

# Classifiers and DP structure in Southeast Asia

Andrew Simpson

In *The Oxford Handbook of Comparative Syntax*, Guglielmo Cinque and Richard Kayne (eds.), 806-838, Oxford University Press: Oxford.

## 1. Introduction

South east Asia is a geographical area which 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. The particular focus of interest of the present paper is the internal structure of DPs and the specific problem of how to account for the considerable amount of cross-linguistic variation which appears to occur in the ordering of constituents in DPs. Some of the patterns found are schematised below 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,<br>Vietnamese: | Num | CL  | N  | Adj | RC  | Dem |

What is of considerable interest is the fact that the variation attested often does not seem to follow or correspond to the apparent headedness of the relevant languages. Thus, for example, Thai and Khmer are both canonical SVO head-initial languages yet at first sight 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 post-nominal 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. The chapter therefore sets 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 south east 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 south east Asian languages, a first important issue which needs to be examined and clarified is the syntactic status of the classifier elements which 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 comprising a single functional head labelled 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 or distinct heads is however an important question with significant consequences and it is therefore important to consider what arguments there are in favour of either of the two possible analyses.

Gil (1994) identifies the following observations as being potential support for the view that numerals and classifiers comprise a single syntactic unit. First of all, as observed in Greenberg (1975), numbers and classifiers commonly occur together as a single *uninterrupted* sequence. Secondly, 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. Thirdly, 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 classifier as a single unit. Such potential arguments in favour of a single head analysis are however 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, Num<sup>0</sup> 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 grammaticalise, as indeed noted in Gil (1994), and again does not rule out the possibility that classifiers might encliticise 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 does not obviously favour a single-head analysis.

The arguments supporting a two-head hypothesis are considerably stronger. First of all 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 south east Asia are also by and large phonologically quite unreduced and so appear to be fully independent functional words rather than inflectional affixes.

Secondly, classifiers are *functional* elements argued by Muromatsu (1998), Cheng & 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 instead be conceived of a non-individuated group whose total is simply specified numerically. This assumed distinction between the functions of numerals and classifiers has clear observable consequences in certain classifier languages and one finds clear evidence of two types in favour of the separation of numerals and classifiers into two formally independent heads. First of all, 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) tus tsov tshaib tshaib plab  
 CL tiger hungry hungry stomach  
 ‘The tiger is/was very hungry.’ (Hmong, Jaisser 1987)

Secondly, one finds the converse situation, that numerals may sometimes occur without any accompanying classifier, typically when the numeral specification is rather vague and individuation is not necessarily implied. Thus in Nung and Burmese classifiers are optional with numbers which are multiples of ten, in Jingpo classifiers are often omitted with numbers over ten, and ‘in Thai classifiers do not occur with large numbers like 1000 unless individuation is implied’ (Aikhenvald (2000), p.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) adalah dua tiga pondok kecil-kecil bersama-sama dekat rumah Temenggong  
 be 2 3 hut small small together near house Temenggong  
 'There were two or three small huts close together near Temenggong's house.'

Finally, Bisang (forthcoming) 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 Lebel (1996) (in (6) no classifier occurs individualizing *phong* 'room'):

- (6) nha ba phong  
 house 3 room  
 'a three room house'

Such patterns are good indication 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 favour 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) an ahn tahng nuhng ma  
 take CL chair one come  
 'Bring a chair.' (Saul & Wilson 1964)

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

- (8) a-mege ' i-cokud a-bae  
 NC-CL GEN NC-orange seed NC-two  
 'two orange seeds' Watters (1981)

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.

Finally, there are languages which 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                      b. liu da jian xingli  
     one small CL book                      6 big CL luggage  
     'one small book'                        'six big pieces of luggage' T'ung & 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<sup>0</sup> and CL<sup>0</sup> 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 south east 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 for by suggesting that (11) is a head-initial DP with Dem/D<sup>0</sup> and Num-CL<sup>0</sup> 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 favour 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<sup>0</sup> is located above Num<sup>0</sup>. 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<sup>0</sup> 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/other quantifier (see Cheng & 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, as 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) cannot be not simply base-generated, it has to be assumed that it results from certain movement. Given that it is the ordering of the elements Num and CL with respect to N which 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, i.e. [N t<sub>i</sub> Num CL<sub>i</sub> ], or that [N Num CL] results from lowering of the Num to a position below CL, i.e. [N Num<sub>i</sub> CL t<sub>i</sub> ]. Neither of these possibilities however seems plausible - 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 which undergoes movement, raising leftwards 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<sup>0</sup> 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 actually movement of the entire NP rather than just the N<sup>0</sup>, as represented in (12):

