Unifying the semantics of class terms and classifiers in Vietnamese

Mike Pham
University of Chicago

1 Introduction

Vietnamese utilizes numeral classifiers in order to combine nouns with numerals, a phenomenon widespread in the region among different language families of East and Southeast Asia. It is well known that that in these languages, classifiers are often obligatory for the purposes of counting nouns, as is shown in (1).[1]

(1) a. tám *(con) chó
eight CLF:ANIM dog
'eight dogs'

b. hai *(cuốn) sách
two CLF:VOLUME book
'two books'

c. bốn *(trái) cam
four CLF:FRUIT orange
'four oranges'

d. bảy *(cây) dù
seven CLF:LONG-SKINNY umbrella
'seven umbrellas'

The examples in (1) also demonstrate that classifiers categorize nouns along various abstract semantic criteria, which can vary, but most often involve animacy, shape, size, and structure (Aikhenvald, 2000), though also function and social status (Denny, 1976). In other words, numeral classifier systems demonstrate an interaction between the syntax and the specific lexical semantics of the words involved. Being able to syntactically combine with numerals appears to be connected to not only the individuating functional semantics of classifiers, but also their lexical non-logical meaning (Roberts, 2010) that is sensitive to the semantic qualities of nouns.

Class term compounds complicate the matter by having mixed properties of both nouns and classifier phrases. Many are directly countable, which, assuming that class term compounds are compound nouns, runs counter to the generalization that nouns always require classifiers for the purposes of counting. Consider the following examples, where a numeral appears to be combining directly with a (compound) noun (highlighted in bold), with no classifier intervening in the canonical position:

(2) a. mười **nhà thương**
ten house injured
'ten hospitals'

b. ba **phòng khách**
three room guest
'three living rooms'

c. năm **bàn cờ**
five table chess
'five chessboards'

d. chín **cây chuối**
nine tree banana
'nine banana trees'

[1] Abbreviations used: CLF:(PROPERTY): classifier with semantic property specified (where no property is listed, I make no commitments to the semantic properties of the classifier); CT: class term; ANIM: animate; (IN)DEF: (in)definite; GEN: generic
Compounds in examples are listed in boldface

Some compounds are capable of taking optional classifiers:

(3) a. ba (chiếc) **máy bay**
three CLF:UNIT machine fly
'three airplanes'

b. sáu (người) **thợ may**
six CLF:PERSON worker sew
'six tailors'

While this makes class term compounds appear like classifier phrases, they share other properties with bare nouns, such as being underspecified for number.

(4) a. chó
dog
'dog(s)'

b. con chó
CLF:ANIM dog
'a/the dog(*s)'

(5) **cây chuối**
tree banana
'banana tree(s)'

In (4), the bare noun *chó* is underspecified for number, denoting both singular and plural sets (as well as having kind/generic interpretations); however the addition of the classifier *con* restricts the denotation of the resulting ClfP to singular sets of atomic dogs. (5) shows that class term compounds behave like bare nouns in this regard.

This variability in classifier omission for compounds seems to hint at a type of language change in Vietnamese where nouns participating in compound derivation are at least partially reanalyzed as classifiers – if not entirely so. DeLancey (1986) provides an analogous diachronic account of compounding in Thai, where he shows that class terms are a major source in the development of new classifiers. I propose that DeLancey's theory of class terms being a source of classifier emergence, which I return to later, also applies to Vietnamese, providing evidence in §3. If this is the case, it is not surprising that class terms show mixed properties of both bare nouns and classifiers. Employing the grammaticalization theory of von Fintel (1995), this occurs by composing in the functional meaning of a classifier – individuation, or access to sets of atomic individuals – into the meaning of the class term in compounds. In other words, as class terms are proto-classifiers, we expect them to share the same semantic function of individuation.

However, if this is the case, then no theory where classifiers enable the individuation or provide the unit of measure for the nominal complement is sufficient. For example, for Chierchia (1998) classifiers form a constituent with the NP in order to convert the kind denotations of nouns into atomic sets so that combination with numerals can occur; Krifka (1995) proposes that classifiers provide a specific measure function, but that they form a constituent with the numeral in order to combine with atomic nominal predicates. What these styles of analyses share is that they treat the entities being counted as being within the denotation of the noun, with the classifier providing the appropriate semantic function to allow composition of the numeral with the noun.

The problem with this view is two-fold given the class term compound data: first, not all class term compounds that can combine directly with numerals involve a nominal modifier. In *máy bay*, 'airplane' (lit. 'machine fly'), *bay* is a verb meaning 'fly'. If we expect that classifiers and class terms either create or access atomic sets of their nominal complements, it is unclear what (non-)atomicity looks like for verb denotations. This leads to the second, major problem with the traditional view of classifiers: what is being counted is not the

atoms in the noun, but rather atoms of what the entire compound denotes. This can be seen in (2d), where what is being counted is not bananas, but trees – specifically banana trees.

In this paper, I argue that what is being counted in Vietnamese numeral phrases is always the individuals denoted by the entire ClfP or class term compound, rather than just the (non-)nominal second element following the classifier/class term. As noted by Thompson (1965), both constructions share a head-modifier relationship. It is this modifier role that is performed by the second element of ClfPs and class term compounds: the head and modifier are combined via the Compound Relator function that takes the set denoted by the Clf/class term and returns a new selective subset. It is the entities in the denotation of this new selective subset that are being counted in numeral phrases.

The remainder of this paper is organized as follows: I provide more background on class terms and class term compounds in §2. Specifically, I demonstrate that these types of compounds have mixed properties of both bare nouns and classifier phrases, with the class term head apparently taking on classifier properties. In §3, I argue that these mixed properties are in fact not surprising if class terms are an intermediary stage of noun to classifier grammaticalization. DeLancey (1986) proposes such a grammaticalization trajectory for Thai, and I provide evidence showing that this trajectory extends to Vietnamese as well. Given that classifiers and class terms are semantically connected via grammaticalization, I provide an analysis in §4 that unifies their semantics: crucially, I show that unlike previous analyses of classifiers, the entities being counted are in the denotation of the ClfP or class term compound as a whole, rather than simply their second elements. This is done via the Compound Relator function, which takes the head and modifier as inputs and returns a new set that is a selective subset of the head. I summarize in §5.

