

Classifiers and DP Structure in Southeast Asia

Andrew Simpson

The Oxford Handbook of Comparative Syntax

Edited by Guglielmo Cinque and Richard S. Kayne

Print Publication Date: Oct 2008

Subject: Linguistics, Morphology and Syntax, Languages by Region

Online Publication Date: Sep 2012 DOI: 10.1093/oxfordhb/9780195136517.013.0018

Abstract and Keywords

This article examines the variations in determiner phrase (DP) and classifiers in Southeast Asian languages. It describes the internal structure of DP and addresses the specific problem of how to account for the considerable amount of cross-linguistic variation that appears to occur in the ordering of constituents in DP. The article suggests that the significant distortion of underlying DP structure is often caused by XP-movement inside the DP, and that variation can essentially be attributed to elements being in different stages of ongoing historical development and reanalysis.

Keywords: Southeast Asian languages, determiner phrase, classifiers, cross-linguistic variation, XP-movement

1 Introduction

SOUTHEAST Asia is a geographical area that is extremely rich from a linguistic point of view, being a Balkan-like region where a wide range of language families meet and interact with each other. This chapter's particular focus of interest is the internal structure of DPs and the specific problem of how to account for the considerable amount of cross-linguistic variation that appears to occur in the ordering of constituents in DPs. Some of the patterns found are schematized in (1)–(3) (RC = relative clause, CL = classifier):

- (1) Thai, Khmer: N Adj RC Num CL Dem
- (2) Burmese: Dem RC N Adj Num CL
- (3) Hmong, Malay, Vietnamese: Num CL N Adj RC Dem

Of considerable interest here is the fact that the variation attested often does not seem to follow or correspond to the apparent headedness of the relevant [\(p. 807\)](#) languages. Thus, for example, Thai and Khmer are both canonical svo head-initial languages, yet at first sight they seem to be head-final in their DPs with elements such as Num, CL, and Dem all following the head noun N (and in the postnominal placement of Num and CL, Thai and Khmer pattern with the canonical SOV language Burmese rather than the svo languages

listed in (3)). Other aspects of (1)–(3) can be similarly argued to be unexpected given the assumed general headedness of the languages in question. In this chapter, therefore, I set out to examine what factors might be responsible for the diversity attested, asking whether there are, indeed, any significant principles regulating the internal structure of DPs in Southeast Asian languages, or whether one has to concede that the patterns are really random and unconstrained.

2 The Status of Classifiers

Considering the syntax of DPs in Southeast Asian languages, a first important issue that needs to be examined and clarified is the syntactic status of the classifier elements that occur in DPs throughout the languages of the region. In the literature concerned with DP syntax in other classifier languages such as Japanese and Chinese, there are actually two quite different assumptions about classifiers and the relation they have to numerals. On the one hand, a number of works consider numbers and classifiers to instantiate distinct functional head positions, Num and CL (e.g., Pan 1990, Tang 1990). In other works, however, numbers and classifiers are treated as constituting a single functional head labeled simply CL, Num, or Q (e.g., Kawashima 1993, Muromatsu 1998). Somewhat surprisingly, there is often little explicit argumentation justifying one of the possible analyses over the other, and either one analysis or the other is frequently simply assumed without further discussion. Whether numbers and classifiers instantiate a single head or distinct heads is an important question with significant consequences, however, and it is therefore important to consider what arguments there are in favor of either of the two possible analyses.

Gil (1994) identifies the following observations as potential support for the view that numerals and classifiers comprise a single syntactic unit. First, as observed in Greenberg (1975), numbers and classifiers commonly occur together as a single *uninterrupted* sequence. Second, numbers and classifiers in many languages pattern phonologically as a single unit, suggesting that the classifier might perhaps be a suffix attached to the numeral. Third, in certain languages, the number+classifier sequence can appear separated and “floated” away from the rest of an NP, indicating (possibly) a particularly close linking of the number and (p. 808) classifier as a single unit. Such potential arguments in favor of a single-head analysis, however, are perhaps not particularly strong. The observation that numbers and classifiers are commonly adjacent and uninterrupted may just as easily be explained by the assumption that numbers and classifiers are perhaps in adjacent functional heads, Num⁰ selecting a complement CLP, and does not force one to assume that classifiers must necessarily be suffixes on numbers. The phonological dependence of classifiers on numbers (where attested) may possibly be attributed to classifiers coming to be enclitics as they grammaticalize, as indeed noted in Gil (1994), and again does not rule out the possibility that classifiers might encliticize from a discrete head position. Finally, the phenomena of numeral and classifier “floating” can, in fact, also be given plausible accounts under a two-head alternative and so do not obviously favor a single-head analysis.

The arguments supporting a two-head hypothesis are considerably stronger. First, there is the simple observation that two distinct morphemes occur in numeral-classifier sequences, which might naturally seem to suggest that two distinct head positions are projected. Classifiers in the many languages of Southeast Asia are also by and large phonologically quite unreduced and so appear to be fully independent functional words rather than inflectional affixes.

Second, classifiers are *functional* elements, argued by Muromatsu (1998), Cheng and Sybesma (1999), and others to have the primary semantic function of *individuating* NPs. Importantly, the two functions of individuation (provided by the classifier) and number specification (provided by numerals) are semantically distinct, and the use of a numeral with an NP does not, in fact, imply that the NP necessarily has to be conceived of as a set of discrete individuated entities. Instead, a numerated NP can be conceived of as a non-individuated group whose total is simply specified numerically. This assumed distinction between the functions of numerals and classifiers has observable consequences in certain classifier languages, and one finds clear evidence of two types in favor of the separation of numerals and classifiers into two formally independent heads. In languages such as Vietnamese, Hmong, and Nung, one finds that a classifier can occur alone *without any numeral*, simply functioning to individuate an NP, as in (4):

(4)

Classifiers and DP Structure in Southeast Asia

Tus	tsov	tshaib	tshaib	plab.
CL	tiger	hungry	hungry	stomach
'The tiger is/was very hungry.' (Hmong; Jaisser 1987)				

The converse situation, that numerals may sometimes occur without any accompanying classifier, typically is found when the numeral specification is rather vague and individuation is not necessarily implied. Thus, in Nung and Burmese, classifiers are optional with numbers that are multiples of ten; in Jingpo, classifiers are often omitted with numbers over ten; and “in Thai classifiers do not occur (p. 809) with large numbers like 1000 unless individuation is implied” (Aikhenvald 2000: 100). Hopper (1986) also points out that in Malay classifiers are omitted with numerals just when approximate and vague numeral reference is made and there is no specific individuation, as in (5):

(5)

Classifiers and DP Structure in Southeast Asia

Adalah	dua	tiga	pondok	kecil-	kecil	bersam a-sama	dekat	rumah	Temeng gong.
Be	2	3	hut	small	small	togeth- er	near	house	Temeng gong
'There were two or three small huts close together near Temenggong's house.'									

Classifiers and DP Structure in Southeast Asia

Finally, Bisang (1999) notes that in Vietnamese the classifier may similarly be omitted when a counted noun is not individualized, as in (6) and various other examples from Label (1996) (in (6) no classifier occurs individualizing *phong* 'room'):

(6)

nha	ba	phong
House	3	room
'a three-room house'		

Such patterns are good indications that classifiers and numerals perform distinct formal functions (numerical specification and individuation) and so should be assumed to occur in separate syntactic heads.

A third argument in favor of the two-head hypothesis comes from the observation that in Nung, a northern Tai language, the number 'one' does not occur adjacent to the classifier at all but is actually separated from the classifier by the noun:

(7)

Classifiers and DP Structure in Southeast Asia

An	ahn	tahng	nuhng	ma.
take	CL	chair	one	come
'Bring a chair.' (Saul and Wilson 1980: 56)				

Classifiers and DP Structure in Southeast Asia

A similar pattern is reported in Ejagham (Benue-Congo), and all numbers may be found nonadjacent to the classifier (NC = noun class marker):

(8)

Classifiers and DP Structure in Southeast Asia

a-mege	'	i-cokud		a-bae.
NC-CL	GEN	NC-orange	seed	NC-two
'two orange seeds' (Waiters 1981: 310)				

(p. 810) If numerical specification by any number is possible in a position distinct from the classifier position, this indicates again that the functions of counting and individuation can be assumed to relate to distinct functional head positions.

