

**UC Berkeley**  
**Dissertations, Department of Linguistics**

**Title**

A Grammatical Sketch of Eastern Kayah (Red Karen)

**Permalink**

<https://escholarship.org/uc/item/5fk0z015>

**Author**

Solnit, David

**Publication Date**

1986

A Grammatical Sketch of Eastern Kayah (Red Karen)

By

David Benedict Solnit  
A.B. (University of California) 1977  
M.A. (University of California) 1979  
C.Phil. (University of California) 1982

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Linguistics

in the

GRADUATE DIVISION

OF THE

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:.....*James A. Trotter*.....*October 14, 1986*  
Chairman Date  
.....*Charles J. Fillmore*.....*October 29, 1986*  
.....*Ken Chang*.....*Oct 27, 1986*

**DOCTORAL DEGREE CONFERRED**  
**DECEMBER 16, 1986**

.....

**A Grammatical Sketch of Eastern Kayah (Red Karen)**

**Copyright © 1986**

**David B. Solnit**

## A Grammatical Sketch of Eastern Kayah (Red Karen)

David B. Solnit

### Abstract

This dissertation describes the Eastern dialect of Kayah (also known as Red Karen), a language of the Karen group of the Tibeto-Burman branch of the Sino-Tibetan linguistic stock. Kayah is the major language of the Kayah State of Burma, and is also spoken in a small area of Mae Hong Son province in northwestern Thailand, where the data for this grammatical sketch was recorded.

Kayah is tonal, monosyllabic (with familiar qualifications), and lacks affixational morphology except in relic form. Compounding, however, is extensive, involving both nouns and verbs.

After brief descriptions of the phonology and the nature of the morpheme and form classes, a fairly detailed description is given of the Verb Complex, a potentially very complicated structure centered around the main verb of the simple clause. Kayah is typical of languages of the mainland Southeast Asia-southern China linguistic area in having verb serialization (also 'verb series', 'verb concatenation'). It is unusual in combining basic SVO typology with extensive use of immediate concatenation of verbs, with no intervening arguments, a trait more typical of the verb-final languages of the area. It is argued that these constructions in Kayah are best analyzed as compounds, formed in the morphology/lexicon, rather than syntactic phrases, whether



base-generated or derived by transformation. The lexical structure of these compound verbs is described in terms of (a version of) current morphological theory.

Other chapters describe clause structure; the syntactic behavior of the NP, PP and Numeral-Classifier construction; sentence types; and an outline of interclausal syntax (nominalized clauses, attributive clauses, and clause sequences).

James A. Matisoff

### Acknowledgements

In various capacities, from my first guide to the linguistic topography of Southeast Asia to chair of my dissertation committee, from the most basic conceptual discussion to the smallest detail (including proposed scatalogical epigraphs), Jim Matisoff has been a constant source of encouragement and valuable advice. I am also indebted to the other members of my committee, Chuck Fillmore and Kun Chang.

I have also benefited greatly from discussions with friends and colleagues at Berkeley and Cornell, including Paul Kay, Farrell Ackerman, Amy Dahlstrom, Knud Lambrecht, Cathy O'Connor, Wayne Harbert, Gennaro Chierchia, and Julian Wheatley.

Many friends in Thailand deserve credit of various sorts, among them Professor Suriya Ratanakul, William Young, Bamrung Ruangsawat (Luka), Cee Yin, Chaiwut Hongthong, Jacques Landy, Pippa Curwen, and Jennifer Ringstad.

For more humble but no less vital types of fortification I would also like to thank the Chiangmai Tea House, Boon Rawt Brewery, Tung Hoo Coffee, and Peet's of Berkeley.

My research was supported by a U.S. Department of Education Fulbright-Hays Doctoral Dissertation Research Abroad grant (\*B008200435). I am also grateful to the National Research Council of Thailand, and to the governments of Mae Hong Son and Chiangmai Provinces for assistance during my stay in Thailand. I am indebted to my major informant Kāmè for long hours of patient work, and to the people of Thā Médx Lē Khā, Rūsō Lē and Thā Médx Lē Chá, who put me up in their homes, told me stories, and allowed me to tape record hours of conversation and song.

Last but not least I would like to thank my parents, who discerned my interest in linguistics long before I did, and Linda Tiffany for support of every sort, both long-distance and close-up.

## Contents

Abstract	
Acknowledgements	i
0. Introduction	1
0.1 Kayah, Karenni, Karen: language and people	1
0.2 Data and theory	4
0.3 Abbreviations, Conventions, Transcription	7
1. Phonology	13
1.1 Syllable structure	13
1.2 Initials	14
1.3 Rhymes	17
1.4 Tones	20
1.5 Phonology of prefixes	24
1.6 Vowel harmony	26
1.7 Outline of historical tone correspondences	27
1.8 Orthographies	28
2. Morphemes, Word Formation, and Form Classes	29
2.1 Monosyllabicity	29
2.2 Compounding	30
2.3 Affixes	33
2.4 Other alternations	40
2.4.1 Tone change	40
2.4.2 Reduplication	42
2.4.3 Allomorphy	43
2.5 Form classes	44

Morphosyntax, Part I	51
3. Theoretical Preliminaries	52
3.1 Introduction	52
3.2 Semantic roles (general)	56
3.2.1 Preliminary remarks	56
3.2.2 Selected and non-selected roles	58
3.2.3 Dependence between role types	63
3.2.4 Specification by verb; by non-verb	66
3.3 Roles for Kayah	71
3.4 Argument structure	76
3.5 Grammatical relations	81
4. The Verb Complex	88
4.1 Introduction	88
4.1.1 Verb serialization: characteristics and problems	88
4.1.2 Verb serialization in Kayah	95
4.1.3 Outline of analysis of the VC	96
4.1.4 Versatility, another possible criterion for analysis	99
4.2 Verb-Verb constructions	102
4.2.1 Resultative	102
4.2.1.1 General	102
4.2.1.2 Directional	106
4.2.2 Sequential	116
4.2.3 Directive	123
4.2.4 Descriptive	131
4.2.5 Modal	138
4.2.6 Summary of V-V's	142

4.3 Verb/Particle constructions	144
4.3.1 Introduction	144
4.3.2 Pre-verbal particles: the <i>khwe</i> -class	146
4.3.3 Quasi-Modal particles	147
4.3.4 <i>ja</i> , a Sequential particle	147
4.3.5 Post-verbal particles: the <i>ra</i> -class	148
4.3.6 Descriptive particles	158
4.3.6.1 Movable	158
4.3.6.2 General	160
4.3.7 Bound result expressions	162
4.4 Interactions	171
4.4.1 V-V and V-V	171
4.4.2 Particles	178
4.4.3 Excursus: noun-incorporating verbs and particles	180
5. Formalizing the Analysis	183
5.1 Introduction: many propositions, one clause	183
5.1.1 Review of arguments for monoclausality	184
5.1.2 V-V as related events	188
5.1.2.1 Predication vs. modification (1)	188
5.1.2.2 Causation vs. sequence	189
5.1.2.3 Common participant	189

5.1.3 Skewing of syntax and propositional semantics	191
5.1.3.1 The problem of skewing	191
5.1.3.2 Multi-level syntax: transformational and control approaches	193
5.1.3.3 Multi-level lexical structure: lexicalist approach	196
5.1.4 Afterword	197
5.2 Multi-level syntax: transformational approach	199
5.2.1 Syntactic base-generation	200
5.2.2 A transformationalist analysis: verb incorporation	204
5.3 Multilevel lexical units: lexicalist (compounding) analysis	215
5.3.1 Characteristics of compounds vs. phrases	215
5.3.2 Some principles of lexicalist morphology	218
5.3.2.1 Headedness and percolation	218
5.3.2.2 Formation rules or not?	219
5.3.2.3 Argument linking	220
5.3.3 Application of lexicalist morphology	224
5.3.3.1 The proposed analysis	225
5.3.3.2 Some remarks on headedness and argument linking	242
5.3.3.3 Afterword	248
Morphosyntax, Part II	253
6. The Clause and Grammatical Relations	254
6.1 Definition and types	254
6.2 Constituents	255
6.2.1 The VP	255
6.2.2 Selected and non-selected grammatical relations	259