2 Class terms

Vietnamese is traditionally considered to be a highly isolating language with little to no morphology other than compounding, which is a highly productive process in the language. One major source of compounding involves the usage of class terms. Class terms are nouns used as (semantic) heads in compounding, where they are phonologically dependent and lexically obligatory (Enfield, 2004). Thompson (1965) calls these constructions in Vietnamese pseudo-compounds, defining them as "morpheme sequences with two immediate constituents at least one of which is bound" (p. 133). Consider the following examples:

(6) xe, ‘vehicle/car’
   a. **xe đạp**, ‘bicycle’ (lit. ‘vehicle step’)
   b. **xe điện**, ‘tram’ (lit. ‘vehicle electricity’)
   c. **xe lửa**, ‘train’ (lit. ‘vehicle fire’)
   d. **xe tăng**, ‘tank’ (lit. ‘vehicle tank’)

(7) máy, ‘machine’
   a. **máy bay**, ‘airplane’ (lit. ‘machine fly’)
   b. **máy lạnh**, ‘air conditioner’ (lit. ‘machine cold’)
   c. **máy giặt**, ‘washer’ (lit. ‘machine wash’)
   d. **máy chữ**, ‘typewriter’ (lit. ‘machine grapheme’)

One thing that becomes apparent is that the meaning of the class term compound, while being potentially idiomatic, is not entirely unpredictable. The entities that a class term compound denotes are always some subkind of whatever the class term denotes. Put another

way, the meaning of class terms have a taxonomic relation to the meaning of the compounds they derive: for example, Enfield (2004) shows the high level of productivity of class term compounding in Lao for referents in the natural domain, where one might intuitively expect taxonomic relationships.

Sometimes in Vietnamese this taxonomic relationship appears to be divided:

(8) bánh, 'cake'

a. **bánh chưng**, 'rice cake' (lit. 'cake stew')
b. **bánh bao**, 'steamed bun' (lit. 'cake bag')
c. **bánh mì**, 'bread' (lit. 'cake wheat')
d. **bánh phở**, 'rice noodles' (lit. 'cake phở')
e. **bánh xe**, 'wheel' (lit. 'cake vehicle')
f. **bánh răng**, 'cog' (lit. 'cake teeth')

In the examples in (8), there is some relevant relation between each compound to the meaning 'cake', whether it has something to do with being food of some (semi-)regular consistency, such as *bánh mì*, 'bread', or being a round object, such as *bánh răng*, 'cog'. However, it seems that the relevant relation with the class term is inconsistent, as bread is not necessarily round and cogs are not food. These relations are discussed further in §4.2.

Class term compounds often have mixed properties of both bare nouns and classifier phrases (ClfPs). Kirby (2006) shows that bare classifier phrases have restricted interpretations compared to bare NPs, as can be seen in the reproduced table below:

(9)

	Bare NP	[Clf+N] (ClfP)	[Num+Clf+N] (NumP)
Definiteness	Indef/Def/Gen	Indef/Def	Indef
Number	Sg/Pl	Sg	–

The table in (9) shows that bare NPs allow the largest range of possible interpretations: they can give rise to indefinite, definite and generic/kind readings, as well as being underspecified for number. On the other hand, bare ClfPs cannot give rise to generic/kind level readings, while also being restricted in meaning to singular entities.

This correlation between syntactic structure and possible semantic interpretation provides some diagnostics for determining whether a class term compound is a bare NP or a ClfP. In addition to these diagnostics, I also consider data dealing with coordination and NP-Ellipsis to determine the categorical status of class term compounds; in the end, they demonstrate mixed properties of both bare NPs and ClfPs, as summed in the table below:

(10) Mixed-properties of class term compounds

NP-like	ClfP-like
generic interpretation	(directly countable)
underspecified for number	(NP-Ellipsis)
no coordination of complements	

Class term compounds always have the NP-like properties: allowing generic/kind level interpretations and being underspecified for number. Additionally, class terms do not allow coordination of their second element, which is not a property of NPs per se, but is one that is expected if they were truly clasifiers; because it provides evidence that class term compounds are not ClfPs, I put this property of non-coordination within the list of NP-like properties.

The ClfP properties include direct countability, which I take to mean appearing adjacent to an overt numeral with no intervening element, and NP-Ellipsis. Class term compounds occassionally have these properties, which is the central puzzle of this paper, though not all of them do. I argue that this is because class term compounds are actually all underlyingly bare NPs, with the classifier-like properties of class terms are the result of grammaticalization.

The following subsections provide explicit data for the mixed properties of class term compounds summarized in (10).

2.1 Noun-like behavior

Kirby's table in (9) shows that bare nouns have a wider range of possible semantic interpretations than bare ClfPs: specifically, they allow generic readings and are ambiguous between singular and plural. Consider the following example, which shows both of these possibilities:

(11) Chó ăn thìt.
dog eat meat
'(The) Dogs eat meat.'
or 'A/The dog eats meat.'

(12) Con chó ăn thìt.
CLF:ANI dog eat meat
'A/The dog eats meat.'

In (11), *chó* can have a reading meaning that dogs in general eat meat. It can also be interpreted as a plural noun, meaning that multiple dogs eat meat. Bare ClfPs, however, do not allow these readings, as (12) demonstrates. When the classifier is present, the denotation of the ClfP is obligatorily singular, and blocks the generic interpretation.

Class term compounds, like bare NPs, allow generic and plural interpretations:

(13) **Máy bay** đi lẹ lắm.
machine fly go quick very
'Airplanes go really quickly.'
or 'The airplanes go quickly.'

As can be seen above, *máy bay* can be interpreted generically or as a plural; as both of these interpretations are impossible for a bare ClfP (12), (13) provides good evidence that class term compounds are not actually ClfPs, and are more likely to be NPs.

Additionally, it is possible for nouns that share the same classifier to be coordinated underneath that classifier (in bold below):

(14) 100 **trái** chuối với cam
100 CLF:FRUIT banana and orange
'100 [oranges and bananas]'

(15) *100 **trái** chuối với chó
100 CLF:FRUIT banana and dog
intended: '100 [bananas and dogs]'

In (14), both bananas and dogs take the classifier *trái* for fruit, and can thus be conjoined underneath it[2]. (15) shows that two nouns that require different classifiers, such as *chuối*, 'banana', and *chó*, 'dog', cannot be conjoined underneath a shared classifier such as *trái*, 'fruit'.

If class terms were actually just full blown classifiers within compounds, then we would expect that their complements should be able to be coordinated. This is not the case:

[2]While it is interesting to note that the resulting interpretation can be either distributive or collective, I leave this fact to be addressed elsewhere

(16) *100 **máy** **bay** với **giặt**
100 machine fly and wash
'intended: 100 [airplanes and washing machines]'

Recall that *máy bay* means 'airplane' and *máy giặt* means 'washing machine'; they both share the class term *máy*, 'machine'. The fact that *bay* and *giặt* cannot be conjoined underneath *máy* provides good evidence that whatever the structure of class term compounds is, it is not a ClfP construction.