There are also languages that allow for a limited range of adjectives to be inserted between numerals and the classifier position, as in Chinese (9), this again indicating that the numeral and the classifier do not occur in a single functional head position:

(9)

a.

yi	xiao	ben	shu
one	small	CL	book
'One small book'			

b.

liu	da	jian	xingli
6	big	CL	luggage
'six big pieces of luggage' (T'ung and Pollard 1982)			

Overall, then, the evidence suggests that numerals and classifiers do not occur together in a single-head position and that there are, instead, two distinct positions projected by numbers and classifiers, Num and CL, each associated with a distinct semantic function.

3 Headedness and Directionality in the DP

The conclusion that Num° and CL° occur as the heads of distinct functional projections can now be shown to have important consequences for the analysis of DP-structure in many of the Southeast Asian classifier languages. Reconsider the ordering of elements in the Thai (and Khmer) DP (10) and compare this with Chinese (11):

(10) Thai: [_{DP} N Adj Num CL Dem]

(11) Chinese: [_{DP} Dem Num CL Adj N]

If one were to assume, contrary to the conclusions of section 2, that Num and CL combine to form a *single* functional head, then (10) and (11) would actually be mirror images of each other. The two orders could then possibly be accounted (p. 811) for by suggesting that (11) is a head-initial DP with Dem/D⁰ and Num-CL⁰ selecting complements to their right, and that the mirror-image pattern in (10) is simply a head-final DP ordered in the opposite way (though such a conclusion might be surprising, as Thai is elsewhere regularly head-initial). Significantly, once the single-head analysis of Num and CL is rejected (for the reasons given) in favor of the assumption that both Num and CL instantiate discrete functional heads, such an analysis of Chinese and

Thai is no longer possible. Consider the Thai sequence in (10) once more. If the ordering in (10) is a result of simple base-generation of a head-final DP and if Num and CL are discrete heads, then it has to be assumed that CL° is located above Num° . If this is so, then it becomes impossible to assume that the head-initial Chinese order in (11) is base-generated, because instead of (11) one would expect the sequence [Dem CL Num Adj N] with CL° selecting $NumP$ to its right. In fact, in all works distinguishing Num and CL as distinct heads, it is commonly assumed that Num/numerals and other quantifiers take scope over CL/ classifiers, this reflecting the assumption that nouns may first be individuated by a classifier and then quantified over by a numeral or other quantifier (see Cheng and Sybesma 1999). If Num and CL project separate heads, such a scope relation therefore suggests that Num should be the higher of the two heads. Reconsidering the Thai order in (10) now, this results in the important conclusion that such a sequence cannot, in fact, be simply base-generated as it appears; in a head-final structure, one would (now) clearly expect the ordering to instead be [N CL Num Dem], with Num selecting CLP to its left.

If the surface linear sequence in Thai (10) is not simply base-generated, it has to be assumed that it results from certain movement. Because the ordering of the elements Num and CL with respect to N seems to be the problematic part of structure in need of explanation, consider now how the order [N Num CL] might be created. Essentially, there are three possibilities. If one assumes that the Thai DP is underlyingly head-final (despite this going against the headedness of Thai elsewhere), one would have to conclude that [N Num CL] sequences arise either via movement of the CL to a higher rightward head position above Num—namely, $[N t_i \text{ Num } CL_i]$ —or that [N Num CL] results from lowering of the Num to a position below CL; namely, $[N \text{ Num}_i \text{ CL } t_i]$. Neither of these possibilities seems plausible, however. The first would be expected to be blocked by the Head Movement Constraint (HMC) and the second barred by general restrictions on lowering. Consequently, it seems that one is forced to assume that Num and CL do not change positions, that the DP is therefore underlyingly head-initial, and that it is the N element that undergoes movement, raising leftward from a position base-generated as the rightward complement of CL. Because this raising might be expected to be blocked by the HMC if just the N_0 moved, and because adjectives and relative clauses also regularly intervene between the N in its DP-initial position and the Num CL sequence, it can be assumed that movement of the ‘N’ is (p. 812) actually movement of the entire NP rather than just the N^0 as represented in (12):

(12)

$[_{DP} [_{NP} \text{dek}$	$\text{naa-rak}]_i [_{NumP}$	$\text{soong } [_{CLP}$	$\text{khon } t_i]]]$
child	lovable	two	CL
‘two cute children’			

Such a conclusion that NP-movement takes place within the DP will, in fact, also be forced by a consideration of the position of NP relative to demonstratives, if one assumes that there should be only a single direction of selection within a language, as in (13):

(13)

[_{DP} [_{NP}	dek	naa-rak]	nii]
	child	lovable	this
'this cute child'			

If one takes the general head-initial property of Thai to indicate that DPs should also be assumed to be head-initial, the sequence in (13) will also not allow for an analysis as being simply base-generated in its surface form. If the demonstrative is in a D^0 head position and Thai is head-initial, it has to be assumed that the NP has been moved leftward from an underlying complement position to the right of D^0 —namely, [_{DP} [_{NP}]_i Dem t_i]. If the demonstrative is alternatively suggested to be in a specifier position (perhaps SpecDP), because specifiers in Thai are projected to the left of phrasal heads, again it would have to be concluded that the NP has undergone leftward movement from a complement position to the right of D^0 —namely, [[_{NP}]_i [_{DP} Dem [_D] t_i]]. The assumption that the demonstrative occurs in DP-final position due to leftward movement of its complement in Thai is one that, incidentally, also has to be made in other languages of Southeast Asia, such as Hmong, Vietnamese, and Indonesian. All of these languages are regular head-initial (and Spec-initial), and all have DP-final demonstratives. Considerations of headedness as noted with Thai therefore lead to the same assumption: that the complement of the D^0 is moved leftward, leaving demonstratives in surface-final position in the DP. what is significant to note about Hmong, Vietnamese, and Indonesian is that, unlike in Thai, Num, and CL, both precede the NP within DPs, and so these languages also appear to be regularly head-initial *inside* the DP, as illustrated in Indonesian (14):

(14)

tiga	buah	sepeda
3	CL	bicycle
'three bicycles'		

(p. 813) As the sole exception to this headedness is the position of the demonstrative, the conclusion that its complement moves leftward leaving the demonstrative in DP-final position is rather straightforward to make, and any other attempted analysis faces serious difficulties in reconciling the different directions of headedness within the DP that would have to be assumed for D^0 versus Num 0 , CL 0 , and N 0 . To the extent that such a conclusion is therefore well justified in Hmong, Vietnamese, and Indonesian, it adds extra plausibility to the similar assumptions made about Thai.

A further important comparative point concerns Nung, a northern Tai language. Significantly, Nung has exactly the same ordering of elements in the DP (and elsewhere in the language) as standard Thai does, with the exception of Num and CL, which *precede* the N (in the positions they are suggested to occur underlyingly in Thai):

(15)

Classifiers and DP Structure in Southeast Asia

Nung (northern Tai):	Num	CL	N	Adj
Thai (southern Tai):	N	Adj	Num	CL

Supposing one were to attempt to argue that the ordering in standard Thai actually was base-generated as a head-final structure (despite all of the arguments against this given previously), it would then be very difficult to suggest that the same order could be base-generated in a related language except with the heads Num and CL located in quite a different position relative to the head-noun. To allow for both standard Thai and Nung, it seems that some kind of movement has to be assumed, and as all other arguments would seem to point toward an analysis of NP-movement in standard Thai, one can suggest that the [Num CL N] order in Nung simply encodes on the surface the underlying order in standard Thai.¹

Finally, an NP-movement analysis of Thai receives further support from the fact that movement of this type is directly observable in certain other languages. For example, in Indonesian it is noted that the neutral order within the DP[Num CL NP Dem] may sometimes be converted into an order with the NP initial in the DP [NP Num CL Dem], as in:

(16)

Classifiers and DP Structure in Southeast Asia

Maka	ada- pun	menge- rjakan	[loban g	sa	buah	itu]	sam- pai	lima	enam	hari.
and	indeed	make	hole	one	CL	that	took	5	6	day
'Indeed it took 5 or 6 days just to dig that one hole.' (Hopper 1986: 317)										

A similar alternation occurs in Vietnamese, as noted in Nguyen (1957). Vietnamese has the neutral order [Num CL NP] just as in Indonesian, but Nguyen points (p. 814) out that in poetry and literature and in “inventory forms,” this may be converted into [NP Num CL], just as in Indonesian.²

The theoretical arguments and empirical support that can be brought together in favor of a head-initial analysis of DPs in Thai therefore turn out to be good. What needs to be done now is to see if there is any plausible motivation for the DP-internal movement of the NP. Before I do this, however, I briefly consider a potential alternative to the conclusion that (NP-)movement is involved and the possibility that Num and CL perhaps modify the NP in some other kind of non-head-complement way. Muromatsu (1998) suggests that rather than being functional categories selecting NP, numerals and classifiers may actually be small-clause *predicates*, predicating onto NP subjects within the DP. Considering Thai within such an approach, one would not be forced to assume movement of the NP to its surface position. Instead, it could be suggested that the NP is simply base-generated DP-initially as the subject of a rightward predicate consisting of Num + CL.