6.3 Subject (and Topic)	265
6.3.1 General	266
6.3.2 In Kayah	269
6.3.3 'Inverted Subjects'	275
6.4 Objects	276
6.5 <i>ne</i> -PP's	283
7. NP, PP, ClfP: Syntax and morpheme classes	296
7.1 NP	296
7.1.1 Lexically-headed NP's	298
7.1.2 Clf-headed NP's	302
7.1.3 Expanded NP's: NP+ClfP	305
7.1.4 Localizers	307
7.2 PP	311
7.3 ClfP	315
7.3.1 Classifiers	315
7.3.2 Quantifiers and counting in Kayah	319
8. Sentence Types and Sentence Particles	326
8.1 Sentence types	326
8.2 Sentence particles	329
8.3 Unique and problematic particles: <i>ke</i> and <i>na</i>	339
8.4 Excursus: questions	344

<b>9. Interclausal Syntax</b>	<b>347</b>
<b>9.1 Non-autonomous nominalized clauses</b>	<b>347</b>
<b>9.1.2 Embedded (non-attributive)</b>	<b>347</b>
<b>9.1.2 Preposed attributive clauses</b>	<b>348</b>
<b>9.2 Postposed attributive clauses</b>	<b>352</b>
<b>9.3 Clause as argument of verb</b>	<b>355</b>
<b>9.3.1 Clausal Objects</b>	<b>355</b>
<b>9.3.2 Clausal Subjects?</b>	<b>356</b>
<b>9.3.3 Quotatives</b>	<b>357</b>
<b>9.4 Clause sequences and <i>ma</i></b>	<b>359</b>
<b>Footnotes</b>	<b>365</b>
<b>References</b>	<b>373</b>



## O. Introduction.

### O.1 Kayah, Karen, Karenni.

The language described in this study is best referred to as Eastern Kayah, indicating that it belongs to the Eastern division of the Kayah group. Kayah in turn is a Central Karen language, and Karen is a major subdivision of the Tibeto-Burman branch of Sino-Tibetan. Kayah is thus very closely related to languages like Padaung and Brè; less closely to other Karen languages like Sgaw, Pho and Pa-O (Taungthu); more distantly to hundreds of languages including Burmese, Tibetan, Lahu, Yi, Lushai, etc., etc.; and most remotely to Chinese.

Eastern Kayah would be more accurately referred to as 'Kayeh', since one of the isoglosses separating the East and West dialect groups is the correspondence East /ɛ/ = West /a/ after the palatal semivowel. The word for 'person, human', also the self-designation of the ethnic group, is thus /kajā/ in the Western group and /kajē/ in the Eastern. I have decided to retain 'Kayah', since it has already some currency in Burma (most saliently, perhaps, in the name of Kayah State) and in the Western anthropological/linguistic literature (especially via the writing of F. K. Lehman). From the linguistic and ethnographic point of view, 'Kayah' is as valid and unitary as most other ethnic groupings; also the people themselves recognize their overall identity. For these reasons it is preferable to use a single term for all subgroups and dialects.

Karen is distinctive among Tibeto-Burman groups in several ways. Its tonal system can be shown to have developed from a two-(possibly three-)tone proto-system, with splits and mergers determined by features of initial consonants. In this it resembles only Lolo-Burmese and possibly Tamang (but the latter is not a major branch of the same degree as Lolo-Burmese or Karen). It resembles no other Tibeto-Burman language in its syntactic typology, which is SVO. In both of these features Karen has affinities with the linguistic stocks to its west, namely Kadai, Miao-Yao and Sinitic; in its SVO typology it also groups with Mon-Khmer. These affinities are surely areal in nature, and only to be expected of the southernmost Tibeto-Burman group (also nearly the easternmost: only Loloish extends further in that direction).

'Central Karen' is essentially a geographic term, which may nevertheless turn out to be a valid unit of linguistic subgrouping as well. It refers to the Karen languages (with the exception of Paku, which is either a dialect of or closely related to Sgaw) spoken in and immediately adjacent to the area now known as Kayah State, formerly Karenni (more fully, 'The Karenni States'; cf. 'Shan States', now Shan State). These include, besides Kayah, languages that have been referred to as Padaung, Brè, Yintale, Palaychi, Mopwa, and many more. Actually the foregoing is more accurately interpreted as a list of designations than of valid, commensurate ethnic or linguistic groupings. The situation in the Central Karen area is complex in the extreme: the ethnolinguistic groupings are in themselves complex, with extensive 'dialect chain' phenomena, and this is complicated

by the nomenclature, which mixes self-designations and exonyms (Matisoff 1986) with abandon. Many of the latter are deictic in nature, with meanings like 'people upstream' or 'westerner', and so naturally change their referents as their users progress upstream, westwards, etc. The reader is directed to Lehman 1979 for some additional discussion of these matters.

The Kayah are numerically the dominant group in the Central Karen area. This area, as its former name 'Karenni states' suggests, has a history of state-level political culture of a sort, and varying autonomy. It has consisted of between three and five states, with the states of Kyebogyi, Bawlakhe, and Kantarawady having the most continuity. Kantarawady comprises the territory east of the Salween, and the closely-similar dialects here referred to as East Kayah could appropriately be identified as the Kantarawady dialect or dialect group.

The variety of Eastern Kayah described here is spoken in several villages just to the south and southeast of Mae Hong Son town in northwestern Thailand; it thus lies on the eastern extreme of the Kayah-speaking area. My division of Kayah into Eastern and Western is really no more than an assumption, and is undoubtedly an oversimplification; it does agree with the distinction used by Eastern speakers themselves (*ʔibe lēbēthe* 'Eastern speech', *ʔibe lēbēnō* 'Western speech'). There are undoubtedly dialectal differences within East Kayah; for instance there is a slight but noticeable difference between the Kayah spoken to the north of Mae Hong Son town and that spoken to the south.

The two dialects are known to the Kayah as *ke' kē* 'upper' and *ke' khu* 'lower' (the language described here is of the 'lower' variety).

## 0.2. Data and Theory.

Data. The language described in this work is that spoken in three villages just to the south of Mae Hong Son town. They are

1. Thā Médx Lē Khā (Thai *khūn hūaj dya* ), at an elevation of about 500 meters on a mountain ( *dɔɔj khūn hūaj dya* ) on the right bank of the Paaj river;
2. Rūsō Lē 'rotten snake creek' (Shan *hɔj sə thaw*, Thai *hūaj syā thāw* 'old tiger creek'), about two-thirds of the way down the mountain;
3. Thā Médx Lē Chá (Shan *waan hɔj ləə*, Thai *hūaj dya* ), on the opposite bank of the river.

Thā Médx Lē Khā is the oldest; people from it later founded Rūsō Lē, and still later some moved farther down to Thā Médx Lē Chá, which I believe was originally a Shan village. In any case the Kayah of Thā Médx Lē Chá live mixed with Shans, and most or all of them speak at least some Shan.

The main informant was 25 years old, a native speaker of Kayah. Her home is in Huai Dya and in addition to Kayah she speaks Shan and Standard Thai, the latter learned during attendance at a government elementary school in Chiang Rai province.

The data was gathered during the period February 1983-March 1984; it consists of about 350 pages of transcribed texts and perhaps 200 pages of notes.

Theory. This study is 'data-oriented'; that is, it makes use of theory as a means of getting at an understanding of Kayah grammar, and as a framework for representing that understanding. Since I do not stick to any one particular theory, my approach might be considered to be of the sort that has been labeled 'atheoretical'; however I do not feel that to be a very accurate label. It is doubtful that it is possible at all to be without a theory, strictly speaking: even the most basic decisions, such as what counts as 'data', and which portions of the data are relevant to which other portions, are nothing if not theoretical decisions.

My approach might be better termed pantheoretical. I hope I have not simply taken a bit from this theory and a bit from that theory, rather I have operated with the feeling that much of what is presented as conflicting views is underlain by a great deal of agreement on basic concepts. I have in mind particularly all theoretical views that have room for the notion that there is some significant aspect of human speech that is best described in terms of abstract formal structure. There are of course approaches that exclude even that moderate a stance, but they have not appealed to me as a grammar-writer.

If one admits the usefulness of formalism, one may still be accused of getting things backwards: clearly a significant proportion of current linguistic research is based on the assumption that the task of linguistics is to discover the best theoretical formalism, in which task data is (only) a means to an end. I may be accused in this study of rather

using theory as a means to the end of description. I plead guilty, but I have not seen anything to convince me that there is any contradiction; to use the words of Chomsky (1982, p. 19), I do not see how one can find out about 'the arrangement of data in the environment' without also contributing to knowledge of 'the nature of the human mind', and vice versa.

This study may be considered an experiment in what happens when one tries to describe the 'whole' grammar of a language without either ignoring the past 25 years of linguistic research or constructing a new theory of one's own that will set it all right. One thing that happens is that it is difficult to keep everything sketchy. Thus the 'sketch portion of this study is made up of Chapters 1-2 and 6-9; Chapters 3-5 are an investigation of multi-verb constructions that goes into some detail of both data and formal interpretation.