The evidence for class term compounds being bare nouns is strong then, as all of them have the noun-like properties mentioned. However, as we have briefly seen, some class terms also have classifier-like behavior.

2.2 Classifier-like behavior

As shown above, a number of class term compounds can be directly counted, which is unexpected for bare nouns, which generally always require classifiers.

(17) mười *(**trái/quả**) cam
ten CLF:FRUIT orange
'ten oranges'

(18) mười **xe đạp**
ten car step
'ten bicycles'

Typical nouns, such as *cam*, 'orange', require classifiers in counting contexts (17). Note here that *trái* and *quả* are synonymous classifiers for fruit, with *trái* being more associated with the Southern Vietnamese dialect and *quả* with the Northern Vietnamese dialect. The class term compound *xe đạp*, 'bicycle', however, can appear without any overt classifier between it and the numeral.

In §2.1 I provide evidence that the class term compound cannot be analyzed as a typical ClfP. There is additional evidence for this, even in the counting context, where class term compounds like *xe đạp* can optionally take classifiers:

(19) mười (chiếc) **xe đạp**
ten CLF:VEHICLE car step
'ten bicycles'

(20) *mười trái quả cam
ten CLF:FRUIT CLF:FRUIT orange
'ten oranges'

As (19) shows, the classifier *chiếc*, which generally means 'piece' but is used specifically for vehicles as well, can be used to count *xe đạp*, showing up in the canonical classifier position between the numeral and the NP. This is in fact the most conservative strategy for counting *xe đạp*, grammatical for all Vietnamese speakers. Compare this to the example in (20), where both fruit classifiers *trái* and *quả* are being used to count the noun *cam*. Classifier recursion, or the ability for a noun to take more than one classifier in a single counting construction, has been argued to be impossible (Jiang, 2012). This is clearly seen in (20), and implies that what is being counted in (19) is not a typical ClfP.

Another property of classifiers is that they license NP-Ellipsis (NPE), wherein ellipsis of the NP complement of a classifier is licensed given a contextual antecedent (Nguyễn, 2004; Alexiadou and Gengel, 2011; Cheng and Sybesma, 2009; Jenks, 2011). Consider the following example, where the classifier is in boldface:

(21) Nếu mày ăn ba **trái** cam, tôi sẽ ăn ba **trái** <*cam*> luôn.
if 2.SG eat three CLF:FRUIT orange 1.SG FUT eat three CLF:FRUIT also
'If you eat three oranges, I will eat three ones as well.'

In (21) above, the entire ClfP [trái cam] is mentioned in the first clause, which licenses ellipsis of the noun *cam* in the second clause. Similarly, certain class terms also seem to license this ellipsis – though it should be noted that unlike ClfPs, the elided element of a class term compound is not necessarily an NP, though I will continue to call it NPE for convenience:

(22) Nếu mày mua ba **máy** **lạnh**, tôi sẽ mua ba **máy** <*lạnh*> luôn.
if 2.SG buy three machine cold 1.SG FUT buy three machine also
'If you buy three air conditioners, I will buy three ones as well.'

In (22), we can see that the class term *máy* licenses NPE: *lạnh* is elided in the second clause, similar to *cam* in (21). However, an important fact to note is that this NPE is not a property of the class term itself, but seems to be a property of the class term within its specific compound context. In other words, whereas *máy* licenses NPE in the compound *máy lạnh* in (22), it does not necessary license it in other *máy* compounds:

(23) *Nếu mày mua ba **máy** **bay**, tôi sẽ mua ba **máy** <*bay*> luôn.
if 2.SG buy three machine fly 1.SG FUT buy three machine also
'If you buy three airplanes, I will buy three ones as well.'

In (23), we can see that the compound *máy bay*, which is formed from the class term *máy* like *máy lạnh*, does not license NPE. However, notice that both compounds are still directly countable (though they can take an optional classifier, such as the generic *cái* for inanimate nouns):

(24) a. mười (cái) **máy** **lạnh**
ten CLF machine cold
'ten air conditioners'

b. mười (cái) **máy** **bay**
ten CLF machine fly
'ten airplanes'

This suggests that the function of classifiers that allow counting is independent from the mechanisms that license NPE, though I do not have more to say on this topic in this paper.

To summarize, class term compounds prove to be problematic for the view that Vietnamese has obligatory classifiers, as they are often directly countable without an overt classifier. Instead, it appears as if the class term head of the compound takes on classifier functions, even licensing other syntactic phenomena such as NPE. However, while CT compounds always have the relevant properties of bare nouns, they vary in the degree to which they behave like ClfPs; in the following section, I show that this is due to the fact that class terms are nouns utilized to derive compounds, which can ultimately lead to their grammaticalization as classifiers.

3 Grammaticalization

Aikhenvald (2000) states that numeral classifiers most often emerge from nouns, and DeLancey (1986) shows more specifically that Thai classifiers emerge from class term compounds. As can be seen in (25) below, the word order of the Thai NP, [N Num Clf], as well as the prominence of the language family's repeater constructions – where a noun can be doubled as its own classifier (repeated form in bold) – provides a pathway for a noun to end up as a classifier in the language.

(25) a. takrâa săam **takrâa**
basket three CLF
'three baskets'

b. ráan-ʔaahăan săam **ráan**
shop-food three CLF
'three restaurants'
(Post, 2007, Thai)

These examples clearly demonstrate DeLancey's claim that Thai classifiers arise from repeated (head) nouns in counting constructions.

While Thai classifiers emerge from its repeater classifer constructions, Vietnamese does not appear to have these constructions, and classifiers emerge from a reanalysis of the surface word order of ClfPs and class term compounds, as well as the influence of prosodic preferences. By stipulating that this grammaticalization cline exists for Vietnamese, we then expect class terms to variably have both noun-like and classifier-like properties as they fall along the cline in different places.

The process of a classifier grammaticalizing from a class term in Vietnamese is due to reanalysis of the surface word order from [Num CT X] to [Num Clf X], where X stands for the non-categorically specified second element of a compound. Recall that Vietnamese is generally head-initial:

(26) a. **cây chuối**
tree banana
'banana tree'

b. cuốn sách
CLF:VOLUME book
'a book'

In each of the examples in (26), the first element of the constituent is the head, and is followed by some modifying or specifying element. In (26a), *cây* means 'tree' while *chuối* specifies that meaning to 'banana tree'; while in (26b) *cuốn* means 'volume' while *sách* specifies that meaning to 'book volume'. The latter case has additional meaning in its functional semantics pertaining to singularity, as we have seen, though this does not bear on the subsective relationship of books to volumes.