- (12) [DP [NP dek naa-rak]<sub>i</sub> [NumP soong [CLP khon t<sub>i</sub> ]]]  
 child loveable 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 only be a single direction of selection within a language, as in (13):

- (13) [DP [NP dek naa-rak] nii ]  
 child loveable 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<sup>0</sup> head position and Thai is head-initial, it has to be assumed that the NP has been moved leftwards from an underlying complement position to the right of D<sup>0</sup>, i.e. [DP [NP ]<sub>i</sub> Dem t<sub>i</sub> ]. 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<sup>0</sup>, i.e. [[NP ]<sub>i</sub> [DP Dem [D ] t<sub>i</sub> ]].

The assumption that the demonstrative occurs in DP-final position due to leftward movement of its complement in Thai is one which incidentally also has to be made in other languages of south east 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 above with Thai therefore lead to the same assumption that the complement of the D<sup>0</sup> is moved leftwards leaving demonstratives in surface final position in the DP. What is significant to note about Hmong, Vietnamese and Indonesian is that unlike 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’

As the sole exception to this headedness is the position of the demonstrative, the conclusion that its complement moves leftwards leaving the demonstrative in DP-final position is rather straightforward to make, and any other attempted analysis faces serious difficulties reconciling the different directions of headedness within the DP that would have to be assumed for D<sup>0</sup> vs Num<sup>0</sup>, CL<sup>0</sup> and N<sup>0</sup>. 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 in underlyingly in Thai):

- (15) 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 above), 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 towards 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.

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) maka adapun mengerjakan [lobang sa buah itu] sampai 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, p317)

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 out that in poetry/literature and in 'inventory forms' this may be converted into [NP Num CL] just as in Indonesian.<sup>1</sup>

The theoretical arguments and empirical support which can be brought together in favour 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 we do this however, we will 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 Thai however, there is a good reason to believe that Num+CL do not in fact modify the NP in any kind of subject-predicate structure. Higginbotham (1987) points out that the possibility for numerals

<sup>1</sup> See also Bhattacharya (1999) on Bangla for other clear evidence of optional NP-movement inside the DP.

and quantifiers to occur as predicates is critically restricted and 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:

(20) 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:

(20) \*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/NP before numeral (and classifier) are found to occur particularly often in written list or 'inventory' forms and 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/food-ordering type situations (as for example in Indonesian (21) below) or alternatively it is only attested in such situations. This sequencing of noun/NP before numeral is furthermore observed (J. Hurford p.c. 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 (22):

(21) saya mau membeli beras dua kilo  
I want buy rice 2 kilo  
'I want to buy two kilos of rice.'

(22) 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 can be found to 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 favoured in the situations noted. A possible answer here is that sequencing the noun/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/NP before Num CL) presents the information in a sequence which 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/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 topicalisation at the sentential level (as in fact hinted at in Greenberg 1975) - in both CP and DP, nominal elements which 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, p.413)):

- (23) you da jiang er ren. Er jiang...  
 be big general 2 CL 2 general..  
 ‘There were two great generals. The two generals...’

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, i.e. [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 post-nominal 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 post-nominal Num-CL sequences

is developed in Fukushima (1991) and Ishii (2000) for modern Japanese, having originally been suggested in Greenberg (1975), and is supported by a variety of evidence. First of all 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 (24) and classical Chinese (25):

(24) mii nisit maa haa khun soong khon  
 be student come find you two CL  
 ‘Two students came to look for you.’

(25) 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)

Secondly, 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 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 (26) noted in Bisang (forthcoming), where the ‘classifier’ is simply a repetition of the first noun:

(26) [NP ren] [NP shi-you-wu ren]  
 person 10-and-5 person  
 ‘15 people

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

(27) Shoonin-ga kinoo [san-nin katsu tashika-ni] sono jiko-o mokugekishi-ta  
 witness-NOM yesterday 3-CL and certainly that accident-ACC witness-PAST  
 ‘Three witnesses certainly witnessed that accident yesterday.’

Finally, 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 following the NP *pisutoru* ‘pistol’ in (28), it cannot be positioned before the NP, even though Japanese otherwise regularly allows the pre-nominal sequence [Num CL no NP]:

(28) a. 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)  
 b. \*Taroo-ga ippatsu-no-pisutoru kinoo utta  
 Taroo-NOM one-CL-GEN-pistol-ACC yesterday shot

The adverb theory of post-nominal 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 re-interpreted 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 re-analyse [NP Num CL] as a single DP will analyse the initial NP as undergoing movement to its surface position from a natural complement position following CL, and will interpret this movement 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 pre-reanalysis 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 re-analysis of a rather different two-NP structure.