In support of the “predicate” theory is the observation that numerals and certain quantifiers seem to be able to occur as predicates in various languages such as English and (classical) Chinese, as noted in Higginbotham (1987) and Pulleyblank (1995):

(17)

- a. The apostles are twelve in number.
- b. They are many (in number).

(18)

Mie-guo-zhe	wu-shi.
destroy-country-NZL	50
'His extinctions of countries were fifty.' (Meng 3B/9, in Pulleyblank 1995: 58)	

In Thai, however, there is a good reason to believe that Num + CL does not, in fact, modify the NP in any kind of subject-predicate structure. Higginbotham (1987) points out that the possibility for numerals and quantifiers to occur as predicates is critically restricted and that only certain *weak* quantifiers are found in subject-predicate structures, where predication is attempted with strong quantifiers as in (18), this is quite unacceptable:

(19) *the men are all/each

Significantly in Thai, the Num position preceding CL also hosts various quantifiers (in alternation with numerals), and these include both weak and strong quantifiers:

(p. 815) (20)

Classifiers and DP Structure in Southeast Asia

dek	soong	khon /	laai	khon /	thuk	khon
child	2	CL	several	CL	every	CL
'two children/several children/all the children'						

The possibility for strong quantifiers such as *thuk* ‘all/every’ to occur in Num preceding CL seems to clearly rule out the plausibility of any subject-predicate analysis of [NP Num CL] sequences. Such strong quantifiers are as impossible as predicates in Thai as they are in other languages:

(21)

*Dek-law-nii	thuk-(khon).
child-group-this	every-CL
Intended: ‘‘These children are all/every.’’	

Consequently, it seems that the NP-movement analysis of Thai DPs has to be maintained.

4 Motivating the Movement

The key to understanding what may cause the movement of the NP to DP-initial position can be suggested to lie in a frequently made observation about the use of [NP Num (CL)] sequences. In Greenberg (1975), Gil (1994), and various other works, it is pointed out that linear sequences of noun or NP before numeral (and classifier) are found to occur particularly often in written list or ‘‘inventory’’ forms, as well as when people are involved in situations such as ordering food in a restaurant or buying commodities in a store. Thus if both [Num CL NP] and [NP Num CL] forms are possible in a language, as in Indonesian, Vietnamese, Chinese, and various other languages, the latter [NP Num CL] ordering is frequently noted to be either a form preferred in lists or buying or food-ordering types of situations (as, for example, in Indonesian (22)), or, alternatively, it is only attested in such situations. This sequencing of noun or NP before a numeral is furthermore observed (J. Hurford, pers. comm., in Gil 1994) to occur as a conventional way of itemizing elements in written shopping lists in languages such as English which otherwise do not permit such orders, as shown in (23):

(22)

Classifiers and DP Structure in Southeast Asia

Saya	mau	membeli	beras	dua	kilo.
I	want	buy	rice	2	kilo
'I want to buy two kilos of rice.'					

(p. 816) **(23)** Sugar, 3 pounds

Bread, 2 loaves

Wine, 4 bottles

Cross-linguistically, then, it can be noted that [NP Num CL] sequences are orders which are either found to be the only possible ordering of Num, CL, and NP in a language (as, e.g., in Thai, Khmer, and Burmese), or they occur frequently in certain situations as alternatives to a possibly more common [Num CL NP] order (as, e.g., in Indonesian, Vietnamese, Chinese, and also Japanese and Korean). [NP Num CL] orders are therefore considerably widespread and occur throughout the major classifier languages in Asia. This now raises the question of why such an ordering should be so widespread and why it should be favored in the situations noted. A possible answer here is that sequencing the noun or NP before the numeral and classifier may well be a natural and useful way of ordering this kind of information in certain types of *presentational* situations, what the placement of the NP in DP-initial position effectively does in [NP Num CL] forms is to ensure that in linear terms information about the identity of the NP is presented before information about its cardinality. Such an ordering is arguably practical and useful at certain times. For example, in the case of a storekeeper receiving information about what goods a customer wishes to purchase, identifying the type of goods before the quantity (i.e., ordering noun or NP before Num CL) presents the information in a sequence that mirrors the actions of the storekeeper, who first needs to identify and locate the required goods and then select a certain quantity of them. The presentation of the information in this way may therefore be both naturally helpful and also efficient and logical. If this is indeed a plausible interpretation of why NP-initial orders are cross-linguistically particularly frequent in lists, ordering, and other presentational situations, the placement of the NP in DP-initial position can be likened to presentational focus or topicalization at the sentential level (as in fact hinted at in Greenberg 1975); in both CP and DP, nominal elements that are being newly presented may be fronted so that they linearly precede additional information being added on about them.

In this regard, note the following sequence from classical Chinese, which at one time had both [Num NP] and [NP Num CL] forms. In the initial presentation of the new referent, the fronted NP-initial order is used, while the [Num NP] order without NP-fronting occurs in the following sentence, as the referent has been established as old, identified information: (from Schafer 1948: 413):

(24)

Classifiers and DP Structure in Southeast Asia

You	da	Jiang	er	ren.	Er Jiang ...	
be	big	general	2	CL	2	general ...
'There were two great generals. The two generals ...'						

(p. 817) In some classifier languages such as Indonesian and classical Chinese, the DP-internal presentational focus is clearly optional, and the languages have both [Num CL NP] and [NP Num CL] orders. In other languages such as Thai, Khmer, and Burmese, there is no alternative to the [NP Num CL] order, and it can be suggested that the presentational focus movement has simply become obligatory, with all DPs having to be formed with the NP in prominent, DP-initial position.

If there is indeed a legitimate presentational focus-related motivation for the NP-movement, this now raises a further question. Supposing that the [NP Num CL] order results from NP-raising, it might be expected that this would have developed from earlier structures with no movement of the NP and that one would therefore find earlier forms with Num, CL, and NP simply remaining in their underlying base-generated positions—namely, [Num CL NP]. However, it does not seem possible to find such forms, and in Chinese, for example, [NP Num CL] sequences seem to have arisen spontaneously without any prior [Num CL NP] forms occurring. This therefore appears to challenge an NP-movement analysis and requires some further investigation.

A reasonable explanation of the lack of early [Num CL NP] forms can be offered here by considering a quite different theory of [NP Num CL] sequences, the “adverb” theory which suggests that postnominal Num-CL elements are actually not inside the DP at all but, rather, are adverbs base-generated in quite distinct adjunct-like positions. This approach to postnominal Num-CL sequences is developed in Fukushima (1991) and Ishii (2000) for modern Japanese, having originally been suggested in Greenberg (1975), and it is supported by a variety of evidence.

First, it is noted that in languages where Num and CL may occur following the NP, the Num and CL may also occur separated from the NP in other VP-/S-final adverbial type positions, as illustrated here in Thai (25) and classical Chinese (26):

(25)

Classifiers and DP Structure in Southeast Asia

Mii	nisit	maa	haa	khun	soong	khon.
be	student	come	find	you	two	CL
'Two students came to look for you.'						