The case of *cây chuối* is similar in that *cây* denotes a superset of trees, which is then specified to the subset of banana trees within that. However, a slight difference is that unlike the ClfP situation, where the second element is a subset of the first element, bananas are not a subset of trees. Rather the meaning of the whole compound, 'banana tree', is a subset of 'tree'. If a ClfP has the structure [A [B]], then B $\subseteq$ A; if a class term compound has the structure [A B], then we instead have the case that AB $\subseteq$ A, where AB is the denotation of the entire compound, which can have varying degrees of idiosyncratic meaning.

Looking at the case of loanwords shows that in some cases, the subsective relationship in class term compounds can be identical to that of ClfPs. The word *tăng* is a loanword from *tank*, and thus enters the language with its fully idiomatic meaning 'tank'. However, it is lexicalized as the compound *xe tăng*, where the class term *xe* means something like 'wheeled land vehicle'. Because compounding in the case of *xe tăng* is no longer creating any idiomatic meaning, there is now redundancy between the meaning of *xe* and that of *tăng*; in other words, *xe* does not appear to be contributing any extra meaning to the overall meaning of the compound, as *tăng* already means 'tank' by virtue of being a loanword. As a result, *xe tăng* now has the same B $\subseteq$ A relationship between its constituent elements as ClfPs have, resulting in reanalysis of *xe* as Clf.

An additional independent factor behind reanalysis of class terms as classifiers might also be a prosodic one. In her dissertation on child acquisition of Vietnamese classifiers, Tran (2011) noted that children presented with disyllabic nouns (including compounds,

loanwords and reduplicated nouns) often made errors in choosing the prescriptively appropriate classifier. She notes that they made four types of mistakes (pp. 357-358):

1. classifier omission
2. usage of the general classifier in place of the more specific target one
3. production of the target classifier with reduction of the disyllabic noun to a single-syllable [I assume this means omission of one of the noun's morpheme rather than any kind of phonological reduction]
4. alternative two- to three-syllable nouns which where either the general classifier was used or the classifier was omitted

Following a numeral, Tran observes that there is a preference for either a disyllabic noun sans classifier, or a classifier with a monosyllabic, reduced/truncated noun. For children at least, there is a strong prosodic effect in which there is the expectation of two syllables following a numeral. This is due to most Vietnamese words being monomorphemic and monosyllabic (not counting compounds), which makes most [Num Clf NP] constituents tri-syllabic.

Though lacking in experimental evidence, there are empirical data of a disyllabic preference for adult speakers of Vietnamese: Thompson (1965) mentions briefly how native Vietnamese speakers diplay a strong preference for disyllabic expressions in cases where a monosyllabic expression is grammatical. He identifies this preference primarily in two places: fragment answers and clause-final position. He also mentions that in at least some cases, focal particles are frequently followed by the relative particle *mà* purely for stylistic/emphatic effect. As he doesn't provide good examples to support this usage, though, I do not include those data here.

If it is indeed the case that classifiers emerge from class terms, which are in turn nouns modified by some other element, then we expect to see certain traces of this trajectory. While I have not conducted a wide-scale study, specific examples do appear to support this hypothesis. Consider the following:

(27) a. Noun: **cây** 'tree'
b. CT: **cây** chuối 'banana tree'
c. Clf: **cây** dù 'a/the umbrella'

As (27) shows, the morpheme *cây* appears to be playing at least three different (though not unrelated) roles. It has the base meaning of 'tree' when used as a bare noun, which is incorporated into the meaning of the compounds it is a CT head for, such as *cây chuối*, 'banana tree'. Additionally, it also functions as a classifier for long and skinny objects, such as umbrellas; we can tell in (27c) that *cây* is a true classifier as it is obligatorily singular.

If the usage of *cây* as a classifier derived from its usage as a CT in compounds, then these are the types of traces of this trajectory that we would expect to find. Note that the classifier's meaning also appears to be more functional in meaning (i.e. denoting singleton sets of atoms) while being semantically bleached in its lexical meaning; this is in line with grammaticalization in general, where semantic bleaching is correlated with a more functional meaning. In sum, though a more comprehensive diachronic study of Vietnamese CTs and classifiers must be done, there is compelling evidence to believe that the latter evolve from the former.

4 Unifying Clf and CT semantics

If new classifiers emerge from nouns functioning as class terms in compounding, then the functional/logical meaning of classifiers must be composed into the meaning of class terms,

following the theory of grammaticalization proposed by von Fintel (1995). If this is the case, then there should be a unified semantics of countability that can account for both classifiers and class terms.

Recall that while standard theories analyze the functional meaning of classifiers as making the individual atoms of the nominal predicate countable (either by modifying the noun's semantics or the numeral's), they always assume that these atoms are within the set denotation of the noun. The case of class term compounds shows that it cannot be the case that the atomic entities being counted are denoted purely by this second element (nouns or otherwise), but that the atoms are rather in the set denotation of the entire compound or classifier phrase itself. To return to an example such as *cây chuối*, 'banana tree' (lit. 'tree banana'), what is being counted are not banana atoms, but rather banana tree atoms. In §4.1, I propose the Compound Relator function as a small modification to standard analyses of classifiers that can make sure that the numeral phrase is counting the right type of entities.

In §4.2, I show that though the denotations of CT compounds are within the taxonomy defined by their respective CT – i.e. *cây chuối*, 'banana tree', is a subkind of *cây*, 'tree' – the modifying element of a ClfP/CT compound does not exactly map the denotation of the head to a subset of that set. This can be seen in compounds like *bánh xe*, 'wheel', composed of the head *bánh* roughly meaning 'cake' and the modifier *xe* meaning 'vehicle/car'; wheels are not strictly speaking a subset of cakes, though they share a property of being round. Modifiers in ClfP and CT compounds don't denote a subset of the head's exact denotation, then, but rather a selective subset defined by some subset of the salient properties of the head: i.e. in the case of *bánh xe*, wheels are not a subset of cakes, but of round things.