If such a re-analysis 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 re-analysis has indeed taken place, as it could be argued as in Fukushima (1991) and Greenberg (1975) that post-nominal Num-CL sequences are in fact still just adverbs and not part of the DP. In what remains of this section, the paper will show that there is indeed evidence of this kind, as well as noting a certain interfering complication.

First of all, 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 final in the DP following Num and CL:

- (29) [DP dek saam khon nan] keng  
 child 3 CL Dem clever  
 ‘Those three children are smart.’

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:

- (30) canaw [DPsaouq hna ouq]-ko weh hta teh  
 I book 2 CL ACC buy Asp NON-FUTURE  
 ‘I bought two books.’
- (31) U-Win-Win [PP [DP meitswee thoun yauq] ne] Yangoun thwaa teh  
 U-Win-Win friend 3 CL with Rangoon go NON-FUTURE  
 ‘U-Win-Win went to Rangoon with three friends.’

Thirdly, there are certain aspects of the interpretation of [NP Num CL] sequences which suggest that the Num and CL are DP-internal elements. Consider the clear contrast in interpretation between Thai (32) where Num CL immediately follow the NP, and (33) where the Num and CL occur sentence-finally in an adverbial type position. Whereas (33) has a partitive type interpretation, in (32) such an interpretation is significantly not available:

- (32) dek saam khon sia chiwit laew  
 child 3 CL lose life ASP  
 ‘The three children died already.’

- (33) 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 (41a) and are blocked when numerals occur under the scope of  $D^0$  inside the DP as in (34b):

- (34) a. [<sub>QP</sub> Three of [<sub>DP</sub> the children]].  
 b. [<sub>DP</sub> The three children]

In Thai (33) the Num and CL are clearly in a DP-external position and the result is a partitive reading. In (32), 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 (32) would therefore seem to be that the Num and CL in such cases are indeed inside the DP, and this therefore blocks the possibility of a partitive interpretation. Note furthermore that the only possible positioning of Num and CL relative to an overt demonstrative is before the demonstrative as in (29) and that attempting to place Num and CL after a demonstrative as in (35) is simply ungrammatical, indicating that post-NP Num and CL in preverbal subject position must indeed be DP-internal:

- (35) \*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. Note that (36) below with the sequence [Dem NP Num CL] only has a non-partitive interpretation. This once again clearly suggests that the Num and CL elements are DP-internal and under the scope of the  $D^0$ :

- (36) canaw eh-dii saouq hna ouq weh hta teh  
 I Dem book 2 CL buy Asp NON-FUTURE  
 ‘I bought those two books/NOT:\*I bought two of those books.’

There is consequently 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 (36) 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:

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

Here Japanese and Burmese might seem to be significantly different, suggesting perhaps that post-nominal 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 (30) that case-markers can occur following the Num CL sequence, it is actually also possible for a case-marker to occur between the NP and Num CL as in (38). 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 (44), suggesting that it is also still possible for Num CL in Burmese to occur and be interpreted outside of the DP:

- (38) U-Win-Win eh-dii saouq-**ko** hna ouq weh teh  
 U-Win-Win Dem book ACC 2 CL buy NON-FUTURE  
 ‘U-Win-Win bought two of those books.’

Consequently while (29-33) allow for the conclusion that adverbial Num CL has indeed allowed for reanalysis in both Burmese and in Thai, examples such as (33) and (38) suggest that it nevertheless may still be possible for Num CL to occur unreanalysed 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 closing this section, one last pattern which can usefully be considered here is the historical development of modern day [NP Num CL] forms in Khmer. 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 which is now the sole modern form, as schematised in (39) below:

- |      |                  |                     |                     |
|------|------------------|---------------------|---------------------|
| (39) | <u>Old Khmer</u> | <u>Middle Khmer</u> | <u>Modern Khmer</u> |
| 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 grammaticalised as a distinct special category. NP CL Num forms were therefore simply sequences of two NPs and Num was attached to the right of a second adverbial NP in the same rightward position that it otherwise attaches in pattern (a). [NP CL Num] is therefore really [NP<sub>1</sub>] [NP<sub>2</sub> Num] essentially in line with Greenberg (1975) and proposals earlier above. Later on in middle and modern Khmer it can be suggested that there occurred natural grammaticalisation of the classifier as a nominal functional category and two-NP structures became re-analysed as single DPs with Num and ‘CL’ internal to the DP. Such a hypothetical grammaticalisation 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 replaced pattern (b) with pattern (c). What such a change seems to suggest is that when grammaticalisation and re-analysis occurred, this significantly forced Num and CL as functional heads within the DP to be re-aligned 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 grammaticalise 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 grammaticalised in a specifically head-initial way reinforces the