There is another sense in which this work is a sketch: while recognizing that delineating an abstract, formal grammar is essential to the description of a language, it is not the end of it. There are also non-abstract, and ultimately non-linguistic factors at work whose effects can be seen at every level of linguistic structure. From that point of view this sketch is the first step in a description of Kayah: the second step awaits an investigation of the discourse, pragmatic and social factors bearing on Kayah grammar, and it is to be anticipated that many points will need to be restated as a result of such an investigation.

### 0.3. Abbreviations, Conventions, Transcription

#### Conventions in Glosses.

Example sentences are cited as in the following example:

- (12) ʔa chuí ʔú            lō    pī chæə    vē te né mī nā pu  
       3 burn smolder use-up complete mine Nε fire two clf  
       She burned up two of mine [blankets] (with fire). (272.3)

The form is: first line: example number, Kayah sentence; second line, morpheme-by-morpheme gloss; third line, free translation followed by a formula indicating the example's source. The example numbering (in parentheses) starts over with each chapter. In interlinear glosses, certain conventions will be followed.

1. Compound words will generally not be analyzed; e.g. *bé se pɔ* will be glossed simply as 'eye' without identifying the components *bé se* 'face' and *pɔ* 'small round object'; similarly *m ca* 'lower-garment + shirt → clothes' is glossed simply 'clothes', *dε sɪ pɔ* 'put + heart → decide' is simply 'decide', and so on.

2. Pronouns are glossed with numerals, plus 's' for singular and 'p' for plural if number is distinguished; thus *vɛ* is '1s' for 'first person singular'. Number is not distinguished in the third person, but there is a special 'obviative' form (see 6.3.2, 7.1), glossed '3OBV'. Gender is not distinguished, so that '3' actually indicates 'he/she/it/they/him/her/them'.

3. Some grammatical morphemes whose meanings are not amenable to a one-word English gloss will be marked by a capitalized version of their phonemic transcription; e.g. *na*, a morpheme with nominalizing, demonstrative and other functions, is glossed as NA.

4. Many-word English glosses of single Kayah morphemes are hyphenated, as *na* 'lower-garment', *de* 'dip-up'(e.g. water).

5. In general, interlinear glosses are intended only to identify morphemes, not to give a fully accurate translation. Thus *na*, strictly meaning any lower garment, may be glossed 'pants'.

In the free translation, items in brackets are supplied to clarify the English, but have no overt equivalents in the Kayah; items in parentheses correspond to elements in the Kayah that are normally not expressed in English. Where English and Kayah items correspond, but with discrepancy in degree of semantic detail, the difference is left for the reader to infer from comparison of morpheme glosses with the free translation; e.g. while the pronoun *na* does not specify gender, it is clear from the context of the above example that the person referred to is female, and this is reflected in the free translation (English has greater detail). Conversely *de* 'cooked rice', *na* 'husked raw rice' and *na* 'unhusked rice' are all rendered as 'rice' in free translation (Kayah has greater detail). Sometimes considerations of space prevent putting these full meanings in the morpheme gloss; thus the three morphemes just mentioned are usually glossed simply 'rice' in the morpheme-by-morpheme gloss as well. It may be pointed out that the



difference will often be evident from context; thus *dī* in *ʔe dī*, glossed 'eat rice', clearly refers to cooked rice.

### Bracketing

When constituent structure seems relevant, the following symbols will be used:

- || . . . || clause boundaries
- | . . . | boundaries of the VC
- | divides NP's, PP, ClfP (NP boundaries not shown when they coincide with clause or VC boundaries)
- [ . . . ] embedded clause (usually relative clause)

example:

ʔa|khé      pò      |thuú| dǎ mi klē    | sō be | bōλ |  
 he shoot additionally bird at forest among 3 clf and-then  
 He shot three more birds in the forest, and then . . .

The bracketing shows a 'flat' structure; in fact there is evidence for more hierarchical structure, as will be discussed, but it is usually derivable from the constituent boundaries as given by the bracketing. The example sentence would be

[<sub>S</sub>[<sub>NP</sub>ʔa] [<sub>VP</sub>[<sub>V</sub>khé pò] [<sub>NP</sub>thuú ]<sub>V</sub>] [<sub>PP</sub>dǎ mi klē ] [<sub>ClfP</sub>sō be]<sub>VP</sub>][<sub>PTC</sub>bōλ ]

Abbreviations:

AMB	ambient noun <i>ke<sup>3</sup> ~ ke<sup>2</sup></i>
CMP	comparative Verb Particle <i>lɔ̄</i> , 'more than'
COM	comitative Verb Particle <i>kā</i>
DCL	declarative Verb Particle <i>wā</i>
DUR	durative Verb Particle <i>pa</i>
IRR	irrealis: Clause Particle <i>pā</i>
NEG	negative: Clause Particle <i>to</i>
NS	new situation: Verb Particle <i>ɿ</i>
OBV	obviative pronoun <i>lū</i>
OS	older sibling
PTC	Verb Particle or Clause/Sentence Particle, with insufficiently-analyzed meaning
sbdy	somebody
stg	something
TRN	transfer of possession: Bound Result Expression <i>pe<sup>3</sup></i>
YS	younger sibling

Sources of examples. The number in parentheses following the free translation indicates the source of the example. Decimal numbers refer to transcribed texts (page number/decimal point/line number); notations with a slash indicate elicited examples (location in my notes identified by date). It has been thought worthwhile to give readers an indication of whether the examples are spontaneous or artificial. Examples with no indication of source are mostly instances of simple, well-supported patterns for which the available examples are unsatisfactory (e.g. contain material that is irrelevant to the point under discussion, but that would require explanation). I have attempted to keep this last type to a minimum.

Transcriptions of non-Kayah languages.

Written Burmese	Okell's (1971) recommended transliteration, but with grave accent (à) for creaky tone; modern spoken pronunciation in IPA
Chinese (Mandarin)	pinyin
Kayoh (Brè Karen)	my notes*
Mien Yao	Downer 1961, with /c ch j/ for /kj khj gj/
Pa-O (Taungthu)	Jones 1961*
Pho Karen	Jones 1961*
Shan	Egerod 1957†
Sgaw Karen	Jones 1961*
Thai (Standard)	Haas†

others are as in cited sources

\*the Karen languages are cited as in the sources except that the tones are converted into the historical categories, essentially as in the works of Haudricourt (see 1.7 for explanation of the categories).

†for Shan and Thai I substitute /p t k/ for Egerod's and Haas' final /b d g/. Note also that the high back unrounded vowel, written /y/ in both; may occur as an off-glide in Shan: thus /lǎy/ 'where' is phonetically [lau<sup>35</sup>].

## 1. Phonology

### 1.1. Syllable Structure

The structure of full (non-prefixial) East Kayah syllables may be represented as follows:

$$\begin{array}{c} T \\ (C_1)(C_2)(G)V \end{array}$$

where T equals a tone;  $C_1$  is any consonant;  $C_2$  is a liquid; G is a glide (w or j); and V is a vowel. The combination of  $C_1 + C_2$  is further referred to as the initial, simple or cluster according to whether the  $C_2$  slot is empty or not. Similarly G + V is known as the rhyme, simple or complex according to whether G =  $\emptyset$  or not. The presence of  $C_2$  entails the presence of  $C_1$ , but it is possible to have G with zero  $C_1$ . That is, the following combinations are possible (T present in all): V, GV, CV, CGV, CCV, CCGV.

1.2. Initials.Simple Initials

	labial	dental	alveopalatal	retroflex	velar	glottal
voiceless unaspirated	p	t	c		k	(ʔ)
voiceless aspirated	ph	th	ch		kh	
voiced	b	d	(j)			
nasal	m	n			ŋ	
voiceless fricative		s				h
voiced continuant	w	l	(j)	r		

## Notes on initials.

1. Aspirated stops and affricate are unit phonemes.
2. /c ch/ are alveopalatal affricates [t<sub>ɕ</sub> t<sub>ɕ</sub><sup>h</sup>]. /ch/ is occasionally realized with no stop component, as a slightly aspirated fricative [ç<sup>h</sup>].
3. /j/ varies between standard palatal glide and voiced palatal fricative, also occasionally appearing as a slightly prenasalized alveopalatal affricate [ʎ<sup>h</sup>], especially in the Low Falling tone.
4. /s/ is an alveolar or dental flat spirant. For some speakers it is strongly dental or even interdental, but this is more characteristic of West Kayah.