4.1 The Compound Relator function

The vast majority of previous approaches to classifiers treats classifiers as enabling reference to singleton sets of atoms within the denotation of the noun of a ClfP. Jenks (2011), for example, provides the following semantics of the Thai classifier, *lûuk*:

(28) $[\![\text{lûuk}_{Clf}]\!] = \lambda k \lambda n \lambda x \,.\, {}^{\cup}k(x) \wedge \mu_{AT}(x) = n$ if ${}^{\cup}k \in \lambda x \,.\, \text{ball-like}(x)$
(Jenks, 2011, p. 81)

This semantic meaning for Thai classifiers can be paraphrased as a function that takes a kind-denoting noun (Chierchia, 1998) k, a numeral n and an individual x, and returns true iff x is within the set denotation of the type-shifted kind (${}^{\cup}k$) and the cardinality of x is equal to n, with the presupposition that individuals in ${}^{\cup}k$ are ball-like – the function is undefined if this presupposition does not hold.

However, as the previous sections show, a semantic meaning of this type does not work for CT compounds. For one, it assumes that the Clf/CT first takes a kind-denoting noun as a complement. CT compounds are not as syntactically constrained as ClfP with respect to the category of the second element: *xe đạp*, 'bicycle' (lit. 'vehicle step'), where the second element is a(n eventive/action) verb; *máy lạnh*, 'air conditioner' (lit. 'machine cold'), where the second element is an adjective (or stative verb – the distinction is trivial here). If Clfs/CTs are to first compose with something of type *k*, i.e. a bare noun, then these CT compounds are problematic, as verbs and adjectives/stative verbs are not typically analyzed as denoting kinds; even if they did, it does not ensure that the kind of entities that the compound denotes is within the set denotation of the type-shifted verb/adjective. In other words, an air conditioner is not a machine that is cold, and a bicycle is not a vehicle that steps. This problem still remains even if we ignore the kind-denoting properties of bare nouns in Vietnamese and other (South)East Asian languages, and treat them as predicates denoting sets.

Secondly, as I have been claiming throughout this paper, the denotation in (28) is ultimately counting the wrong kinds of atomic entities when it comes to CT compounds. This is true even in cases where we avoid the aforementioned type problem: consider a CT compound such as *cây xoài*, 'mango tree', where the second element is a noun. We can see that giving *cây* the semantics above results in an undesireable prediction that what we are counting is mangos rather than mango trees:

(29) False prediction:
$[\![\text{cây xoài}]\!] = \lambda n \lambda x \, . \, {}^{\cup}\text{MANGO}(x) \wedge \mu_{AT}(x) = n$ if ${}^{\cup}\text{MANGO} \in \lambda x \, . \, \text{long-skinny}(x)$

Not only does this type of denotation count the wrong kind of atomic entities in CT compounds, but it also presents problems for the presuppositional (lexical) content of the CTs themselves. Under the denotation above, we would in fact have an undefined meaning for *cây xoài*, as mangos are not long and skinny, thus failing the presupposition. Analyses that treat the semantic properties that classifiers lexically specify as a part of the asserted truth-conditional meaning (Rothstein, 2010) do no better, and in fact are arguably worse, as they predict falsehood rather than being undefined.

Nomoto (2013) has yet another analysis following McCready (2009, 2012) where the lexical meaning of classifiers is neither truth-conditional nor presuppositional, but is rather captured best via conventional implicatures within the framework developed by Potts (2005):

(30) $[\![\text{Clf}]\!] = \lambda P \lambda x \, . \, P(x) \wedge [\neg \exists y \in P \, . \, y < x] \blacklozenge \lambda P \, . \, P \subseteq \text{CLASS}$

Here, the formalization is slightly different than Jenks', but the fundamental idea is the same: the Clf takes a predicate argument P and an individual argument x, and is true if x is in P and is atomic. The ClfP also simultaenously has the conventional implicature (indicated by the material following $\blacklozenge$ that the set P is a subset of a set with the properties specific to the classifier being used (with CLASS representing these properties). Once again, however, this denotation results in counting the wrong type of atomic entities for CT compounds (we would once again be counting mangos instead of mango trees).

In sum, previous analyses of classifiers cannot be extended to the case of CT compounds, as they all assume that what is counted is a nominal second element of a ClfP, rather than what is denoted by the entire ClfP/CT compound. For the purposes of this paper, I adopt and modify Nomoto's formalization of classifier semantics, though any other formalization must also take into account my argument that what is being counted in ClfPs and CT compounds is not simply atoms in the denotation of the second element.

In order to specify that what is being counted is the entire ClfP/CT compound, rather than just the second element, I propose the Compound Relator function that maps the set denoted by the head noun to another set specified by the modifier; this new set is not a strictly intersective/subsective set, but rather a selective subset, further detailed in §4.2. The Compound Relator (CR) function thus takes two arguments and returns a set that is related to both arguments; a subset relation satisfies this relation.

(31) Compound Relator function: $\text{CR}(\alpha, \beta)$
   a. if $\beta \subseteq \alpha$ then $\text{CR}(\alpha, \beta) = \beta$
   b. else $\text{CR}(\alpha, \beta) = \gamma$, where γ is related to both α and β in some way

The condition in (31b) that the new set γ is related to both α and β 'in some way', is left intentionally vague, as the exact semantic relation between compound elements is notoriously unpredictable: *firetruck* in English refers to a vehicle that is utilized in putting out

fires, while Vietnamese *xe lửa*, lit. 'vehicle fire', refers to a train, which at least historically used to run on an engine that was primarily fueled by fire. While I do not attempt to capture the nature(s) of relations between the two elements of a compound and its meaning as a whole, I also show in the next section that these relations are not completely arbitrary either: specifically, γ is a selective subset of α, though not necessarily a strict subset.

The condition in (31a) is crucial for the case of ClfPs, where the noun always denotes a subset of the classifier: *con* refers to animals while *chó* refers to dogs, which comprise a subset of animals. Thus CR(ANIMAL, DOG) refers to dogs ($\gamma = \beta$), because $\beta \subseteq \alpha$. In other words, when there is a subset relation between the two elements of a ClfP (or CT compound), then the set that is returned by the Compound Relator function is trivial, as it is equivalent to the more specific modifier set. Consequently, in the case of true ClfPs, the kind of atoms denoted by the entire ClfP is identical to the kind of atoms comprising the nominal complement; as such, it makes sense that previous approaches in the literature have simply assumed that the atoms being counted is solely dependent upon the noun.

Thus, we have the updated semantics for classifiers and CTs:

(32) $[\![\text{Clf/CT}]\!] = \lambda P \lambda x \,.\, \text{CR}(\text{CLASS}, P)(x) \wedge [\neg\exists y \in (\text{CR}(\text{CLASS}, P) \,.\, y < x]$
♦ $\lambda P \,.\, \text{CR}(\text{CLASS}, P) \subseteq \text{CLASS}$

All that has changed from Nomoto's original semantics here is that rather than looking at the set denoted by $P(x)$ to evaluate set membership and atomicity of x, we are now looking at the set denoted by CR(CLASS, P), where CLASS represents the set denoted by the specific lexical properties of the classifier: long and skinny things, fruit, animals, etc. For true Clfs, then, CR(CLASS, P)(x) trivially reduces to the same thing as P(x): counting dog-animals or mango-fruit is the same as counting dogs or mangos, respectively.