view that grammaticalisation and reanalysis does not occur in any random fashion and that there are clear principles of headedness governing the organisation 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 in south east 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 paper now moves on to consider other aspects of the structure of DPs in south east Asian languages and suggests how  $X^0$ -movement and grammaticalisation may also in certain cases 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 south east Asian area, there is an interesting classifier pattern which 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 (47-49) from Vietnamese, Hmong and Nung:

- (40) Nguoi chong rat tot (Vietnamese)  
 CL husband very good  
 ‘The husband was very good.’ (Daley 1998)
- (41) Tus tsov tshaib tshaib plab (Hmong)  
 CL tiger hungry hungry stomach  
 ‘The tiger was very hungry.’ (Jaisser 1987)
- (42) leo tu me da tu po va.. (Nung)  
 then CL wife scold CL husband say  
 ‘Then the wife scolded the husband and said...’ (Saul & Wilson 1980)

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.

Considering the general patterns found in Vietnamese, Hmong and Nung, an important observation which has often been made 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 (40-42) (see Daley (1998), L bel (1996), Bisang (forthcoming) and Nguyen (1997)). Cheng & Sybesma (1999) also note and investigate similar patterns in Cantonese and suggest that classifiers in Cantonese can be assumed to be inherently definite elements like determiners in other languages. As classifiers are however assumed to occur in  $CL^0$ , it is argued that bare classifier nominal expressions with definite interpretations in Cantonese are simply CLPs and do not project any higher functional structure such as NumP or DP. Turning to examine the south east Asian languages now the paper will explore a slightly different approach and suggest that bare classifier structures may in fact be regular DPs resulting from the raising of the classifier up to  $D^0$ .

A critical aspect of the general classifier patterning which needs to be accommodated in any analysis is the fact that although bare [CL NP] forms may naturally be interpreted as being definite,



The suggestion that CL may (sometimes) move to D and cause definite interpretations of the DP will certainly account for the basic patterns observed above, and would also seem to be a reasonable way of explaining the ambivalent nature of CL, classifiers sometimes being associated with definite interpretations and sometimes not. However, in order to be fully convinced of the plausibility of a CL-to-D approach one might hope to find further empirical evidence of the higher D position. Such evidence significantly 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:

- (46) 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.' (Nguyen 1997)
- (47) cai chiec ban nay  
 CL CL table Dem  
 'this table' (Nguyen 1997)

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 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 for it to be re-analysed as (potentially) just a  $D^0$ -element permitting simple insertion into  $D^0$  and allowing for  $CL^0$  to be lexicalised and instantiated by a second classifier as in (46-47). In Simpson (1998), Wu (2000), Simpson & Wu (2000) and also Roberts & Roussou (1999) it is suggested that grammaticalisation 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 re-analysis 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 (48), the re-analysis of *ge* in  $D^0$  now allows for the  $CL^0$  position to be instantiated by a new classifier:<sup>3</sup>

- (48) 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 (46-47). 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

---

of a DP-shell structure as suggested in Simpson & Wu (2000) to account for languages such as Spanish which allow the clear co-occurrence of determiners and demonstratives. In such a DP-shell both the determiner (or classifier) and the demonstrative may be generated in  $D^0$  positions.

<sup>3</sup> The fact that classifiers may be re-analysed as either definite or indefinite determiners would seem to indicate again that such elements are not inherently definite or indefinite but come to be associated with either definiteness or indefiniteness due to their frequent use in certain constructions.

CL-to-D to give rise to the definite interpretation of a DP as suggested.<sup>4</sup> Consequently then, either CL-to-D, or demonstrative insertion into D<sup>0</sup>/SpecDP or general classifier insertion into D<sup>0</sup> (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, we will now consider another revealing 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 (49). The number one however is commonly placed *following* the classifier in Thai and can *only* occur in DP-*final* position in Nung (50a/b):