5. /ŋ/ has allophone [ŋʰ~ɲ] (fronted velar or palatal nasal) before front vowels and glide /j/. Otherwise it is velar [ŋ].

6. /b d/ are voiced stops, with little or no implosion. They may be prenasalized in connected speech (i.e. intervocally, since all East Kayah words end with a vowel).

7. /w/ is usually labiodental [ɸ], and will be so written henceforth for convenience.

8. /r/ is usually a retroflex approximant similar to Mandarin Chinese /r/. In emphatic speech it may be an alveolar trill. As C<sub>2</sub> in clusters it is largely or completely devoiced by simultaneous aspiration, approaching [s].

9. Two types of free variation seem confined to certain morphemes: [ph~h~ 0] in /phé/'only'(a Verb Particle); and [k~ʔ~0] in /kuklɔ/'head' and /kúkɫē/'swidden'. The former type is also found in [hu~u] 'diminutive suffix', related to /phú/'child', but here there is a difference in grammatical category.

10. Zero initial: in non-particles this is realized as either glottal stop or zero, the former more common after pause or in emphatic speech, the latter more common in connected speech. In particles it is always zero: the clitic nature of particles finds phonetic expression in their being fused to the preceding morpheme if they begin with a vowel. There are two possible treatments of these facts. First, the phonology could be made sensitive to the morphosyntax by making the required (vs. optional) realization of /0/ as zero to be a consequence of the grammatical category of the morpheme in question. There are similar differences of particles vs. other classes in the realization

of tone, and it is not unheard of cross-linguistically for such a form class to have its own phonological peculiarities. Alternatively, a contrast could be recognized between phonemic glottal stop and phonemic zero: /ʔ/ > [ʔ ~ 0] and /0/ > [0]. This would be slightly more cumbersome notationally, but would not require any grammatical category of morphemes to be directly hooked up with the phonology. (on the other hand, the phoneme /0/ would occur only in particles). Actual minimal pairs are few; so far only three are known:

ʔɔ 'pond; to bark'

ɔ 'Sentence particle (prompt-questions)'

ʔé 'many; to call'

é 'Sentence particle (meaning unknown)'

ʔú 'classifier for books'

ú 'diminutive suffix'

Examples:

ma thā ʔɔ 'it's a pond'

ma thā ɔ 'it's water, huh?'

kəjē li ʔú 'Kayah book'

kəjē li ú 'the Red Kayah' (full self-designation of the Kayah).

### Cluster Initials

pl          phr

kl          khr

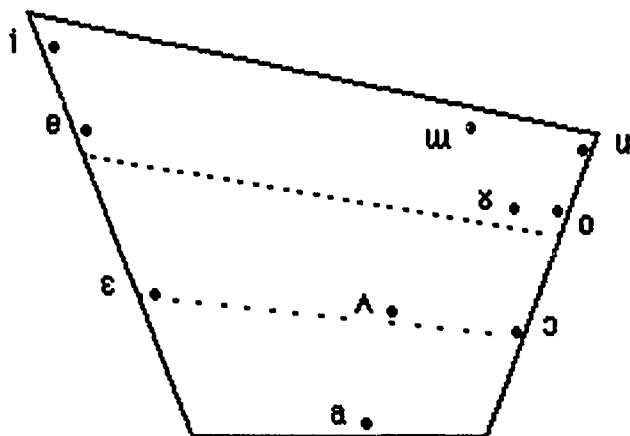
In clusters, there is complementary distribution between aspiration and the l/r contrast; either may be treated as redundant. If such mutual determination is considered undesirable, there is one bit of evidence for choosing the l/r difference as fundamental:



simple aspirates never occur in the low falling tone, but [phr khr] do quite frequently. In other words, aspiration in clusters has different phonological behavior from aspiration in simple initials. Historically it is in fact the /r/ contrast that is significant, with aspiration being a side effect of the medial /r/ (cf. the Tai languages, where /h/ is common as a reflex of \*r).

### 1.3 Rhymes

#### Simple Rhymes (the vowel system)



#### Notes

1. /i u/ are cardinal [i u].
2. /a/ is a low central [A].
3. /e o/ are slightly higher than cardinal.

4. /ɛ ɔ/ are cardinal [ɛ ɔ]; /ɛ/ is not as low as the [æ] of Thai which is often transcribed as (ɛ).
5. /ɯ/ is a centralized high back unrounded vowel, more central than the similar vowel of standard Thai, but not a fully central [ɨ].
6. /ɤ/ is a slightly centralized upper-mid back unrounded vowel, very similar to the /ɤ/ of Thai and Shan (which is often transcribed 'ə'), and occurring almost exclusively in Thai or Shan loanwords.
7. /ʌ/ is a centralized version of cardinal [ʌ], similar to the English vowel often transcribed with the same symbol (the vowel of *but, gum, hug* etc.)
8. All vowels except /i u a/ are raised under the low-level and low-falling tones. /ɯ/ is slightly lowered under the high tone.
9. The phonemic status of /ɤ/ is solid only in the high tone, cf.

dú cut, slice	dɛ at(preparation)	dá give
tú just now	tɛ chest, box	—
sú wrong	sɛ insert	—
chú kindle	—	chá clear; ten
jú shrink	—	já (a particle)

The total number of occurrences of these three vowels in high-tone words is (from about 1800 words) 17 ɯ : 10 ɤ : 6 ʌ . Three of the words with /ʌ/ are particles.

In other tones, words with /x/ are rare, and include many Shan/Thai loans, e.g. mǎ  
'city' (Shan máŋ = Thai myaŋ), mé dx 'Ficus' (Thai madỳa = Shan màak lǎ).

### Compound Rhymes

wi	wa	we		
	ja		jo	(jw)

The onglides /j- w-/ are usually closer to [-e-] and [-o-] respectively. There is at least one form /pja/ which may result from a fusion of /pe/ + /a/ (see 4.5).

/jw jo we/ are relatively rare. The first occurs only in one word so far recorded: thā khjw 'the Salween river'. /jo/ similarly occurs only in the common word mjō 'type, kind' (from Shan/Burmese). /we/ is more common, but seems confined to Shan and Burmese loanwords, e.g.

cwe habituated	W[ritten] B[urmese] cwài stick fast in, adhere, use habitually, chronic.
pwe celebrate, festival	WB pwài id.
jwe small change	Shan jɔj id. (cf. Thai jɔj 'break up into small particles')

Note that the rhyme spelled '-wai' in WB is pronounced [-wɛ ] in modern Rangoon Burmese.

**/wa/** occurs with all initials (including 0) except **/v/**. **/wi/** occurs with all simple initials except **/v t n h 0/**. **/ja/** occurs only after labial obstruents and simple(non-cluster) velars.

#### 1.4. Tones

There are four major tones plus one marginal one.

mid	├	33	ā
low level	└?	22	a (unmarked)
low falling	└ <sub>?</sub>	21	à
high	└ <sup>?</sup>	55	á
high falling	└ <sub>?</sub>	52	ae (see below)

Examples:

rē across	kō blow away
re all over, at random	ko general classifier
rè trellis	kò wear on head
taré wax	kó do temporarily
cókare <sup>52</sup> otter	chiko <sup>52</sup> shrimp

The forms given above are those found in isolation or before pause. In this position all but the mid tone end in glottal stop. The low-falling tone goes to the bottom of the vocal

pitch register and thus terminates with a very brief stretch of creaky voice shading immediately into glottal stop. Mid tone syllables are slightly longer than those with other tones, but this difference disappears in connected speech. In connected speech pitch and contour are the functioning cues; the final glottalization disappears, except that there is a slight tendency for low-falling tone words to retain it, in the form of creaky voice, probably as a concomitant of that tone's very low terminal pitch.

The sentence particles have special characteristics relating to the glottalization feature of tones, as they do with zero initial. The distribution of glottalization is often reversed: the mid tone often has final glottal stop, especially in the Sentence Particle *pā* 'irrealis', while other tones seldom do. There is also a Sentence Particle, meaning 'only', with the anomalous phonetic shape [to:<sup>22</sup>]; i.e. low level tone with long vowel and no final glottal stop. This seems to contrast with the negative morpheme /to/, also a Sentence Particle, but with the short vowel and pre-pause glottal stop standard to the low level tone. One might consider setting up a distinctive final consonant /-ʔ/ on the basis of this contrast, but it would occur only in the first morpheme mentioned. There is also some evidence that the two particles are in complementary distribution syntactically; see 6.6. In the meantime I will adopt the spelling /too/ for the particle 'only'. This may be taken to reflect a speculation that the form is actually a fusion of two syllables /to/+o/; if there were such a combination, it would in fact have the phonetic shape described, given what was said above about the zero initial.