It is in the case of CT compounds where the CR function makes a difference, though. Recall that *cây xoài* denotes mango trees rather than mango fruit. Because MANGO $\not\subseteq$ TREE – that is, mangos are not a subset of trees – CR(TREE, MANGO) must return a new set that is related to both TREE and MANGO in some way. In this case particular case, this new predicate is the set of mango trees:

(33) [[cây xoài]]

a. $= \lambda x \,.\, \text{CR}(\text{TREE}, \text{MANGO})(x) \wedge [\neg\exists y \in \text{CR}(\text{TREE}, \text{MANGO}) \,.\, y < x]$
♦ $\lambda P \,.\, \text{CR}(\text{TREE}, \text{MANGO}) \subseteq \text{TREE}$

b. $= \lambda x \,.\, \text{MANGO-TREE}(x) \wedge [\neg\exists y \in \text{MANGO-TREE}(x) \,.\, y < x]$
♦ $\lambda P \,.\, \text{MANGO-TREE} \subseteq \text{TREE}$

The relation between mango trees and mangos and trees is relatively transparent, but this is not always the case; a compound can refer to something not entirely predictable in relation to its parts, as with the *firetruck – xe lửa* example earlier. It is in those cases where the output of the CR function is subject to the idiomatic interpretation of a CT head and its modifier.

(34) [[xe lửa]]

a. $= \lambda x \,.\, \text{CR}(\text{VEHICLE}, \text{FIRE})(x) \wedge [\neg\exists y \in \text{CR}(\text{VEHICLE}, \text{FIRE}) \,.\, y < x]$
♦ $\lambda P \,.\, \text{CR}(\text{VEHICLE}, \text{FIRE}) \subseteq \text{VEHICLE}$

b. $= \lambda x \,.\, \text{TRAIN}(x) \wedge [\neg\exists y \in \text{TRAIN}(x) \,.\, y < x]$
♦ $\lambda P \,.\, \text{TRAIN} \subseteq \text{VEHICLE}$

Having the CR function thus lets us unify the semantics for classifiers and class terms, which is motivated by the evolution of classifiers from nouns utilized as class terms in

compounds. While the CR function is essentially redundant in the case of true classifiers, it is necessary in order to return a new set in CT compounds where the two elements are not directly related to each other (i.e. subset relation). I take this redundancy to imply that when the two elements already have a direct relation to each other, there is no need to create a new set that relates the two arguments that the CR function takes; this relation already exists. This captures the intuition that there is little to no ambiguity in the denotation of a novel ClfP, but the denotation of novel compounds is not always predictable – the speaker/hearer must find a way to relate the two compound elements to get the denotation of the compound as a whole, though this relation is subject to variation, evident in English *firetruck* and Vietnamese *xe lửa*.

4.2 Deriving selective subsets with the Compound Relator function

This section provides a more in-depth look at how the Compound Relator function returns a new set that is related to both compound elements. Specifically, the relation between the set γ is not arbitrary with respect to the head CT α, but as I demonstrate below also not necessarily a strict subset/intersective set; rather if α is actually treated as the set defined by various semantic qualities (i.e. fruit, round, etc), then γ must be a subset of the set described by at least one of these qualities, but not necessarily all of them. In other words, the set γ returned by the CR function is a subset of at least one of the CT's supersets (that is a proper subset of the domain), but not necessarily all of them. I call this the selective subset relation:

(35) **Selective subset ($\subseteq_S$):** A $\subseteq_S$ B holds iff B is the intersection of a contextually-given set of sets Z, such that each set in Z is a superset of B and is a proper subset of the domain, and A $\subseteq$ Y where Y $\in$ Z

Note that each set in Z must be a proper subset of the domain, to prevent any set trivially being a selective subset of any other set by virtue of the domain being a superset by necessity.

Given the selective subset definition in (35), the result of CR(α, β) where $\beta \nsubseteq \alpha$ is the set γ, such that $\gamma \subseteq_S \alpha$. In other words, the set denoted by a CT compound is a subset of at least one superset of the class term that heads it, but isn't necessarily a subset and doesn't necessarily intersect with the set denoted by the class term's bare noun equivalent.

Let's consider a more concrete example: recall that CT compounds comprise a taxonomy of the CT that heads them. The example from (8) is reproduced below:

(36) bánh, 'cake'
  a. **bánh chưng**, 'rice cake' (lit. 'cake stew')
  b. **bánh bao**, 'steamed bun' (lit. 'cake bag')
  c. **bánh mì**, 'bread' (lit. 'cake wheat')
  d. **bánh phở**, 'rice noodles' (lit. 'cake phở')
  e. **bánh xe**, 'wheel' (lit. 'cake vehicle')
  f. **bánh răng**, 'cog' (lit. 'cake teeth')

The dictionary definition of *bánh* is 'cake', and indeed compounds derived from this morpheme's use as a class term, such as *bánh bao* seem cake-like enough in being round, edible things with some sort of internal consistency. However, these properties that are present in the meaning of *bánh* show up more irregularly as one works down the list of CT compounds in (36): *bánh phở*, 'rice noodles', are certainly edible and internally consistent, but are not

saliently round in any way.[3] Contrast this to *bánh xe*, which refers to wheels (usually in reference to a vehicle's wheel), which are round, but not edible – whether they are conceptualized as having regular internal consistency is arguable.

If *bánh* denotes the set of cakes, then it seems incorrect to consider rice noodles and wheels as proper subsets of cakes. However, at the same time, it is clear what the relation of these things to cakes is: rice noodles are edible and wheels are round, and both generally have some internal consistency. In other words, they are selective subsets of cakes: RICE-NOODLES $\subseteq_S$ CAKE; WHEEL $\subseteq_S$ CAKE.