- |      |    |                  |      |      |        |    |                                     |       |       |        |
|------|----|------------------|------|------|--------|----|-------------------------------------|-------|-------|--------|
| (49) | a. | dek              | saam | khon | (Thai) | b. | slam                                | ahn   | vet   | (Nung) |
|      |    | child            | 3    | CL   |        |    | 3                                   | CL    | spoon |        |
|      |    | 'three children' |      |      |        |    | 'three spoons' (Saul & Wilson 1980) |       |       |        |
| (50) | a. | dek              | khon | nung | (Thai) | b. | ahn                                 | tahng | nuhng | (Nung) |
|      |    | child            | CL   | one  |        |    | CL                                  | chair | one   |        |
|      |    | 'one/a child'    |      |      |        |    | 'one/a chair' (Saul & Wilson 1980)  |       |       |        |

Considering the placement of 'one' either before 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 (re-)positioning of the number 'one' is that 'one' is now commonly placed in the DP-final position that demonstratives otherwise occur in, as in (51). The Thai example (52) shows that 'one' in this final position is indeed in complementary distribution with demonstratives, though it may nevertheless occur in the regular pre-CL Num<sup>0</sup> position with a demonstrative following CL:

- |      |    |                        |      |      |     |        |    |   |      |       |     |        |
|------|----|------------------------|------|------|-----|--------|----|---|------|-------|-----|--------|
| (51) | a. | dek                    | song | khon | nii | (Thai) | b. | slong                                     | ohng | dehk  | te  | (Nung) |
|      |    | child                  | 2    | CL   | Dem |        |    | 2   | CL   | child | Dem |        |
|      |    | 'these three children' |      |      |     |        |    | 'those two children' (Saul & Wilson 1980) |      |       |     |        |
| (52) | a. | *dek                   | khon | nung | nii |        | b. | dek                                       | nung | khon  | nii |        |
|      |    | child                  | CL   | one  | Dem |        |    | 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 which contrasts in its indefinite specification with the definiteness encoded by demonstratives.

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 we have argued at length for the occurrence of NP-movement over Num and CL in Thai, i.e. [NP<sub>i</sub> Num CL t<sub>i</sub>]. When a demonstrative occurs in DP-final position in Vietnamese, Nung, Indonesian, Thai and Khmer we have also suggested that the elements found to the left of the demonstrative are raised to this position as a single constituent from an underlying position to the

<sup>4</sup> Another possibility is that individual speakers in Vietnamese consistently use different strategies. Nguyen (1997) notes that the classifier-doubling strategy is found more in older generation speakers.

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 leftwards movement involved in the derivation, as illustrated in (53). First of all the NP will raise over Num and CL to a position between these elements and Dem, and then the  $[\text{NP Num CL}]$  constituent will raise over Dem itself (as in Vietnamese, Nung etc), resulting in the surface linearisation of examples such as (51):<sup>5 6</sup>

- (53) underlying structure:  $[\text{Dem Num CL NP}]$   
 NP movement:  $[\text{Dem NP}_i \text{Num CL } t_i]$   
 movement over Dem:  $[[\text{NP}_i \text{Num CL } t_i]_k \text{Dem } t_k]$

Concerning the occurrence of the number ‘one’ in the Dem position, it can be suggested that ‘one’ is first base-generated in  $\text{Num}^0$  and then undergoes head-raising up to  $D^0$ . Note that it is not possible to have any lexicalisation of the Num position when ‘one’ is in the Dem position, suggesting that ‘one’ has indeed raised up from  $\text{Num}^0$  (and semantically there is no reason why numbers should not co-occur with an indefinite determiner, as in Chinese (48)):

- (54) \*dek saam/nung khon nung  
 child 3/one CL ‘one’

The derivation which results in ‘one’ occurring in  $D^0$  in Thai can therefore be suggested to be as schematised in (55) below with ‘one’ originating in  $\text{Num}^0$ . ‘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, 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:

- (55) underlying structure:  $[\text{‘one’ CL NP}]$   
 ‘one’ raises from  $\text{Num}^0$  to  $D^0$ :  $[\text{one}_i [ t_i \text{ CL NP } ]]$   
 NP movement over Num CL:  $[\text{one}_i [ \text{NP}_k t_i \text{ CL } t_k ]]$   
 movement over the Dem position:  $[\text{NP}_k t_i \text{ CL } t_k]_m [\text{one}_i t_m]$   
 (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 of all, the patterns can be argued to provide good support for the general underlying structures and movement suggested to occur in Thai, Khmer and other 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  $\text{Num}^0$  (thus accounting for the fact that  $\text{Num}^0$  may not be independently lexicalised when ‘one’ occurs in  $D^0$ ), this indicates again strongly that Thai DPs with the surface order  $[\text{N Num CL D}]$  ( $D=\text{Dem}$ ) cannot in fact

<sup>5</sup> Note that if no NP over Num CL raising occurs then the surface order is that of Vietnamese, Nung, and Indonesian, and if NP-movement but no movement over Dem occurs, the order is that found in Burmese.