Note that the final glottal stops are considered features of the tone. They are not etymological: neither the low-level nor the low-falling tones descend from \*stop-final syllables, while the mid tone does include reflexes of such syllables (see below).

In a sequence of two mid tone words, the first may have a slightly lowered tone. The beginning of the low-level tone may be raised by a closely-preceding high-tone syllable; thus the pitch contours of *kúma* 'fingernail' and *kúmi* 'tail' are very similar, and the two words are distinguished (apart from the vowel difference) by the presence of creaky voice in the latter as much as anything else<sup>1</sup>.

There is no tone sandhi properly speaking; that is, shift between tones conditioned by the tones of adjacent syllables, often with morphological relevance.

The relative lexical frequency of the tones is as follows: low-level > mid > high > low-falling > high-falling. The high-falling tone is quite rare, occurring most often in animal names and some other polysyllabic morphemes. But minimal pairs can be found:

pe <sup>52</sup> bat	təpe over (turn __, once __)
	təpè one layer; one arrow
di he <sup>52</sup> kind of frog	dí hé frog says
di be <sup>52</sup> paper wasp	dí bé yellow rice
se <sup>52</sup> same as	sé back at, in response

Since an apparent high-falling tone can be produced by a high tone syllable followed by a low-level or low-falling tone syllable with the same vowel and zero or /ʔ/ initial,

(e.g. *le ʔe* 'forage for stg' in connected speech is phonetically identical to a hypothetical \**le*<sup>52</sup>), and since the high-falling tone is lexically infrequent, it would be possible to analyze the high-falling tone as a high tone syllable followed by a low-level tone suffix with zero initial and vowel copying the vowel of the preceding syllable. Vowel harmony of this sort occurs in other affixes (see 1.5); on the negative side, the proposed suffix would only occur after high tone syllables, as there are no examples of mid or low-falling tones with this low-level copy vowel following; i.e. no instances of tones [31] or [21:]. These facts could indicate that addition of the proposed suffix also conditions a change to high tone for the preceding full syllable. Henceforth I will represent the high falling tone as a suffixed /ə/: this vowel will also be used to write the vowel of prefixes which also copies the following vowel (see next section). For example, 'bat' is written /təpleə/, 'cockroach' is /lɔ kiə/, and so on; see 2.3 below for further discussion.

Support for this analysis can be found in cognate forms in the 'upper' (*ke<sup>h</sup> khu*) dialect of East Kayah (spoken to the north of Mae Hong Son town, and otherwise very close to the dialect described here) and in West Kayah (*ʔ* signifies breathy voice:

	East	East(upper)	West
crab	chwa khri <sup>52</sup>	chwa khreo <sup>52</sup>	(ɕuo)
bat	təple <sup>52</sup>	təpleu <sup>52</sup>	pié
otter	có kərə <sup>52</sup>	cɔ kərəu <sup>52</sup>	tsó kəŋ̄
shrimp	chi kɔ <sup>52</sup>	--	si kó
cockroach	lɔ ki <sup>52</sup>	--	lɔ k̄i

The 'upper' East Kayah forms show the suffix in a less reduced form, with its own vowel quality (back rounded non-low; the difference [o~u] is probably just hesitation during transcription). The West Kayah cognates show that the preceding full syllables originally had a variety of tones; the regularly corresponding (and synchronically underlying?) forms in East Kayah would be təplé 'bat', cɔ kərə 'otter', chi kò 'shrimp', and lɔ k̄i 'cockroach' (the West Kayah 'crab' is cognate only to the first syllable of the East Kayah forms).

#### Phonology of Prefixes.

Prefixes are proclitic syllables with a characteristically reduced range of phonological values, always preceding a full syllable, with which they form an iambic rhythm (unstressed-stressed). There are two subtypes:

1. ?i (with tone restricted to mid, high and low level)
2. CV, where C = /p t k/, and V is a copy of the vowel of the following syllable, tending towards schwa in connected speech. We write this type as Cə, directly preceding



its main syllable: pəthə 'upwards'; təmjō 'one sort'; kəjē 'person, Kayah'. Note the minimal pair kədā 'door' (prefix plus main syllable) vs. kɿ dā 'spaces between figuration'(two full syllables).

This latter type of prefix, with non-contrastive vowel and tone, is found elsewhere in Tibeto-Burman (e.g. Jinghpaw) and is probably to be reconstructed for the proto-language. It is also typical of the Mon-Khmer languages, where it is commonly termed 'minor syllable'. In both Tibeto-Burman and Mon-Khmer the prefixes/minor syllables as attested in the modern languages do not always have the morphological/derivational function implied by the term 'prefix', but such functions can often be seen in relic form even where they are no longer fully productive. Incorporation of the CV type of prefix in the syllabic canon would now give the 'sesquisyllabic' structure below:

$$\begin{array}{cc} T_p & T_f \\ (C_1\text{ə})(C_2)(C_3)(G)V \end{array}$$

$T_p$  is prefix tone (always low-level for CV prefix),  $T_f$  is full-syllable tone. We may also use schwa for the copy-vowel suffix posited in (1.4) above; it could be incorporated as:

$$\begin{array}{cc} T_p & T_f \\ (C_1\text{ə})(C_2)(C_3)(G)V(\text{ə}) \end{array}$$

We could also pursue this with inclusion of the ?i prefixes, as follows: if  $C_1 = /p t k/$ ,

/ə/ is copy-vowel ~ [a],  $T_p$  = low-level; if  $C_1 = /ʔ/$ , /ə/ is realized as [i] and  $T_p$  = high, mid, or low-level.

### Vowel Harmony

The vowel we are symbolizing as /ə/ is found only in affixes: prefixes pə-, tə-, kə-, tə- 'one', suffix -ə, and a suffix -ə possibly found in tə<sup>13</sup> 'some' (see 2.3 above for further discussion of affixes). In all instances it copies the vowel of the full syllable to which it is affixed. In prefixes this vowel is reduced towards schwa as speech tempo increases; in suffixes it merges completely with the vowel of the full syllable.

Vowel harmony in prefixes is found elsewhere in Tibeto-Burman, e.g. Bodo-Garo and Mikir (Benedict 1972, 97); within Karenic 'White Karen'<sup>2</sup> is reported to show vowel harmony in pronouns (Grierson). Outside Tibeto-Burman, the vowel of the 'minor syllable' of Khmu<sup>?</sup> (Smalley, 16)<sup>3</sup> shows vowel harmony under certain conditions.

1.7. Outline of Historical Tonal Correspondences.

mid	A <sub>1</sub> , D <sub>2</sub>
low level	A <sub>2</sub> , B <sub>1</sub>
low falling	B <sub>2</sub>
high	D <sub>1</sub> , B'

The proto-tone categories used here are those proposed by Haudricourt (1946, 1953, 1975). The following table, adapted from Mazaudon 1985, shows the equivalent categories in the correspondence sets of Luce (Roman numerals), followed by comma and the reconstruction of Jones 1961:

proto-initial	proto-tones			
	*A	*B	*B'	*D
aspirate	III, `h(asp)	VI, ´q(asp)	Va	VIII, ´?
1 voiceless, glottalized	II, `´	VIa, ´q	V, `h	VIII, ´?
2 voiced	I, `´(asp)	IV, ´´(asp)	--	VII, `?

### 1.8. Orthographies

Kayah in general has no established written form, although orthographies for Kayah languages have been devised. A Roman-letter orthography has been developed by Catholic missionaries for a language of Western Karenni, and several prayer books exist in this orthography. The language is either a type of West Kayah or one of the transitional dialects between Kayah and Kayoh (Lehman 1967 mentions a Catholic script used for Manaw; this is probably a reference to the same orthography). There also exists a script invented recently by members of the indigenous self-determination organization, the Karenni National Progressive Party. This script is in the Indic style, consisting of main graphs for initial consonants and secondary graphs, superscripts and subscripts for vowels and tones. It is not obviously derivative of Burmese, Thai, Shan, or any other pre-existent script known to me, which may well be by intention. It seems accurate for the West Kayah dialects, but it includes graphs for sounds not found in West Kayah and seems to be meant to be applicable to all Karen languages (for instance it includes a spelling for the voiced velar fricative found in Sgaw and Pwo). Unfortunately it lacks means of writing two of the phonemes of East Kayah, namely the low-falling tone and the vowel /ɤ/. It is probably not known or used by more than some two hundred people connected with the K.N.P.P.