Classifiers and DP Structure in Southeast Asia

Xi	sang	di	yu	Qin	qi-bai-li.	
west	lose	land	to	Qin	700	li
'On the west we lost 700 li's land to the Qin.' (Meng 1A/5, in Pulleyblank 1995: 58)						

Second, it is well observed that classifiers develop from other independent nouns. Consequently, it can be suggested that in the earliest classifier-type constructions before the category of classifier may have formally been established as (p. 818) a DP-internal functional-category, “classifiers” would have actually been the heads of independent NPs occurring as adverbial-type elements modified by numerals. This is represented by the bracketing given to the early (oracle bone inscription) Chinese example (27) noted in Bisang (1999), where the classifier is simply a repetition of the first noun:

(27)

[_{NP} ren]	[_{NP} shi-you-wu		ren]
Person	10-and-5	person	
'15 people'			

Third, Fukushima (1991) points out that Num and CL in Japanese can actually be coordinated with adverbs, suggesting that Num and CL constitute separate adverbial elements base-generated outside the DPs they numerically modify:

(28)

Classifiers and DP Structure in Southeast Asia

Shoonin- ga	kinoo	[san-nin	katsu	tashika-ni]	sono	jiko-o	mokugek- ishi-ta.
witness- NOM	yesterday	3 rd CL	and	certainly	that	accident- ACC	witness- PAST
'Three witnesses certainly witnessed that accident yesterday.'							

Fourth, there are instances of “floated” Num-CL pairs, which could not have been base-generated with the associated preceding NP and so which must be assumed to be base-generated independently. Although the Num-CL pair *ippatsu* ‘one blast’ can occur after the NP *pisutoru* ‘pistol’ in (29), it cannot be positioned before the NP, even though Japanese otherwise regularly allows the prenominal sequence [Num CL no NP]:

(29)

a.

Classifiers and DP Structure in Southeast Asia

Taroo-ga	pisutoru-o	ippatsu	kinoo	utta.
Taroo-NOM	pistol-ACC	one-CL(blast)	yesterday	shot
'Yesterday Taroo shot off one blast of his pistol.' (Fukushima 1991: 73)				

b.

*Taroo-ga	ippatsu-no-pisutoru	kinoo	utta.
Taroo-NOM	one-CL-GEN-pistol-ACC	yesterday shot	

The adverb theory of postnominal Num-CL can now be suggested to provide a way of understanding how [NP Num CL] DP sequences might arise without there being any previous stage of [Num CL NP]. It can be suggested that under simple linear adjacency early sequences of argument NP and numerically quantified adverbial NP (the Num-CL pair) may have come to be significantly reinterpreted as parts of a single DP. when such a hypothetical reanalysis takes place in a language such as Thai, because of the ordering of Num before CL and the general headedness of the language, speakers who reanalyze [NP Num CL] as a (p. 819) single DP will analyze the initial NP as undergoing movement to its surface position from a natural complement position following CL and interpret this movement as an instance of simple DP-internal presentational focus, effectively achieving the same linear ordering effect as the occurrence of an argument NP before a numerically quantified adverbial NP in prereanalysis structures. The “occurrence” of NP-movement in the DP can consequently be suggested to arise not from an earlier “unmoved” [Num CL NP] source but may actually be the direct and instantaneous result of the reanalysis of a rather different two-NP structure.

If such a reanalysis approach can be maintained, it will clearly explain why one does not find earlier sequences of [Num CL NP] in languages with [NP Num CL] orders. Ironically, though, what one now needs to find is good evidence that the suggested reanalysis has indeed taken place, as it could be argued as in Fukushima (1991) and Greenberg (1975) that, in fact, postnominal Num-CL sequences are still just adverbs and not part of the DP. In what remains of this section, this chapter shows that there is indeed evidence of this kind, as well as noting a certain interfering complication.

First, in Thai and Khmer there is clear evidence from the placement of demonstratives that Num and CL do occur inside the DP. As noted in an earlier section, demonstratives in Thai (and Khmer) occur in final position in the DP, following Num and CL:

(30)

Classifiers and DP Structure in Southeast Asia

[_{DP} Dek	saam	khon	nan]	keng.
child	3	CL	Dem	clever
'Those three children are smart.'				

Second, in Burmese, demonstratives are DP-initial and consequently cannot be used to provide arguments for the occurrence of Num and CL inside DP. However, the presence of case markers and postpositions following the Num and CL in DP-final position again suggests that these elements are DP-internal rather being base-generated as some kind of adverbial unit:

(31)

Classifiers and DP Structure in Southeast Asia

Canaw	[_{DP} saouq	hna	ouq]	-ko	weh	hta	teh.
I	book	2	CL	ACC	buy	Asp	NON-FU-TURE
'I bought two books.'							

Classifiers and DP Structure in Southeast Asia

U-Win-Win [_{PP} [_{DP}	meitswee	thoun	yauq]	ne]	Yangoun	thwaa	teh.
U-Win-Win	friend	3	CL	with	Rangoon	go	NON-FU- TURE
'U-Win-Win went to Rangoon with three friends.'							

(p. 820) Third, certain aspects of the interpretation of [NP Num CL] sequences suggest that Num and CL are DP-internal elements. Consider the clear contrast in interpretation between Thai (33), where Num and CL immediately follows the NP, and (34), where Num and CL occur sentence-finally in an adverbial type of position. Whereas (34) has a partitive type of interpretation, in (33) such an interpretation is significantly not available:

(33)

Classifiers and DP Structure in Southeast Asia

Dek	saam	khon	sia	chiwit	laew.
child	3	CL	lose	life	ASP
'The three children died already.'					

Classifiers and DP Structure in Southeast Asia

Dek	sia	chiwit	laew	saam	khon.
child	lose	life	ASP	3	CL
'Three of the children died already.'					

Partitive readings are generally assumed to be possible when numerical quantification is applied to a definite DP from a DP-external position, as in (35a) and are blocked when numerals occur under the scope of D^0 inside the DP as in (35b):

(35)

- a.** [_{QP} three of [_{DP} the children]].
- b.** [_{DP} the three children]

In Thai (34) the Num and CL are clearly in a DP-external position, and the result is a partitive reading. In (33), where the Num and CL occur adjacent to the NP, no partitive reading is possible. The natural conclusion to be made about [NP Num CL] sequences in examples such as (33) would therefore seem to be that Num and CL in such cases are indeed inside the DP, and this therefore blocks the possibility of a partitive interpretation. Furthermore, the only possible positioning of Num and CL relative to an overt demonstrative is before the demonstrative as in (30) and attempting to place Num and CL after a demonstrative as in (36) is simply ungrammatical, indicating that post-NP Num and CL in preverbal subject position must indeed be DP-internal:

(36)

Classifiers and DP Structure in Southeast Asia

*dek	nan	saam	khon	keng
child	Dem	3	CL	clever

Turning to Burmese, there is similar interpretative evidence that the Num CL is inside the DP. In (37), with the sequence [Dem NP Num CL] there is only possible a nonpartitive interpretation. This once again clearly suggests that the Num and CL elements are DP-internal and under the scope of D^0 :

(p. 821) (37)

Classifiers and DP Structure in Southeast Asia

canaw	eh-dii	saouq	hna	ouq	weh	hta	teh.
I	Dem	book	2	CL	buy	Asp	NON-FU-TURE
'I bought those two books. / NOT:*! bought two of those books.'							

Consequently, there is a range of evidence suggesting that if Num-CL sequences originated as adverbial elements, as seems likely, in Burmese and Thai, they have now allowed for reanalysis within the DP as suggested. Cross-linguistically, however, not all [NP Num CL] forms pattern in the same way, and there is a further complication. If one compares Burmese (37) with an apparently similar string of [Dem NP Num CL] in Japanese, one finds that the Japanese sequence actually has a different and opposite interpretation from the Burmese, as indicated:

(38)

Classifiers and DP Structure in Southeast Asia

Jiro-wa	sono	hon-o	san	satsu	katta.
Jiro-TOP	DEM	book-ACC	3	CL	bought
'Jiro bought three of those books. / NOT:*Jiro bought those three books.' (Muromatsu 1998: 65)					

Here Japanese and Burmese might seem to be significantly different, suggesting perhaps that postnominal Num CL sequences in Japanese are not inside the DP but adverbial, as argued in Fukushima (1991) and Ishii (2000). Interestingly, further patterns in Burmese show that Burmese may also still allow for such an adverbial possibility under certain explicit circumstances. Although it has been noted in (31) that case markers can occur after the Num CL sequence, it is actually also possible for a case marker to occur between the NP and Num CL, as in (39). Significantly, when this does occur (and case markers are used sparingly in Burmese), it allows for the same partitive-type interpretation that occurs in Japanese (38), suggesting that it is also still possible for Num CL in Burmese to occur and be interpreted outside of the DP:

(39)

Classifiers and DP Structure in Southeast Asia

U-Win-Win	eh-dii	saouq	-ko	hna	ouq	weh	teh.
U-Win-Win	Dem	book	ACC	2	CL	buy	NON-FU-TURE
'U-Win-Win bought two of those books.'							