We can see that it is not just CT compounds in Vietnamese that have this selective subset relation between the head and modifier, but also ClfPs. Consider the following examples of the classifier *trái*:

(37) a. *trái cam*, 'an/the orange'
b. *trái chuối*, 'a/the banana'
c. *trái banh*, 'a/the ball'

The classifier *trái* is ordinarily used for fruits, as can be seen in the case of oranges in (37a). It is also unsurprisingly used for bananas as well, as they are also fruit. However, what is interesting is that *trái* also serves as the classifier for *banh*, 'ball', which is not a fruit. The relation is clear, though: balls are spherical, as are many fruits, such as oranges. Crucially, bananas are non-spherical fruits and balls are spherical non-fruits, so there is no single unified semantic property set for *trái* that can capture both bananas and balls without ruling out the other. Rather, while the CR function returns the set denoted by the nominal complement in ClfPs, they are still subject to the selective subset relation.

A natural question that arises is given the set Z of supersets of B, why can't A $\subseteq_S$ B with respect to any Y $\in$ Z, no matter how obscure or irrelevant? Put another way, cakes are also often sweet, so there is a superset of cakes that is the set of sweet things. However, I am unaware of any compounds derived from the CT *bánh* that denote sweet things that are inedible, non-round, etc. (such as antifreeze, for example). Such a compound would be quite surprising, as sweetness is not a highly salient property of Vietnamese *bánh* – in fact, of the examples listed in (36), not one of the food items is sweet.

This is where the the condition that the set Z (set of supersets) is contextually given comes into play: only the most salient properties of a CT are generally at play when deriving selective subsets via the CR function. For a CT like *bánh*, it is salient that cakes in Vietnamese are generally round, edible and have internal consistency, but not that they are (sometimes) sweet, and it is these properties that will be targeted by a novel compound headed by *bánh*. So of all the countless supersets of B that can be in Z, we only consider the ones that are contextually provided, i.e. highly salient.

Given that the context still provides a number of supersets comprising Z, there is also some ordering of salience within Z. In the case of *bánh*, for example, the superset of edible things is probably more salient than the property of being round, as exemplified by the compounds denoting non-round foods: i.e. *bánh phở*, 'rice noodles'; *bánh chưng*, 'rice cake' (typically rectangular). Thus, there appears to be at least a partial ordering of salient, contextually-provided properties that comprise the CT/Clf; a higher degree of salience will correlate to that particular superset being targeted when CR derives a selective subset.

Nomoto (2013) provides an analysis of this ordering of importance in classifier properties in a Linear Optimality Theory (LOT) framework. Recall that CLASS in his denotation

[3] It should be noted that rice noodles of the sort described by *bánh phở* are generally sold in plastic packs where they are dried and wound into 'bricks'. This packaging no doubt has some impact on the choice of the CT *bánh*, but as mentioned, these packaged noodles are box-shaped, rather than being round

in (30) is actually shorthand for the various properties specified by a classifier that are a part of the classifier's conventionally implicated meaning. In his analysis, these properties are constraints that are weighted, with violations incurring penalties against the aggregated harmonic score H; consider the following tableau for the Japanese classifier *dai*, used for land vehicles (superficially comparable to Vietnamese *xe*):

(38)

dai kuruma/honbako/neko CLF car/bookcase/cat	INANIMATE 34.81	MECHANICAL 3.9	ON THE GROUND 0.5	DETACHED 1.3	CARRIES THINGS 0.2	*H*
car						0
??bookcase		*		*		-4.2
*cat	*	*			*	-38.91

(Nomoto, 2013, p. 85)

The harmonic score H is calculated according to the following formalization:

(39) $H_i = -\sum w_j v_j$

a. H_i is harmony score of candidate i

b. w_j = weight of constraint j

c. v = number of violations of j

(Nomoto, 2013, p. 52)

In (38), we can see that using *dai* with *kuruma*, 'car', incurs no violations, resulting in a perfect H score of 0; on the other hand, using *dai* with *neko*, 'cat', incurs more significant violations, with a low H score of -38.91. This is interpreted as *dai* being a perfectly grammatical classifier for cars, but unacceptable for cats; *honbako*, 'bookcase', has a score of -4.2, which is not perfect, but also not as bad as -38.91 for *neko*, thus resulting in questionable acceptability.

I assume that whatever weighting mechanism is at play for determining classifier-noun harmony in Nomoto's LOT formalization is the same mechanism at play for ranking saliency of the contextually provided (super)sets comprising Z in the CR function. The context provides only the salient properties (supersets) of a given CT/Clf, forming the set of sets Z; these sets in Z are then weighted for relative salience with respect to each other. In this way, when a new compound is formed via the CR function, it will usually be a subset of the most salient superset (property) of the head CT/Clf. This also ensures that the derived compound has a taxonomic relationship to the head that it is derived from.

In summary, what is being counted in the case of class term compounds and ClfPs is not the atoms denoted by the second element of each construction, but rather the atoms denoted by the entire constituent as a whole. In the case of ClfPs, this is a trivial difference, because the Compound Relator function will return the same set denoted by the Clf's nominal complement; however, it is in CT compounds where we see idiomatic meaning creep back in, making it clear that we must consider the meaning of the compound as a whole in order to count the right type of things.

The CR relation derives and returns a new set that is related to both of the elements that comprise the compound, and while this set need not be strictly intersective with the set denoted by the head, we see that it is also not a completely arbitrary set: CR returns a selective subset of the head. The selection of which head property to be a selective subset with respect to is further constrained to the contextually provided head supersets – in other words, only the salient properties of the head matter. These contextually provided supersets are also then weighted such that the set returned by the CR function will tend towards being a subset of the most salient contextually provided superset, as this will result in the best harmonic score. In plain words, a novel compound will often incorporate the meaning of the head's most prominent quality/s.

5 Conclusions

In this paper, I show that previous analyses of classifiers are insufficient to handle CT compounds in Vietnamese. These CT compounds are often directly countable without classifiers, which is problematic at first glance for Vietnamese, which is ordinarily considered an obligatory classifier language – and indeed, many nouns in the language require classifiers in order to be counted, making it different from languages where classifiers appear to be optional with all nouns (c.f. Indonesian (Chung, 2000); Malaysian (Nomoto, 2013)).