## 2. Morphemes, Word Formation, Grammatical Categories.

### 2.1. Monosyllabicity

Southeast Asian languages have often been characterized as monosyllabic. It is now generally recognized that a truer formulation would be, 'Southeast Asian languages are monosyllabic, but...' with particular languages requiring various amounts of qualification to the 'monosyllabic myth'. For Kayah we can say that the great majority of morphemes are monosyllabic, and that there is no certain instance of a morpheme of less than a full syllable (the uncertain instances are the affixes, see 2.3 below).

There is a sizeable number of truly polysyllabic morphemes, some of which fall into semantically definable classes. Unanalyzable loan words include *khanakhá* 'king' < Shan *khǎn há khám*; *hōsaphō* 'airplane' < Shan *há* 'boat' + Burmese *saj-bhau* 'ship', perhaps influenced by Kayah *phō* 'winged insect'. Another group consists of names of plants, animals, and insects: *mēleké* 'pineapple', *lehá* 'teak', *kúpé* 'butterfly', *phébná* 'mantis'. Others fall into no obvious class: *níné* 'real', *jéjo* 'shadow, image', *lǎi* 'yet'.

Kayah morphemes are usefully divided into Free and Bound types. Free morphemes are those capable of functioning as a major clause constituent such as Subject, Object, or main verb (=head of VP; all these terms will be discussed below); Bound morphemes cannot thus function alone, but must combine with some other morpheme. A 'word' can then be defined as a minimal Free form.

Two types of word-forming processes can be distinguished in Kayah, affixation and compounding. The former is not in fact a productive process but refers rather to a collection of relic forms. The latter is highly productive, playing a central role in all types of nominal constructions. It is also possible to analyze the multi-verb constructions of the Verb Complex as instances of compounding. This chapter will describe principally nominal compounding; an extended discussion of the putative verbal compounds will occupy Chapters 4 and 5.

## 2.2. Compounding.

Certain grammatical classes are Bound by definition (e.g. Prepositions, Classifiers); others include both Free and Bound members (Verbs and Nouns). Compound expressions may contain all possible combinations of Free and Bound morphemes:

F + F sīne gun + thi penis → trigger

hí house + khu upper surface → roof

F + B thā<sub>1</sub> water + mē bamboo-section → water container

pù ox + po enclosure → cattle-pen

so tree, wood + klā boat → boat

B + F tē fish + bū white → kind of (large white) fish

## B + B

sí heart + plo small round thing → heart(the organ)

tē fish + ú (a suffix) → fish(general term)

The above examples are restricted to noun-noun compounds except for *tē bū* (*bū* is classified as a verb); for another example involving a verb cf. *mi dā* 'lighter, flint-and-steel', consisting of the noun *mi* 'fire' and the verb *dā* 'to forge, to strike a light'. For further discussion of the syntax of modification in the Noun Phrase, see Chapter 7.

We may also mention the existence of Verb-Object compounds, which are made up of verb plus NP that seem to relate to each other as main verb and Object, but whose meaning is specialized. With *nō* 'enter' there are

nō hóhé 'attend school';

nō jechua 'be a Christian' (*jechua* 'Jesus')

for a non-idiomatic version of the former cf. *kanō dā hóhé kū* 'go into the school building'.

Cross cutting the Free/Bound contrast is that of versatile versus restricted. Of the preceding examples, *pú**pó* consists of two versatile morphemes: *pú* 'ox' could also combine with *ja* 'flesh', *nó* 'horn', etc., while *pó* could be preceded by *the* 'pig' *chā* 'chicken', and so on. On the other hand *kā* 'boat' never occurs without *só* 'wood', which in this word takes on the character of a prefix; *kā* is a (highly) restricted morpheme.

Another example of this type is *lè sw* 'shed, granary': *lè* is a highly versatile element meaning basically 'place for \_\_\_'. *sw*, on the other hand, is a morpheme: it occurs nowhere but in this compound noun, and if asked to gloss it we can only reply that it stands for that part of the meaning of 'shed' that remains after the meaning of *lè* is subtracted. The same can be said of *kɪ̄* in 'boat'.

When restricted morpheme combines with restricted morpheme it becomes increasingly difficult to identify the meanings of the component parts, and the expression verges on being a single polysyllabic morpheme. Often it is not possible to be completely certain that recurrent elements in compounds represent instances of a single morpheme. One example is the syllable *do* in *donè* 'tell legends', *dodè* 'hold out, offer in the hand', and *domé* 'show'. Of the second elements, *mé* is 'to look', and *de* may be related to the second part of *sedè* 'come forth (as new fruit)' (*se* 'to fruit'); *do* might then be said to have a meaning like 'set forth, offer'. Again, consider

<i>súplɔ</i> 'rope'	<i>plɔ</i> 'rope, string' (Bound)
<i>súba</i> 'harvested hemp'	<i>ba</i> 'classifier for sheets, flakes, mats'
<i>súse</i> 'bamboo splint'	<i>se</i> ? (distinct from 'fruit')

A common meaning assignable to *sú* in these three words might be 'fiber'. This would be distinct from the second syllable of *rasú* 'oil'; but which of the two occurs in *kúbīsúse* 'peanut'? *kúbɪ* is the general term for 'bean', but is the word to be analyzed as 'bean + oil + fruit' (*kúbɪ-sú-se*) or as 'bamboo-splint + bean' (*kúbɪ-súse*)?



Comparative study may reveal that components of currently unanalyzable polysyllables derive from older full morphemes; e.g. *jòlem̄* 'squirrel' consists of *jò* 'rodent, rat' plus unanalyzable *lem̄*; *lɛ* however is probably cognate to Pa-O *li<sup>A1</sup>*, a Free noun, while *m̄* is cognate to the second syllable of Kayoh (Brè) *ji<sup>B2</sup> m̄<sup>D2</sup>* 'squirrel'. Conversely, bound morphemes that may seem identical in modern Kayah may represent a merger of historically distinct forms, as in *tapleə* 'bat' and *jòlepleə* 'flying squirrel', which seem to share a morpheme *pleə*, perhaps meaning 'flying mammal'. Actually 'bat' is cognate to Pa-O *pla<sup>A1</sup>* and Pwo *phla<sup>B1</sup>*, while 'flying squirrel' relates to Pa-O *phre<sup>D2</sup>*, Pwo *phlai<sup>D2</sup>*, and Sgaw *pli<sup>D2</sup>*. Thus 'bat' and 'flying squirrel' reconstruct as contrasting forms, something like *\*(?)pla<sup>B'</sup>* and *\*ble<sup>?</sup>* respectively.

### 2.3. Affixes

Prefixes in General. Prefixes occupy a borderland between phonology and morphology. Phonologically they have a precise definition involving a reduction of possible phonemic distinctions and clitic-like attachment to a following full syllable. Their status as morphemes is much fuzzier, mainly because of the difficulty in assigning them any sort of precise meaning or morphological function, and also because of their low productivity. Given a list of all the occurrences of a particular prefix, it is always possible to separate out a group in which the prefix seems to have a common meaning, but there is always also a residue that does not fit.

Prefixes are in fact relic forms in Kayah; prefixes of this sort are a hallmark of the Tibeto-Burman languages, although similar phenomena can be seen in the other linguistic stocks of Southeast Asia/China. Also characteristic of Tibeto-Burman is cyclical prefixation: as old prefixes are lost or fuse with initials, new ones arise. Karenic shows this feature: e.g. the *th-* of Kayah *the* 'pig', *thē* 'bear' and *thwi* 'dog' is in origin an old prefix, the Tibeto-Burman roots being something like \*wak, \*wam, and \*kwi (in the last case the initial \*k- was reanalyzed as a prefix, then dropped before addition of the new prefix t(h)- ). But even by proto-Karen times this \*th- was assimilated into the initial consonant system; the prefixes of Kayah (and of other modern Karenic languages), now on their way to moribundity, must have been an innovation of the proto-Karen stage or possibly later.

#### Prefixes Listed

ʔi. Often found in names of tools. At times it has a nominalizing function:

mū to hammer, strike                      ʔimū a stick (for beating)

thá to plow                                      ʔithá a plow

ci cut with scissors                        ʔici scissors

pɔ to thresh                                    ʔipɔ a hammer

but cf. chá 'hit with the fist', ʔichá 'pound in a mortar'. Other verbs containing ʔi- are ʔilò 'to plant (seeds)', ʔikhré 'to winnow', ʔiché 'to tell', ʔiʔɛ 'dirty', ʔipia 'narrow'.