Consequently, while (30–34) allow for the conclusion that adverbial Num CL has indeed allowed for reanalysis in both Burmese and Thai, examples such as (34) and (39) suggest that it nevertheless may still be possible for Num CL to occur unreanalyzed and adverbially in DP-external positions in both languages (VP-finally in Thai and following case-marked DPs in Burmese), this having clear effects on the interpretations of such forms.

Before this section closes, one last pattern that can usefully be considered here is the historical development of modern-day [NP Num CL] forms in Khmer. (p. 822) Briefly, old Khmer had two possible forms with numerals, either simply [NP Num] or alternatively [NP CL Num] with a classifier. In middle Khmer, as well as the latter [NP CL Num] forms, a second sequence [NP Num CL] is found, and it is this sequence that is now the sole modern form, as schematized in (40):

(40)

	<i>Old Khmer</i>	<i>Middle Khmer</i>	<i>Modern Khmer</i>
a.	NP Num		
b.	NP CL Num	NP CL Num	
c.		NP Num CL	NP Num CL

Such a sequence of development is both interesting and revealing and seems to suggest the following explanation. In old Khmer when “classifiers” first begin to occur, these elements were really just nouns and not grammaticalized as a distinct special category. NPCL Num forms were therefore simply sequences of two NPs, and Num was attached to the right of a second adverbial NP in the same right-ward position that it otherwise attaches in in pattern (a). [NP CL Num] is therefore really [NP₁] [NP₂ Num], essentially in line with Greenberg (1975) and earlier proposals. Later on, in middle and modern Khmer, it can be suggested that there occurred natural grammaticalization of the classifier as a nominal functional category, and two-NP structures became reanalyzed as single DPs with Num and “CL” internal to the DP. Such a hypothetical grammaticalization process can now be suggested to be directly responsible for the otherwise puzzling and important change in CL Num word order which occurred, reversing the linear order of CL and Num and effectively replacing pattern (b) with pattern (c). what such a change seems to suggest is that when grammaticalization and reanalysis occurred, this significantly forced Num and CL as functional heads within the DP to be realigned in a *head-initial* order, following the general direction of headedness found elsewhere in Khmer. If correct as an explanation of the switch in CL Num word order, such a change is important in clearly showing the strong pressure that languages may be under to adopt consistent head-initial orders. One could imagine, for example, that it might be simpler for the original linear order [NP CL Num] to grammaticalize as a head-final structure without any reversal of the orders of Num and CL. That this did not happen, and CL seems to have grammaticalized in a specifically head-initial way, reinforces the view that grammaticalization and reanalysis do not occur in any random fashion and that there are clear principles of headedness governing the organization of DPs in the languages examined here.

Summarizing briefly the conclusions of sections 2–4 now, the general goal of the investigation here was to see whether there is any real regularity in DP structure (p. 823) in Southeast Asian languages and what factors might be responsible for surface variation. A significant cause of cross-linguistic word-order variation has now been identified as the obligatory application of DP-internal NP-movement in certain languages but not others. It has been suggested that such movement is essentially the result of the reanalysis of an earlier adverbial form, that similar adverbial forms may still exist in certain languages, and that the switch in CL Num word order in middle Khmer is understandable once one adopts such an account of the development of classifiers in DPs. In sections 5–7, the chapter now moves on to consider other aspects of the structure of DPs in Southeast Asian languages and suggests how, in certain cases, X⁰-movement and grammaticalization may also be responsible for further surface variation attested.

5 Bare Classifiers and Definiteness

Although the use of classifiers with numerals is common and highly developed throughout the Southeast Asian area, one interesting classifier pattern is found in just a subset of the languages of the region: the occurrence of bare classifier-NP sequences without any accompanying numeral as in (48–50) from Vietnamese, Hmong, and Nung:

(41)

Nguoi	chong	rat	tot. (Vietnamese)
CL	husband	very	good
‘The husband was very good.’ (Daley 1998: 65)			

(42)

Classifiers and DP Structure in Southeast Asia

Tus	tsov	tshaib	tshaib	plab. (Hmong)
CL	tiger	hungry	hungry	stomach
'The tiger was very hungry.' (Jaisser 1987: 171)				

Classifiers and DP Structure in Southeast Asia

Leo	tu	me	da	tu	po	va ... (Nung)
then	CL	wife	scold	CL	husband	say
'Then the wife scolded the husband and said ...' (Saul and Wilson 1980: 160)						

As such patterns do not occur in Thai, Khmer, Burmese, or Indonesian, this raises the question of whether classifiers should be assumed to have a different syntactic status in different languages and how the bare classifier phenomenon should be accounted for.

(p. 824) Considering the general patterns found in Vietnamese, Hmong, and Nung, an important observation is that bare classifier-NP sequences are commonly associated with referentiality and definiteness effects, so that when a DP has a definite interpretation, a bare classifier is generally found to occur with it, as in (41–43) (see Daley 1998, Löbel 1996, Bisang (1999), and Nguyen 1975). Cheng and Sybesma (1999) also note and investigate similar patterns in Cantonese and suggest that classifiers are used to express definiteness in Cantonese in a way similar to the use of definite determiners in other languages. As classifiers are assumed to occur in CL^0 , however, it is argued that bare-classifier nominal expressions with definite interpretations in Cantonese are simply Classifier Phrases (CLPs) and do not project any higher functional structure such as NumP or DP. Turning to examine the Southeast Asian languages now, this chapter will explore a slightly different approach and suggest that bare-classifier structures in these languages may be regular DPs resulting from raising of the classifier from CL^0 up to D^0 .

A critical aspect of the general classifier patterning that needs to be accommodated in any analysis is the fact that although bare [CL NP] forms may naturally be interpreted as being definite, where a numeral occurs preceding the classifier, as in (44), such sequences have only indefinite interpretations:

(44)

Classifiers and DP Structure in Southeast Asia

Toi	mua	tam	cai	ghe. (Vietnamese)
I	buy	8	CL	chair
'I bought eight chairs.'				

Such a patterning is perhaps unexpected if classifiers are taken to regularly cause interpretations of definiteness when inserted into CL⁰. One might expect that either the definiteness of CL in (44) should become a property of the whole nominal expression leading to a definite interpretation or, if the numeral has quantificational scope over the definiteness of its CLP [CL NP] complement, one might alternatively expect a partitive interpretation ‘three of the books’, yet this is also not a possibility. Basically, there would seem to be a need to capture in some way the fact that classifiers in these languages may be definite or may express definiteness at some times but not at other times. A further observation relevant here is that when an otherwise indefinite sequence of [Num CL NP] is combined with a demonstrative, the result is that the whole expression is clearly definite:

(45)

goh	saam	bo	sue (Cantonese)
Dem	3	CL	book
‘those three books’			

(46)

slong	ahn	sleng	te (Nung)
2	CL	province	Dem
‘those two provinces’ (Saul and Wilson 1980)			

(p. 825) The definiteness effects induced by demonstratives in some position above Num and CL might seem to indicate that the locus of definiteness should indeed be taken to be a D⁰ position above NumP, as often assumed, and that it is the occurrence of a particular morpheme in either the D⁰ position or SpecDP which results in the interpretation of the DP as being definite. If this is indeed reasonable to assume, and if it is also reasonable to generalize further from structures with demonstratives to structures without demonstratives so that definiteness is assumed to be encoded in a constant position within nominals, it can now be suggested that bare classifier [CL NP] forms may perhaps be interpreted as being definite via the overt association of the classifier with the D⁰ position through raising of the classifier up to D⁰ in instances when a DP has a definite interpretation. It can be suggested that in the languages in question, either the D⁰ or SpecDP must be overtly instantiated by a certain lexical element (a demonstrative or the classifier) for the definite interpretation to be triggered or signaled and that otherwise the DP will be interpreted as having a default indefinite value. [Num CL NP] forms such as (44) will consequently be automatically interpreted as indefinite, as the classifier remains in situ in CL and has clearly not been raised to D. Furthermore, when Num does occur, it will not be possible to raise CL (over Num) to D because of the Head Movement Constraint, explaining the nonoccurrence of forms such as [^{*}CL_i Num t_i NP], and instead a demonstrative has to be inserted either into D⁰ (or possibly SpecDP) to trigger a definite interpretation.³