Furthermore, typologically distinct classifier languages also demonstrate similar problems in counting the wrong sorts of things in compound (or compound-like) constructions. Consider the following data from Chol (Mayan):

(40) a. cha'-ts'ijty ja'as
two-CLF:LONG-SKINNY banana
'two bananas'

b. cha'-pajl ja'as
two-CLF:BUNCH banana
'two bunches of bananas'

c. cha'-tyek ja'as
two-CLF:TREE banana
'two banana trees'

(Chol, (Bale and Coon, to appear, p. 8))

Though they relocate the heavy lifting of individuation and atomicity away from the classifier and onto the numeral, Bale and Coon (to appear) provide the following semantic meanings for Chol classifiers and numerals (specifically *ux*, '3', here), which are still problematic with respect to what is being counted:

(41) a. Denotation of a numeral that requires a classifier:
$[\![\text{ux}]\!] = \lambda m \lambda P : \text{ATOMIC}(P) . \{x : {}^*P(x) \wedge m(x) = 3\}$

b. Denotation of the classifier:
$[\![\text{p'ej}]\!] = \mu_{\#}$ (Bale and Coon, to appear, p. 7)

In the semantics above, the numeral takes a classifier, which provides a cardinality measure function $\mu_{\#}$, and a nominal predicate P as arguments; as long as P is atomic, then what is returned is the set such that each sum, x, of entities that can be formed within that set is in P, and the cardinality of x (presumably with respect to the classifier's specific properties) is equal to some number – 3 in the case of *ux*.

Despite the difference in formalization, what is crucial to notice about Bale and Coon's proposal for Chol classifiers is that it suffers from the same problem of paying too much attention to the entities denoted by the nominal predicate and their atomicity. Consider (40c), where what is being counted is banana trees. According to the denotations in (41), we check for atomicity in the set of bananas, and then look at sums of individuals that are both bananas and tree-cardinalities of 3, which will likely yield an empty set, as nothing is going to be simultaneously a banana and a tree. The problem is that by having only P show up in the denotation, we are caring too much about the entities denoted by the noun, instead of what the noun and Clf together denote. In short, Chol provides another example where the standard analysis of classifiers (and numerals) leads to the prediction that we are counting bananas, and not banana trees.

These data are less problematic if we view CTs as part of the grammaticalization path from nouns to classifiers, as proposed by DeLancey (1986) for Thai, and appears to also be the case for Vietnamese; doing so, however, motivates a unified semantic account of both classifiers and class terms. To do this, I have outlined a proposal in which it is always the entities denoted by the entire compound (used loosely to include both CT compounds and ClfPs), rather than just entities denoted by the second element. This set of entities is returned by the Compound Relator function, which either returns the same set denoted by the nominal complement in ClfPs (as they satisfy a subset relation already), or returns a selective subset of the head CT/Clf.

While the semantic relations between the elements comprising compounds is as notoriously multi-headed as the mythical Hydra, they are not entirely without order. The selective subset relation seen in the CR function shows that the taxonomic relationship between compounds and their heads is systematic in a way that the grammar on a larger scale is sensitive to (i.e. direct countability). While semantic analyses of classifiers have generally focused largely on the logical, individuation-oriented meaning of classifiers, I hope to show in this paper that looking at the non-logical, lexically specific meaning of classifiers and class terms can give us a fuller, more nuanced view of the classifier landscape.

Acknowledgements: I would like to especially thank Itamar Francez and Karlos Arregi for their invaluable help with this paper. I would also like to thank my primary consultant Huonglien Pham, as well as Chieu Nguyen, Karyn Mai. All mistakes are my own.

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford studies in typology and linguistic theory. New York: Oxford University Press Inc.

Alexiadou, Artemis, and Kirsten Gengel. 2011. Classifiers as morphosyntactic licensors of NP ellipsis: English vs. Romance. In *Proceedings of NELS 39*, 15–28. Manuscript. Universität Stuttgart.

Bale, Alan, and Jessica Coon. to appear. Classifiers are for numerals, not for nouns: Consequences for the mass-count distinction. *Linguistic Inquiry* .

Cheng, Lisa Lai-Shen, and Rint Sybesma. 2009. De as an underspecified classifier: first explorations. *Yǔyánxué lùncóng* 39.

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

Chung, Sandra. 2000. On reference to kinds in Indonesian. *Natural Language Semantics* 8:151–171.

DeLancey, Scott. 1986. Toward a history of Tai classifier systems. In *Noun Classes and Categorization*, ed. Colette Craig, volume 7 of *Typological studies in language*, 437 – 452. Amsterdam/Philadelphia: John Benjamins B.V.

Denny, J. P. 1976. What are noun classifiers good for? In *Proceedings from the twelfth regional meeting of the Chicago Linguistics Society*, 122–123.

Enfield, Nick J. 2004. Nominal classification in Lao: a sketch. *Sprachtypologie und Universalienforschung* 57:117–143.

von Fintel, Kai. 1995. The formal semantics of grammaticalization. In *NELS 25 Proceedings*, volume 2: Papers from the Workshops on Language Acquisition and Change, 175–189. GLSA, UMass Amherst.

Jenks, Peter Spiros Eric. 2011. The hidden structure of Thai noun phrases. Doctoral Dissertation, Harvard University, Cambridge, Massachussetts.

Jiang, Li. 2012. Nominal arguments and language variation. Doctoral Dissertation, Harvard University, Cambridge, Massachussetts.

Kirby, James. 2006. Vietnamese and the structure of NP. U. Chicago manuscript.

Krifka, Manfred. 1995. Common nouns: A contrastive analysis of English and Chinese. In *The generic book*, ed. G. N. Carlson and F. J. Pelletier, 398–411. Chicago, IL: The University of Chicago Press.

McCready, Eric. 2009. Conventional implicature and classifiers. *Journal of Cognitive Science* 10:195–208.

McCready, Eric. 2012. Classification without assertion. In *Proceedings of the Fifth Formal Approaches to Japanese Linguistics Conference*, ed. Matthew A. Tucker, Anie Thompson, Oliver Northrup, and Ryan Bennett, volume 64 of *MIT Working Papers in Linguistics*, 141–154.

Nguyễn, Tuong Hung. 2004. The structure of the Vietnamese noun phrase. Doctoral Dissertation, Boston University.

Nomoto, Hiroki. 2013. Number in classifier languages. Doctoral Dissertation, University of Minnesota.

Post, Mark W. 2007. Grammaticalization and compounding in Thai and Chinese. *Studies in Language* 31:117–175.

Potts, Chris. 2005. *The logic of conventional implicatures*. New York: Oxford University Press.

Roberts, Ian. 2010. Grammaticalization, the clausal hierarchy and semantic bleaching. In *Gradience, gradualness and grammaticalization*, ed. Elizabeth Closs Traugott and Graeme Trousdale, 45–73. Amsterdam, Netherlands: Benjamins.

Rothstein, Susan. 2010. Counting and the mass/count distinction. *Journal of Semantics* 27:343–397.

Thompson, Laurence C. 1965. *A Vietnamese grammar*. Number XIII-XIV in Mon-Khmer Studies. University of Hawai'i Press.

Tran, Jennie. 2011. The acquisition of Vietnamese classifiers. Doctoral Dissertation, University of Hawai'i at Manoa.