ʔi-. Found in many verbs, often those denoting body movements or functions: ʔitā 'get down, as from a vehicle' (tā 'descend, fall'), ʔícha 'jump', ʔílò 'bathe', ʔívī 'to whistle', ʔíchλ 'urinate', ʔínē 'to fart', ʔírō 'to sing', ʔípu 'to crow'. It also occurs in nouns: ʔíthw 'post', ʔíbè 'bamboo shoot', ʔíja 'flesh', ʔíphē 'my father', ʔímò 'my mother' also

ʔíkhu<sub>1</sub> 'earth, the world' (khu<sub>1</sub> 'on the upper surface of' [Localizer, cf. 7.1])

ʔíkhu<sub>2</sub> 'to wind, as thread on a spool' (khu<sub>2</sub> Classifier for spools of thread)

ʔi-. Less common, but with a few apparently-derivational uses:

dū sweep ʔidū broom

chē hurt sí ʔichē afraid (sí 'heart')

ché sew sé ʔiché sewing machine (sé 'large machine')

Other examples are ʔitē 'what?', ʔiilū 'the Kayah New Year festival', ʔikē 'shawl, blanket', ʔithó cover as with a blanket', ʔihí 'to spin (thread)'; cf. also ʔithé 'Crataeva' (Thai *phāk kum*, a plant with edible sour leaves), possibly related to tathé 'centipede' (see *tə-* below).

tə- (with echo-vowel). Common in both nouns and verbs. Two groupings of common meanings can be found.

1) directional: təva 'encircling', təka 'in a curving path', təlwá 'past', təphā 'out of the way', təja 'past going in opposite directions'. These are members of a subclass of directional verbs, but not all members of that class have prefix tə-.

2) undesirable personal qualities: təmwī 'crazy', təklw̄ 'stunted', təro 'timid', təké 'dwarfed', təklē 'lazy', təkhr̄ō 'stupid', təkhwā 'speechless'.

Other occurrences of tə-: təpē 'kick stg', təkli 'gnaw', təlū 'roll stg up', təné 'steep', təcá 'cool', təmε 'tusk (of pig)', təkhwá 'lizard', təphé 'cotton', təchē 'elephant' təkō 'box', təpwī 'longan'(a fruit), təcá 'anthing', təmò 'sun', təthé 'centipede' (cf. ?ithé 'Crataeva' above).

tə- 'one'. Homophonous with the preceding, but sufficiently distinct and unitary semantically to warrant separation. See 7.3 below on numerals and counting.

kə-. Relatively rare; no discernable semantic common denominator. kajē 'person', kəlō 'hill', kədó 'lid' (dó 'a wall, to enclose'), kəne 'almost', kəp̄ō 'noisy, deafening', kəlwa 'slanted', kasé 'itch'.

kə- and pə-. These occur with direction verbs only. kə- means 'Subject changes location', thus ?a phjá the 'he picks it up' (phjá 'take', the 'ascend') versus ?a phjá kəthe 'he takes it and goes up'. In the first sentence only the (unspecified) Object/Patient moves upwards, while in the second the Subject/Agent moves upwards (and incidentally carries the Object along). (There is also a difference in the grammatical structure: phjá the is a Resultative V-V, while phjá kəthe is a Sequential V-V; see Chapter 4).

pə- means 'orientation', as in jò pəthe kə 'raise the leg', mé pəthe 'look upwards', pəhē 'up ahead, in front' (hē 'go from home'). In general the meaning of verbs with pə- does not include a change in location of any entity (in 'raise the leg' the emphasis is on the fact that the leg ends up pointing upwards, not on the motion that results in that state). pə- occurs with most directionals except lə 'descend'; the equivalent of the non-occurring \*pələ is tālē, with tā 'fall' and tone-change on lə.

?a. Primarily a third person pronoun 'he/she/it/they/', but with several derivationally-flavored functions. It may precede certain Bound nouns to produce a Free noun; e.g. plə 'classifier for small round things', ?aplə 'a seed'; kū (a) 'classifier for holes' (b) 'the inside of', ?akū 'a hole'. But these could also be analyzed as noun-noun compounds, see (7) for further discussion. ?a also has an apparent nominalizing function with stative verbs: bū 'white', ?abū 'the white one'. This also has an alternative analysis,

namely as a verb modifying a preceding noun, more or less as in the English translation.

Suffixes. Suffixation is far less evident in Kayah than prefixation. There is indeed no shortage of Bound morphemes that must be postposed to some other morpheme (what Chao would call 'start-bound'), but they do not parallel prefixes either phonologically (they show no restrictions on possible occurrence of phonemes) or morphologically (they have precisely describable functions). Here we will mention three suffix-like elements.

(a)

phú is a noun meaning 'child'. It is Bound, requiring a preceding personal pronoun if it has the sense 'offspring' (as Thai lûuk): vĕ phú 'my child', ?a phú 'his/her/somebody's child'; with the sense 'immature person' (as Thai dĕk) it must be followed by cĕ 'soft, tender, young'. Related to this noun is a morpheme meaning 'small': dĕ phú 'small village', ngĕ phú 'small banana plant'. Although this phú is verb-like in following a modified noun, it never occurs in a predication; the equivalent verb is patí ~ patĕ. A second probable derivative of 'child' is ú, with zero initial; in at least some uses of this phú is also possible: klāú ~ klāphú 'soldier' (klā 'army'), kajĕ li phú ~ kajĕ liú 'the red Kayah' (kajĕ 'person, Kayah', li 'red'). Here there is no diminutive meaning; the sense seems rather to be 'member of a class'<sup>1</sup>.

Note especially *thu* 'bird' and *tĕ* 'fish', both Bound nouns that are first syllables in dozens of names for species of bird and fish; the generic terms for 'bird' and 'fish' as Free nouns are *thuú* and *tĕú* respectively. *ú* thus has the reduced, clitic nature of an

affix, and its meaning, while discernable, is not very substantial.

(b)

The suffix -ə posited as underlying the high-falling tone (1.4) is quite parallel to the prefixes in that it is highly reduced phonologically and has a largely unapparent semantic function, with a subgroup of occurrences that could be called 'animal suffix'. However, while there are abundant examples of main-syllable morphemes occurring both with and without a prefix (which provides the best clues to the prefix's meaning), only the

following can be offered as such examples with -ə:

chi kó s̄	shrimp paste, <i>kapi</i>	chi koa	shrimp
	(s̄ 'rotten')		

ho	stealthily	hoa	hidden, out of sight
----	------------	-----	----------------------

bē	yellow	(?) dī beə	paper wasp
		(cf. phō dī 'housefly')	

(c)

From the prefix/numeral tə- 'one' is derived tə<sup>24</sup> 'some, certain ones', which could be analyzed as tə- plus a suffix /ə/ (or /á/; the tone of 'some' seems to rise beyond the level of the mid tone, but does not go as high as the high tone). This would amount to the joining of a suffix to a prefix, something not otherwise known to occur in Kayah.

## 2.4 Other alternations.

2.4.1. Tone Change. Under this heading we include two different phenomena: the expressive use of high tone and high falling 'tone', and a number of sets of morphemes, homophonous except for tone, of which some seem clearly related and some are perhaps simply coincidence. Many, though by no means all, of this type include the high tone as one of the pair, suggesting a former derivational function for that tone. Examples:

thō enclosure, container	thó cover as with a blanket
bōklé blink	bókhri close the eyes
bō classifier for lengths	khe bó leggings <sup>2</sup> (khe 'leg')
tacē a bird trap	ché to trap
du big	dúlōá be older of siblings
ma steady, fixed	rā má write down
le descend	tālē downwards
de put, place	dodē hold out, offer
	sedē form, as fruit (se 'to fruit')
ro be early (in the day)	rō classifier for mornings
chī nā chī sē	chī nā chī sē
both: all day and all night	(nā 'day', sē 'night'; both are classifiers)
ro other, another	rò particle denoting plural action by animates (see 4.3.7)

Note the additional alternation of initial in tacē/ché 'a trap/to trap'; this type of alternation is not common in Karen.