The suggestion that CL may (sometimes) move to D and cause definite interpretations of the DP will certainly account for the basic patterns observed here and would also seem to be a reasonable way of explaining the ambivalent nature of CL, classifiers sometimes be-

ing associated with definite interpretations (when they raise to D^0) and sometimes not (when they remain in CL^0).⁴ However, to be more convinced of the plausibility of a CL-to-D approach, one might hope to find additional empirical evidence of the higher D position. Such evidence interestingly exists in Vietnamese, and one finds examples where a second general classifier element occurs preceding the regular classifier, resulting in sequences with clear definite interpretations:

(47)

Classifiers and DP Structure in Southeast Asia

Cai	con	dao	[anh	cho	toi	muon],	no	that	sac.
CL	CL	knife	you	give	me	borrow	it	real	sharp
'The knife you gave me is really sharp.' (Le 1968)									

(48)

cai	chiec	ban	nay
CL	CL	table	Dem
'this table' (Nguyen 1975)			

(p. 826) What this shows is that there is indeed another X^0 head-position above CL^0 and, importantly, it is a head-position of just the type suggested, one that is both instantiated by a classifier and specifically associated with definite interpretations of the DP. Such patterns therefore seem to provide clear empirical support for the CL-to-D hypothesis and indicate that classifier elements may indeed sometimes occur in higher D^0 type heads in definite DPs. Historically, it can be suggested that the possibility of inserting a general classifier directly into the higher posited D^0 position has resulted from a sequence of movement and reanalysis. After a certain initial period of simple CL-to-D movement, with the classifier instantiating both heads CL and D, it can be suggested that frequent raising of the general classifier to D^0 may have allowed it to be reanalyzed as (potentially) just a D^0 element permitting simple insertion into D^0 and allowing CL^0 to be lexicalized and instantiated by a second classifier as in (47–49). In Simpson (1998), Simpson and Wu (2000), Wu (2004), and also Roberts and Roussou (2003), it is suggested that grammaticalization may indeed frequently consist in just this kind of movement and reanalysis sequence. In Wu (2000), one particularly relevant example of this is argued to be the reanalysis of the general classifier *ge* in Mandarin as an *indefinite* determiner in D^0 after similar raising from the CL^0 position. As shown in (49), the reanalysis of *ge* in D^0 now allows for the CL^0 position to be instantiated by a new classifier:5

(49)

Classifiers and DP Structure in Southeast Asia

he	ge	san	ping	jiu
drink	GE	3	CL	wine
'do a drinking of three bottles of wine'				

Synchronically, it can be assumed that if the general classifier in Vietnamese occurs selected in the numeration, together with a second regular classifier, the general classifier will be inserted into D^0 , and there will be no CL-to-D movement (47–48). Otherwise, however, if an additional general classifier is not selected, it can be argued that the classifier base-generated in CL^0 will instead have to undergo CL-to-D to give rise to the definite interpretation of a DP as suggested.⁶ Consequently, then, CL-to-D, demonstrative insertion into D^0 /SpecDP, and general classifier insertion into D^0 (where available) can all be suggested to achieve the same basic goal of overtly specifying the DP as being definite.

6 Num-Raising and Indefiniteness

Having argued that Vietnamese has come to have double classifier sequences as the result of head movement and reanalysis, I now consider another revealing (p. 827) case where head movement results in variation in DP surface structure: the patterning of number ‘one’ in Thai and Nung. In both these Tai languages, the regular position for numerals is preceding the classifier, as in (50). The number ‘one,’ however, is commonly placed *following* the classifier in Thai and can occur *only* in DP-final position in Nung (51a–b):

(50)

a.

dek	saam	khon (Thai)
child	3	CL
‘three children’		

b.

slam	ahn	vet (Nung)
3	CL	spoon
‘three spoons’ (Saul and Wilson 1980)		

(51)

a.

dek	khon	nung (Thai)
child	CL	one
‘one/a child’		

b.

ahn	tahng	nuhng (Nung)
CL	chair	one
'one/a chair' (Saul and Wilson 1980)		

Considering the placement of 'one' either before or after CL in Thai, it is unlikely that this results from a linear inversion rule, and simple inversion of the classifier with the number "one" clearly cannot account for the related postposing of 'one' in Nung. Importantly, what both Thai and Nung can be shown to share in their repositioning of the number 'one' is that 'one' is now commonly placed in the DP-final position where demonstratives otherwise occur, as in (52). The Thai example (53) shows that 'one' in this final position is indeed in complementary distribution with demonstratives, though it may nevertheless occur in the regular pre-CL Num⁰ position with a demonstrative following CL:

(52)

a.

dek	song	khon	nii (Thai)
child	2	CL	Dem
'these two children'			

b.

slong	ohng	dehk	te (Nung)
2	CL	child	Dem
'those two children' (Saul and Wilson 1980)			

(53)

a.

*dek	khon	nung	nil
child	CL	one	Dem

(p. 828) **b.**

dek	nung	khon	nil
child	one	CL	Dem
'this one child'			

It can therefore be argued that the number 'one' in both Thai and Nung is coming to be an indefinite determiner that contrasts in its indefinite specification with the definiteness encoded by demonstratives.⁷

Before we see how 'one' comes to be in the Dem position, let us consider again briefly how the demonstrative and other DP-internal elements achieve their surface order in Thai and Nung. In sections 3 and 4, I argued at length for the occurrence of NP-movement over Num and CL in Thai—that is, $[NP_i \text{ Num CL } t_i]$. I have also suggested that when a demonstrative occurs in DP-final position in Vietnamese, Nung, Indonesian, Thai, and Khmer, the elements found to the left of the demonstrative are raised to this position as a single constituent from an underlying position to the right of the demonstrative, hence $[[\text{Num CL NP}]_i \text{ Dem } t_i]$. In Thai and Khmer, when a demonstrative occurs, there will effectively be two applications of leftward movement involved in the derivation, as illustrated in (54). First, the NP will raise over Num and CL to a position between these elements and Dem, and then the $[NP \text{ Num CL}]$ constituent will raise over Dem itself (as in Vietnamese, Nung, etc.), resulting in the surface linearization of examples, such as (52):⁸

(54)

Underlying structure:	$[\text{ Dem Num CL NP}]$
NP movement:	$[\text{ Dem NP}_i \text{ Num CL } t_i]$
movement over Dem:	$[[NP_i \text{ Num CL } t_i]_k \text{ Dem } t_k]$

Concerning the occurrence of the number 'one' in the Dem position, it can be suggested that 'one' is first base-generated in Num^0 and then undergoes head-raising up to D^0 . It is not possible to have any lexicalization of the Num position when 'one' is in the Dem position, suggesting that 'one' has indeed raised up from Num^0 (and semantically there is no reason that numbers should not co-occur with an indefinite determiner, as in Chinese (49)):

(55)

*dek	saam/nung	khon	nung
child	3/one	CL	'one'

The derivation that results in 'one' occurring in D^0 in Thai can therefore be suggested to be as schematized in (56) with 'one' originating in Num^0 . 'One' will first raise up to D^0 , and then this will be followed by regular movement of the NP over Num and CL to its landing site below the Dem position. Finally, there will occur raising of the constituent to the right of Dem over Dem to its left, (p. 829) resulting in 'one' becoming final in the DP in a way entirely similar to the DP-final occurrence of demonstratives. Note that with the single exception of the suggestion that 'one' raises up to D^0 , the analysis needs to make no new assumptions and simply makes use of mechanisms already argued for:

(56)

Underlying structure:	['one' CL NP]
'One' raises from Num^0 to D^0 :	[one_i [t_i CL NP]]
NP movement over Num CL:	[one_i [NP _k t_i CL t_k]]
Movement over the Dem position:	[NP _t t_i CL t_k] _m [one_i t_m] (Containing 'one')

In addition to showing how the occurrence of head-movement results in further clear-surface variation in the DP, these patterns with 'one' in Thai and Nung also allow for three more general conclusions. First, the patterns can be argued to provide good support for the general underlying structures and movement suggested to occur in Thai, Khmer, and similar languages. If one assumes that simple classifier-number linear inversion is not possible as an analysis (and made more unlikely by the facts in Nung), and if it is assumed that 'one' raises up to D^0 from Num^0 (thus accounting for the fact that Num^0 may not be independently lexicalized when 'one' occurs in D^0), this indicates again strongly that Thai DPs with the surface order [N Num CL D] (D = Dem) cannot, in fact, be simple head-final structures. If such an ordering were to directly reflect the base-generated position of heads in the language, raising of 'one' from Num^0 to D^0 should be straightforwardly blocked by the intervening CL^0 head and the HMC. The fact that raising from Num^0 to D^0 does seem to occur can arguably be accounted for *only* if Num^0 occurs higher than CL^0 , which itself then entails that the N(P) must be assumed to have undergone leftward raising from some other position lower than CL^0 .