<sup>6</sup> With regard to the two applications of leftward XP-movement posited here, if these are both taken to be DP-internal topicalisation-type movements, this may suggest that there are DP-internal topic-like positions both below and above Dem/ $D^0$  in a way similar to Rizzi’s (1997) suggestion that there are iterated topic positions both below and above the Focus position within the clause.

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<sup>0</sup> to D<sup>0</sup> should be straightforwardly blocked by the intervening CL<sup>0</sup> head and the HMC. The fact that raising from Num<sup>0</sup> to D<sup>0</sup> does seem to occur can arguably *only* be accounted for if Num<sup>0</sup> occurs higher than CL<sup>0</sup>, which itself then entails that the N(P) must be assumed to have undergone leftward raising from some other position lower than CL<sup>0</sup>.

The patterning here also provides further good support for the general assumption that Num<sup>0</sup> and CL<sup>0</sup> 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<sup>0</sup>) position, this suggests that classifiers are not simply suffixes attached to numerals in a single head, as affixes/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.

Thirdly, the fact that ‘one’ in Thai and Nung is becoming an indefinite determiner and is targeting the same position as definite D<sup>0</sup> elements (demonstratives) strongly supports the assumption that both definiteness and indefiniteness are encoded in the same position D<sup>0</sup> in DPs, and hence that indefinite nominal expressions are indeed DPs (at least in the Tai languages). For a number of years there has been frequent discussion in the literature about whether indefinite nominal expressions might possibly be constituents which are smaller than full DPs, being NPs or alternatively NumPs, however often it is difficult to find conclusive empirical evidence in favour of either a DP or an NP/NumP analysis. Here though in Thai and Nung because of the interesting surface linearisation of the DP we find that the D<sup>0</sup> position is actually not adjacent to the NP or the NumP in Thai. Consequently it is possible to see the number one developing as an indefinite determiner in a position 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 most definitely can be full DPs and the same syntactic size as their definite counterparts. Furthermore, if the indefinite nature of *some* nominal expressions is expressed overtly in the D<sup>0</sup> position, it might be natural to assume that the D<sup>0</sup> 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 (49), for example.<sup>7</sup> 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, we will 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<sup>0</sup> and Num<sup>0</sup>.

## 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 of all, in a wide range of languages it is found that certain *particular* nouns are commonly quantified without any apparent classifier. Frequently the subset of nouns which pattern in

---

<sup>7</sup> Note that it is possible that the number ‘one’ in English also raises to D<sup>0</sup>. In contrast to other numerals, ‘one’ cannot co-occur with other instantiations of D<sup>0</sup> unless it is interpreted like the adjective ‘unique’ and is followed by a relative clause as in (iii):

- (i) The two/three/\*one man left.
- (ii) These two/three books are mine/\*/?this one book is mine.
- (iii) The unique/one person \*(I like amongst them) is David.

this way is very similar and includes words for units of time e.g. ‘year’, ‘day’, ‘time’ etc, sometimes the word for ‘person’ and certain other terms used with fairly high frequency as illustrated in (56):

- (56) a. tu nam                      b. ba lan      (Vietnamese)  
      4 year                        3 time  
      ‘four years’                ‘three times’

The second way that classifier-less forms manifest themselves is not lexically restricted in the above way and it is noted that in certain languages (e.g. colloquial Minangkabau and Indonesian) the use of a classifier appears to be quite optional with a fully wide range of nouns, as e.g. in Nung (57):

- (57) slam (tew) kha-lo      (Nung)  
      3      (CL) road  
      ‘three roads’ (Saul & Wilson 1980)

There might seem to be two possible analyses of the patterns here. Considering just the former lexically-restricted type in (56), if the forms in (57) are taken to be genuinely classifier-less and without any CL position, this might be suggested to indicate that there is a significant distinction amongst nouns in classifier languages and that some nouns simply require no individuation. However, such an approach would not seem to be able to generalise further to cover the second type of classifier-less forms in (57) 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/NP in such constructions. However, this is somewhat unlikely as the non-occurrence of an N-position and an NP-complement to CL would entail that there are classifiers which classify/individuate nothing. As classifiers are essentially by definition functions which apply to some second complement term, the presence of a classifier would therefore seem to require the presence of an N/NP. Furthermore, such an approach will again not generalise to cover the second type in (57) as the overt elements here are clearly nouns.

