications for Vietnamese language acquisition and maintenance in the United States." *JSAAEA* 2:1-33.
- Thompson, Lawrence. 1965. *A Vietnamese Grammar*. Seattle, WA: University of Washington Press.
- Wilson, Andrew. 2006. *Corpus Linguistics Around the World*. New York: Rodopi.

Received: 12 September 2007

University of Minnesota
Department of Speech-Language-
Hearing Sciences,
University of Minnesota
<tangx098@umn.edu>