Second, the patterning here also provides further good support for the general assumption that Num^0 and CL^0 are independent functional heads. If the number 'one' is able to raise out of the regular position of numerals and up to a higher (D^0) position, this suggests that classifiers are not simply suffixes attached to numerals in a single head, as affixes and subparts of words are not normally stranded by operations of movement. Instead, it would seem to indicate that numerals and classifiers are independent words in discrete functional head positions.

Third, the fact that 'one' in Thai and Nung is becoming an indefinite determiner and is targeting the same position as definite D^0 elements (demonstratives) strongly supports

the assumption that both definiteness and indefiniteness are encoded in the same D^0 position in DPs and, hence, that indefinite nominal expressions are indeed DPs in at least the Tai languages. In recent years there has been regular discussion about whether indefinite nominal expressions might possibly (p. 830) be constituents which are smaller than full DPs, being NPs or alternatively NumPs; however, often it is difficult to find clear empirical evidence in favor of either a DP or an NP or NumP analysis. Here in Thai and Nung, though, because of the interesting surface linearization of the DP, we find that the D^0 position is actually not adjacent to the NP or the NumP. Consequently it is possible to see the number ‘one’ developing as an indefinite determiner in a position that is quite distinct from the numeral position, and as this position otherwise hosts demonstratives, it is straightforward to make the important conclusion that indefinite nominal expressions can indeed be full DPs and of the same syntactic size as their definite counterparts. Furthermore, if the indefinite nature of *some* nominal expressions is expressed overtly in the D^0 position, it might be natural to assume that the D^0 position is indeed regularly the locus for the specification of indefiniteness and hence that there is a phonetically null counterpart to overt ‘one’ occurring with other numerals in indefinite DPs such as (50), for example.⁹ The ‘one’ paradigm in Thai and Nung thus supports a variety of possible insights into the structure of DPs and provides important information about the derivation underlying DP surface order.

In the final section of the chapter now, I continue to examine how variation in surface structures may relate to the way that DP-internal heads are physically instantiated, this time looking further down in the DP below D^0 and Num⁰.

7 N-to-CL

A last set of clear classifier-related variation in the DP which remains in need of explanation is the occurrence of ‘classifier-less’ DPs—instances where no classifier occurs with a numerically quantified NP despite individuation being implied. This phenomenon essentially occurs in two basic forms. First, in a wide range of languages, certain *particular* nouns are commonly quantified without any apparent classifier. Frequently, the subset of nouns that pattern in this way is very similar and includes words for units of time such as ‘year’, ‘day’, and ‘time’; sometimes the word for ‘person’, and certain other terms used with fairly high frequency, as illustrated in (57):

(57)

a.

tu	nam
4	year
‘four years’	

b.

ba	Ian (Vietnamese)
3	time
'three times'	

The second way that classifierless forms manifest themselves is not lexically restricted in this way, and in certain languages (e.g., colloquial Minangkabau and (p. 831) Indonesian) the use of a classifier appears to be quite optional with a fully wide range of nouns, as, for example, in Nung (58):

(58)

slam	(tew)	kha-lo (Nung)
3	(CL)	road
'three roads' (Saul and Wilson 1980)		

There might seem to be two possible analyses of the patterns here. Considering just the former lexically restricted type in (57), if the forms in (58) are taken to be genuinely classifierless and without any CL position, this might indicate a significant distinction among nouns in classifier languages and suggest that some nouns simply require no individuation. However, such an approach would not seem to be able to generalize further to cover the second type of classifierless forms in (58) as here the nouns certainly do have classifiers. The observation about this second type is just that the classifiers are used optionally, not that they are inherently individuated. A second approach to the former lexically restricted type might alternatively be to suggest that the elements 'year' and 'time' are actually base-generated in CL and that there is no N or NP in such constructions. However, this is somewhat unlikely as the nonoccurrence of an N-position and an NP-complement to CL would entail that there are classifiers which classify or individuate nothing. As classifiers are essentially by definition functions that apply to some second complement term, the presence of a classifier would therefore seem to require the presence of an N or NP. Furthermore, such an approach will again not generalize to cover the second type in (58), as the overt elements here are clearly nouns.

In addition to these two possible analyses of cases such as (57), there is a third potential analysis of these patterns suggested by the phenomena examined in sections 5 and 6. If the analysis of Thai and Nung 'one' moving from Num to D, and of Vietnamese CL moving from CL to D is correct, it could be that similar movement takes place in instances such as (57), with the elements 'year and time' being base-generated in the N position and then raising to the CL position, instantiating both in the same way that 'one' instantiates both Num and D in Thai and Nung, and that Vietnamese classifiers sometimes instantiate both CL and D. Considering the simple patterns in Vietnamese, Indonesian, and Chinese

languages where CL and N are positions immediately adjacent to each other in the tree, there is no obvious way of finding empirical evidence in favor of one hypothesis rather than the other. However, Thai and Khmer, with their critically different surface linearization of DP structure, do provide a clear and useful means to check these hypotheses, as in these languages N and CL are not linearly adjacent in surface forms but commonly are separated by numerals in Num (i.e., in the order [N(P) Num CL]). As the Thai data in (59) show, in fact, elements such as (p. 832) ‘year’ and ‘time’ consistently appear to the right of Num in the CL position and not in the N-position, which occurs to the left of Num in surface word order:

(59)

a.

soong	pii
2	year
‘two years’	

b.

saam	khraang
3	time
‘three years’	

c.

sii	khon
4	person
‘four people’	

This indicates that the first possible analysis that no CL position occurs in ‘classifierless’ forms cannot be correct: Thai and Khmer show that with elements such as ‘year’ and ‘time’ a CL position is indeed present and overtly occupied by these elements. As the second possibility that there is no N position in such structures is rather implausible for the reasons noted previously, and as ‘year’ and ‘person’ and the like also do appear very nominal and likely to be N-heads, it would seem that the most likely explanation of the forms in (57) and (58) is indeed the third possibility: that the patterns found in (57) and (58) are ultimately the simple result of movement from N to CL. Once again, then, as with the Num-to-D and CL-to-D raising considered in sections 5 and 6, it would seem that the surface variation attested (i.e., the difference between (57) and (58) and more regular forms such as, for example, (50)) is simply due to the application of overt

head movement in certain instances and a single lexical element coming to instantiate two discrete syntactic head positions in the DP-tree.

The N-to-CL analysis of (57) is also given further plausibility and support by certain other patterns found. Such an analysis basically suggests that both N and CL positions are always projected in the DP structure and that various nouns raised from N to CL instantiate both N and CL positions in a single derivation. In some sense, this amounts to suggesting that some nouns might be able to function as their own classifiers and lexicalize the CL head, as well as the N head. Interestingly, in a range of languages such as Thai, Burmese, and Lahu, one finds exactly this possibility in a slightly different guise and the frequent occurrence of “self-classifiers” or “repeaters”—the simple repetition in CL of the element in N as in (60):

(60)

a.

hoong	saam	hoong (Thai)
room	3	room/CL
‘three rooms’		

b.

cun	ta	cun (Burmese)
island	one	island/CL
‘one island’		

The idea of an element being used to classify or individuate itself is consequently both plausible and commonly attested. Assuming the “copy theory” of movement (Chomsky 1995), one might also suggest that in these repeater cases there may again quite possibly be movement between N and CL, with the difference between (p. 833) repeater and nonrepeater cases such as (57) being that in (60) the copy left in N by movement to CL is not deleted and is actually spelled out phonetically.