In addition to the above two possible analyses of cases such as (56) there is also a third potential analysis of these patterns suggested by the phenomena examined in the sections 5 and 6. If the analysis of Thai/Nung ‘one’ moving from Num to D and Vietnamese CL moving from CL to D is correct, it could be suggested that there is also similar movement taking place in instances such as (56), with the elements ‘year/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/Nung, and 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 favour of one hypothesis rather than the other. However, Thai and Khmer with their critically different surface linearisation of DP structure do provide a clear and useful means to check the above hypotheses as in these languages N and CL are not linearly adjacent in surface forms, but commonly separated by numerals in Num (i.e. in the order [N(P) Num CL]). As the Thai data in (58) show, elements such as ‘year’ and ‘time’ in fact 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:

- (58) a. soong pii                      b. saam khrang                      c. sii khon  
       2 year                              3 time                              4 person  
       'two years'                        'three years'                        'four people'

This indicates that the first possible analysis that no CL position occurs in 'classifier-less' 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 above and 'year' and 'person' etc also do appear very nominal and likely to be N-heads, it would seem that the most likely explanation of the forms in (56) and (57) is indeed the third possibility that the patterns found in (56) and (57) 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 (56/57) and more regular forms such as for example (49)) 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 (56) 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/classifiers for themselves 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 below:

- (59) a. hoong saam hoong (Thai)                      b. cun ta cun (Burmese)  
       room 3 room/CL                                      island one island/CL  
       'three rooms'    'one island'

The idea of an element being used to classify/individuate itself is consequently both plausible and commonly attested. If one assumes the 'copy theory' of movement (Chomsky 1995), it might also be suggested that in these repeater cases there may again quite possibly be movement between N and CL, with the difference between repeater and non-repeater cases such as (56) being that in (59) the copy left in N by movement to CL is not deleted and is actually spelt-out phonetically.

Finally, unlike the other two possible analyses of (56) considered above, the N-raising analysis of cases such as (56) will importantly also extend in a natural way to cover the second type of lexically unrestricted classifier-less Ns (i.e. (57)), 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 lexicalise 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.<sup>8</sup> Furthermore, it can be suggested that if such N-to-CL movement does occur regularly and if the 'movement and reanalysis' approach to grammaticalisation is correct, one might expect that the frequent association of certain 'N's with the CL position due to regular N-to-CL raising might ultimately allow for the reanalysis of such 'N's as simple direct instantiations of CL, base-generated

---

<sup>8</sup> In this sense classifiers pattern a little like expletives – when an expletive such as 'there' is selected in a numeration, it will be inserted into SpecIP, but when no such expletive is selected, the subject of a clause has to raise to instantiate the SpecIP position.

in CL *without* having first been raised there from N<sup>0</sup> (just as Vietnamese *cai* (46-47) and Mandarin *ge* (48) both originally CLs can now be base-generated in D). As noted earlier, it is well-observed that classifiers do in fact very frequently develop from elements that were originally nouns. The assumption of N-to-CL movement together with a movement and re-analysis view of grammaticalisation 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 relexicalisation of the lower N position with new nominal elements.

## 8. Concluding remarks

Concluding now, this paper has been concerned with attempting to understand how apparent variation in DP structure may be accounted for across a broad range of south east Asian languages. In the first major part of the paper, sections 2-4 concluded that significant distortion of underlying DP structure is often caused by XP-movement inside the DP (i.e. NP-fronting). The second half of the paper in sections 5-7 has now considered other aspects of DP-internal variation, and has argued that in each case the patterns result from simple X<sup>0</sup>-movement and the common raising of DP-internal heads to lexicalise and instantiate other higher head positions, a process which naturally lead to reanalysis and full grammaticalisation of a head in a higher DP-internal position. In both parts of the paper, variation has therefore essentially been attributed to elements being in different stages of ongoing historical development and reanalysis, with both XP-movement and X<sup>0</sup>-movement becoming regularised, reanalysed and grammaticalised. Throughout the paper we have also 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 south east Asian languages is far from random, but in fact arguably seem to indicate the opposite conclusion and suggest that DPs in the variety of languages examined here may actually share in common a single basic highly regular underlying structure.<sup>9</sup>

## References

- Aikhenvald, Alexandra (2000) *Classifiers*. Oxford University Press.
- Bhattacharya, Tanmoy. (1999) *The structure of the Bangla DP*. PhD dissertation, University College London.
- Bisang, Walter. (1993) Classifiers, Quantifiers and Class Nouns in Hmong. *Studies in Language* 17:1-51.
- Bisang, Walter. (forthcoming) Classifiers in East and South East Asian languages: Counting and Beyond. in J. Gvozdanovic (ed.), *Changes in Numeral Systems*. Mouton de Gruyter, Berlin.
- Cheng, Lisa & Rint Sybesma. (1999) Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistics Inquiry* 30: 509-542.
- Chomsky, Noam (1995). *The Minimalist Program*. MIT Press, Cambridge, Mass.
- Cinque, Guglielmo. (1996) The 'Antisymmetric' Programme: Theoretical and Typological Implications. *Journal of Linguistics* 32:447-464.
- Daley, Karen. (1998) *Vietnamesenamese Classifiers in Narrative Texts*. Summer Institute of Linguistics, University of Texas at Arlington.