Finally, unlike the other two possible analyses of (57) I considered, the N-raising analysis of cases such as (57) will importantly also extend in a natural way to cover the second type of lexically unrestricted classifierless Ns (i.e., (58)), and it can be suggested that rather than assume that the CL position is not present in such cases, the N may instead be taken to raise up to CL to lexicalize this position. The optionality here would then relate to whether a classifier is selected from the lexicon for use or not, and if one is not selected, the N will simply be used to instantiate the CL position.¹⁰ Furthermore, it can be suggested that if such N-to-CL movement does occur regularly and if the “movement and reanalysis” approach to grammaticalization is correct, one might expect that the frequent

association of certain Ns with the CL position due to regular N-to-CL raising might ultimately allow for the reanalysis of such Ns as simple direct instantiations of CL, base-generated in CL *without* having first been raised there from N⁰ (just as Vietnamese *cai* (47–48) and Mandarin *ge* (49), both originally CLs, can now be base-generated in D). As noted earlier, it is well observed that in fact, classifiers do frequently develop from elements that were originally nouns. The assumption of N-to-CL movement, together with a movement and reanalysis view of grammaticalization, therefore provides a clear way of understanding how the creation of new classifiers from nouns may indeed occur. Continued use of N-to-CL results in the reanalysis of nouns in CL and allows for relexicalization of the lower N position with new nominal elements.

8 Concluding Remarks

This chapter has been concerned with attempting to understand how apparent variation in DP structure may be accounted for across a broad range of Southeast Asian languages. Sections 2–4 concluded that significant distortion of underlying DP structure is often caused by XP-movement inside the DP (i.e., NP-fronting). Sections 5–7 considered other aspects of DP-internal variation and argued that, in each case, the patterns result from simple X⁰-movement and the common raising of DP-internal heads to lexicalize and instantiate other higher head positions, a process which naturally leads to reanalysis and full grammaticalization of a head in a higher DP-internal position. In both parts of the chapter, then, variation has essentially been attributed to elements being in different stages of ongoing historical development and reanalysis, with both XP-movement and X⁰-movement becoming regularized, reanalyzed, and grammaticalized. Throughout

(p. 834) the discussion, I attempted to show how important a role the classifier may often have in leading to possible explanations of the variation found. Finally, on a general level, the patterns investigated here not only suggest that the structure of DPs in Southeast Asian languages is far from random but also, in fact, arguably seem to indicate the opposite conclusion and suggest that DPs in the variety of languages examined here may actually share a single, basic, highly regular underlying structure.¹¹

References

- Aikhenvald, Alexandra (2000) *Classifiers*. Oxford University Press, Oxford.
- Bhattacharya, Tanmoy (1999) “The Structure of the Bangla DP.” Ph.D. diss. University College London.
- Bisang, Walter (1999) “Classifiers in East and Southeast Asian Languages: Counting and Beyond,” in J. Gvozdanovic (ed.) *Numeral Types and Changes Worldwide*, Mouton de Gruyter, Berlin, 113–185.
- Cheng, Lisa, and Rint Sybesma (1999) “Bare and Not-So-Bare Nouns and the Structure of NP.” *Linguistics Inquiry* 30: 509–542.

-
- (p. 837) Chomsky, Noam (1995) *The Minimalist Program*. MIT Press, Cambridge, Mass.
- Cinque, Guglielmo (1994) "On the Evidence for Partial N-movement in the Romance DP," in G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, and R. Zanuttini (eds.) *Paths towards Universal Grammar: Studies in Honor of Richard Kayne*. Georgetown University Press, Washington D.C., 85–110.
- Cinque, Guglielmo (1996) "The 'Antisymmetric' Programme: Theoretical and Typological Implications." *Journal of Linguistics* 32: 447–464.
- Daley, Karen (1998) *Vietnamese Classifiers in Narrative Texts*. Summer Institute of Linguistics, University of Texas at Arlington.
- Fukushima, Kazuhiko (1991) "Generalized Floating Quantifiers." Ph.D. diss., University of Arizona.
- Gil, David (1994) "Summary: Numeral Classifiers." *Linguist List* 5: 466.
- Greenberg, Joseph (1975) "Dynamic Aspects of Word Order in the Numeral Classifier," in C. Li (ed.) *Word Order and Word Order Change*. University of Texas Press, Austin, 27–46.
- Higginbotham, James (1987) "Indefiniteness and Predication," in E. Reuland and A. ter Meulen (eds.) *The Representation of (In-)definiteness*. MIT Press, Cambridge, Mass., 43–70.
- Hopper, Paul (1986) "Some Discourse Functions of Classifiers in Malay," in C. G. Craig (ed.) *Noun Classes and Categorization*. John Benjamins, Amsterdam, 309–325.
- Ishii, Yasuo (2000) "Floating Quantifiers in Japanese: what Are They and what Do They Tell Us?" Talk presented at School of Oriental and African Studies, University of London.
- Jaisser, Annie (1987) "Hmong Classifiers." *Linguistics of the Tibeto-Burman Area* 10: 169–176.
- Kawashima, Ruriko (1993) "The Structure of NPs: Arguments for Quantifier Phrase and Number Phrase." *Cornell Working Papers in Linguistics* 11: 56–72.
- Le Van Ly (1968) *So-thao ngu-phap Viet-Nam* [Sketch of Vietnamese grammar]. Trung-tam Hoc-lieu, Saigon.
- Löbel, Elisabeth (1996) "Klassifikatoren: Eine Fallstudie am Beispiel des Vietnamesischen." Master's thesis, University of Cologne.
- Muromatsu, Keiko (1998) "On the Syntax of Classifiers." Ph.D. diss., University of Maryland.
- Nguyen Dinh Hoa (1957) "Classifiers in Vietnamese." *Word* 13: 124–152.

Nguyen Tai Can (1975) *Ngu-phap tieng Viet: tieng, tu ghep, doan-ngu*. [Vietnamese grammar: morphemes, compounds and phrases]. Dai-hoc and Trung-hoc Chuyen-nghiep, Hanoi.

Pan, Hai-Hua (1990) "Head Noun Movements in Chinese, the ECP and Minimality." Unpublished ms., University of Texas at Austin.

Pulleyblank, Edwin (1995) *Outline of Classical Chinese Grammar*. University of British Columbia Press, Vancouver.

Rizzi, L. (1997) "The Fine Structure of the Left Periphery," in L. Haegeman (ed.) *Elements of Grammar*. Kluwer, Dordrecht, 281–337.

Roberts, Ian, and Anna Roussou (2003) *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge University Press, Cambridge.

Saul, J. E., and N.F. Wilson (1980) *Nung Grammar*. Summer Institute of Linguistics, University of Texas at Arlington.

(p. 838) Schafer, Edward (1948) "Noun Classifiers in Classical Chinese." *Language* 24: 408–413.

Simpson, Andrew (1998) "On the Re-analysis of Nominalizers in Chinese, Korean and Japanese." Paper presented at the University of Southern California Symposium on East Asian languages, November.

Simpson, Andrew, and Zoe Wu (2000) "The Grammaticalization of Formal Nouns and Nominalizers in Chinese, Japanese and Korean", in T. E. McAuley (ed.) *Language Change in East Asia*. Curzon, London, 250–283.

Simpson, Andrew, and Zoe Wu (2002) "Agreement, Shells and Focus." *Language* 78 2: 287–313.

Sproat, Richard, and Chin Shih (1991) "The Cross-linguistic Distribution of Adjective Ordering Restrictions", in C. Georopolous and R. Ishihara (eds.) *Interdisciplinary Approaches to Language*. Kluwer, Dordrecht, 565–593.

Tang, C.-C. Jane (1990) "Chinese Phrase Structure and the Extended X'-Theory." Ph.D. diss., Cornell University, Ithaca, N.Y.

Tang, Sze-Wing (1999) "Some Speculations about the Syntax of Noun Phrases." *UCI Working Papers in Linguistics* 5: 135–154.

T'ung, P. C., and D. E. Pollard (1982) *Colloquial Chinese*. Routledge, London.

Wahba, Wafaa (1984) "Wh-constructions in Egyptian Arabic." Ph.D. diss., University of Illinois at Urbana-Champaign.

Classifiers and DP Structure in Southeast Asia

Walters, J. (1981) "A Phonology and Morphology of Ejagham: With Notes on Dialect Variation." Ph.D. diss., University of California Los Angeles.

Wu, Zoe (2004) *Grammaticalization and Language Change in Chinese: A Formal View*. Routledge Curzon, London.

Andrew Simpson

Andrew Simpson, School of Oriental and African Studies, London