---

<sup>9</sup> See also Cinque (1996) who suggests the possibility of such a conclusion universally for DPs, and Tang (1999) who assumes such a conclusion in a rather different feature-based approach to DP-structure variation.

- Fukushima, Kazuhiko. (1991) Generalized floating quantifiers. PhD dissertation, U. of Arizona.
- Gil, David. (1994) Summary: Numeral Classifiers. *Linguist List* Vol 5: 466
- Greenberg, Joseph. (1975) Dynamic Aspects of Word Order in the Numeral Classifier. in C. Li (ed.) *Word Order and Word Order Change*, 27-46, U. of Texas Press, Austin.
- Higginbotham, James. (1987) Indefiniteness and predication. in *The representation of (in-)definiteness*, ed. E. Reuland & A. ter Meulen, 43-70. Cambridge, Mass, MIT Press.
- Hopper, Paul. (1986) Some Discourse Functions of Classifiers in Malay. in *Noun Classes and Categorization.*, (ed.) C.G. Craig, pp 309-25, John Benjamins, Amsterdam.
- Ishii, Yasuo. (2000) Floating quantifiers in Japanese: What are they and what do they tell us? talk presented in SOAS, University of London.
- Jaisser, Annie. (1987) Hmong Classifiers. *Linguistics of the Tibeto-Burman Area* 10:169-75.
- Kawashima, Ruriko. (1993) The Structure of NPs: Arguments for Quantifier Phrase and Number Phrase. *Cornell Working Papers in Linguistics* 11: 56-72.
- Loebel, Elisabeth. (1996) Klassifikatoren. Eine Fallstudie am Beispiel des Vietnamesischen. Universitaet K In: Habilitationsschrift. Ms.
- Muromatsu, Keiko. (1998) On the syntax of classifiers. PhD dissertation, U. of Maryland.
- Nguyen, Dinh Hoa. (1957) Classifiers in Vietnamese. *Word* 13:124-52.
- Nguyen, Dinh Hoa. (1997) *Vietnamese*. John Benjamins, Amsterdam.
- Pan, Hai-Hua. (1990) Head noun movements in Chinese, the ECP and minimality. Ms. U. of Texas at Austin.
- Pulleyblank, Edwin. (1995) *Outline of classical Chinese grammar*. Vancouver: UBC Press.
- Rizzi, L. (1997). The fine structure of the left periphery. *Elements of grammar*, ed. L. Haegeman, 281-337. Dordrecht: Kluwer.
- Roberts, Ian & Anna Roussou. (1999) A Formal Approach to “Grammaticalisation”. Ms, University of Stuttgart and University of Cyprus.
- Saul, J.E. & N.F. Wilson. (1980) *Nung Grammar*. Summer Institute of Linguistics, U. of Texas at Arlington.
- Schafer, Edward. (1948) Noun Classifiers in Classical Chinese. *Language* 24:408-13.
- Simpson, Andrew. (1998) Empty Determiners and Nominalisation in Chinese, Korean and Japanese. paper presented at the USC Symposium on E. Asian languages, Nov. 1998.
- Simpson, Andrew & Zoe Wu. (2000) Focus, Shells and Agreement. paper presented at WCCFL 19, UCLA, January 2000.
- Tang, C.-C. Jane. (1990) Chinese phrase structure and the extended X'-theory. PhD dissertation, Cornell, Ithaca.
- Tang, Sze-Wing. (1999) Some Speculations about the Syntax of Noun Phrases. in F. del Gobbo & H. Hoshi (eds.) *UCI Working Papers in Linguistics* 5:135-154.
- T'ung, P.C. & D.E. Pollard (1982) *Colloquial Chinese*. Routledge, London.
- Watters, J. (1981) *A phonology and morphology of Ejagham – with notes on dialect variation*. PhD dissertation, UCLA.
- Wu, Zoe. (2000) The development of functional categories in Mandarin Chinese. PhD dissertation, USC.