Two body-part terms require further description. *khε* 'leg' is found in the compounds *khε mā* 'knee' (*mā* 'joint'), *khε kē lē* 'hollow of the knee', and *khε rē* 'paw'. But the form *khī* is found in many other compounds: *khē do* 'lower leg', *khēka* 'thigh', *khē le* 'foot', *khē khi* 'shin'. Incidentally *khε*, the less frequent form, is the one that is etymologically regular. There also seems to be some sort of generic body-part element, with tonal variation, in the following:

<i>kūʔu</i>	mouth
<i>kuklís</i>	head
<i>kukhā</i>	tooth
<i>kúkhu</i>	hand
<i>kújā</i>	palm
<i>kúma</i>	finger nail ( <i>ma</i> 'shell')
<i>kúmì</i>	tail
<i>kúdū</i>	crown of the head
<i>kūce</i>	earring, long silver Kayah-style
<i>kusē</i>	earring, gold Shan-style
<i>kúlā</i>	necklace

As was pointed out in the discussion of compounds (2.2.1), the history of these sets of forms may show either descent from a formerly productive and clear-cut process, or coincidental convergence of formerly-distinct forms. An example of the latter may be

*d̥* 'village' and *ɔ* 'to enclose; a wall', which seem to share a notion of enclosure (this would involve the assumption that the Kayah or their predecessors built walled or stockaded villages). In fact these two words probably descend from syllables that contrasted segmentally as well as tonally; compare Pa-O *ɔɔŋ*(A1) 'village' and *ɔɔʔ*(D1) 'to cover, close'<sup>3</sup>.

**2.4.2. Reduplication.** Reduplication in Kayah is a morphophonemic process that copies the last syllable in a clause, with the meaning 'also, too, either'. Examples:

(1) *vē ma ʔe k̄l̄ phé thé ja ja*

1s be-so eat COM only pig flesh.

I ate only pork, too (as did he).

(2) *ʔa cwá k̄l̄ k̄l̄*

3 go COM

He went along too.

(3) *vē cwá to to*

1s go NEG

I won't go either.

(4) *mm, síʔichē ké r̄ò he he*

afraid AMB cold LEST

Mm, I'm afraid it'll be cold, too (e.g. in addition to raining).

The process is largely a simple matter of copying whatever syllable happens to be clause-final, regardless of either form-class or syntactic function. The reduplicated syllables in the above examples include:

ja 'flesh', a Free Noun, the head of an NP functioning as DO

kā 'with, also', a Verb Particle

to 'negative', a Clause Particle

he 'lest, possible bad situation', another type of Clause Particle

The only exceptions to this simple rule are: 1) certain clause-final particles may not reduplicate (see 6.2 below); 2) the occurrence of:

(5) thé phra kā ke ke ~ thé phra phra ke

pig to-sound COM PTC

(both:) It might also be a pig making noise.

Here it seems possible to reduplicate the lexical morpheme phra 'to sound', passing over a grammatical morpheme, the Clause Particle ke 'possible non-future situation'. More data is needed on this point.

2.4.3. Allomorphy. Kayah morphemes are invariant as a rule, with only a few exceptions which will be listed here.

1. The numeral 'ten' varies in tone: in the numerals 10-19 it is *chí*, while in 20-99 it is *chī* (see 7.3 for details). Note that this parallels English *ten*, which has *-teen* for the former and *-ty* for the latter. This variation differs from the type listed under 'Tone Change', in which the variants differ in category (e.g. *tacē* 'trap'[N], *chē* 'trap'[V]) and/or meaning (*dē* 'put', *do dē* 'offer'). The two forms of 'ten' cannot be said to vary in meaning also the choice of form is entirely predictable from features of its co-constituent.

2. The classifier for humans has the curious suppletive forms *phre~sí*, depending on the co-occurring numeral (cf. 7.3).

### 2.5. Form Classes

Notions of the form class ('word class', 'grammatical category') that have been useful in linguistics may be separated into three general categories; in roughly historical order they are:

(1) affixal (classical, Greco-Latin grammar). Classes are identified by characteristic

sets of affixes or inflections, e.g. verbs by person-number agreement marking.

(2) distributional (structuralist). Classes are defined by occurrence in 'frames or

slots constituted by other classes' (Chao,5); e.g. verbs in most Southeast Asian

languages can be defined as any and all forms that can cooccur with the negative morpheme.

(3) clause-functional (generative, X-bar theory) (a better label may be possible).

Classes are terminal symbols in phrase-structure rules; major form-classes

(Noun, Verb, Preposition) define the clause constituents (NP, VP, PP)--that is, they

are their sole required constituents ('heads'), at least in the deep structure.

All three of these also include semantic notions which are auxiliary rather than criterial: Nouns are the names of things, Verbs refer to actions, events, or states, and so on. Of the three, the first is of no use in Kayah, which like Southeast Asian languages in general lacks productive affixation or inflection. The second has been widely and

successfully used in analyses of Southeast Asian languages, but for reasons that will become apparent, it is not quite sufficient for Kayah grammar. The third, although it is in intention not 'discovery procedure'-oriented, can be adapted to such use.

Regarding distributional definitions of form classes, Kayah is unusual among Southeast Asian languages in that two near-universal definitions of Noun and Verb do not work well.

For Verbs negatability has been used as at least one of the defining characteristics in Chinese (Chao), Thai (Noss), Lahu (Matisoff), and Khmu<sup>2</sup> (Smalley). But the Kayah negative morpheme *to* is a constituent, not of the VP (or predicator), but of the clause, in which it occupies one of the final slots. It *can* be said that only Verbs can be inserted in the frame \_\_\_\_*to* and result in a complete clause; since the Verb is the only non-optional constituent of the clause, what possible occurrence of *to* really proves is clausehood, not Verbhood. But *to* cannot be used to pick out the Verb among doubtfully-identified morphemes within a clause (this contrasts with Thai, in which the occurability and position of the negative is a very useful analytic test). There does exist a listable class of morphemes, with abstract meanings, that may always follow Verbs and must precede any following Noun. Occurability before these 'Verb Particles' is then one defining feature of Verbs; some Verb Particles are *ɿ'* 'new situation', *pa'* 'durative, V's onward', *po'* 'additionally' (similar to Thai *iik*), and *la'* (obligatorily co-occurring with *to*) '(not) yet'

For Nouns in Southeast Asian languages a useful distributional test has been

co-occurrence with, and modification by, a numeral-classifier (or demonstrative-classifier) expression. This holds only in a weakened form in Kayah: the most usual place for the ClfP (Classifier Phrase, the numeral-classifier construction) is second-to-last in the clause, always potentially separated from the non-oblique nominal arguments by a locative Prepositional Phrase (see below for clause constituents and grammatical relations). Example:

- (6) ʔa khé thuú dǝ́ mi klē sō be  
 3 shoot bird at forest midst three CLF  
 He shot three birds in the forest. (5/6)

Thus, exactly as the Verb may be directly followed by the negative if all intervening slots happen not to be occupied, so a nominal Direct Object may be directly followed by a ClfP if the clause happens not to contain a Prepositional Phrase. But even in the latter case the ClfP does not form a unit with the fortuitously adjacent NP: just as the negative relates to the clause as a whole rather than to the main Verb the ClfP is best seen as just another nominal argument, bearing a grammatical relation to the predicating verbal expression (which we will call the Verb Complex). An important subclass of classifiers, the one whose members allude to notions of physical configuration (flat things, long things, etc.), does have a *semantic* relation to co-occurring nominal arguments, by virtue of the fact that all Nouns include in their lexical entries a notation of the classifier(s) that are used to count them. We may then say that any morpheme that may have such a

semantically-related classifier (= any morpheme whose lexical entry includes one of the aforementioned type of classifiers), is shown to be a Noun. While such a formulation seems valid, it must be recognized as being different from the purely-syntactic type of distributional definition used in classical structuralist studies.

The third, 'clause-functional', type of form-class definition will be used for Kayah in the following manner. The structure of the Kayah clause may be symbolized as

$$NP_1 [_{VP} [_{\bar{V}} VC NP_2 NP_3 ]_{\bar{V}} PP ClfP ]_{VP} Ptc$$

Nouns, as the heads (only non-optional constituents) of NPs, typically function as Subject, Direct Object, and Indirect Object. Verbs, the heads of VCs (Verb Complexes), typically function as main Verb or Predicator. The final Particles are a listable class. Classifiers are those morphemes that may be the co-constituents of Quantifiers (the listable class that includes numerals). Prepositions are those morphemes that may begin a PP.

The converse is also true--Subjects and Objects must be Nouns, Predicators must be Verbs--with the important proviso that only non-complex clauses are at issue; that is, clauses that do not contain other, embedded clauses. The existence (and prevalence) of nominalization means that Verbs and clauses may function as Object, although this phenomenon is less prevalent in Kayah than might be expected. Clauses that may seem to function as Subject are usually treated here as Topics, which are distinct from Subjects and considered to be outside of the clause proper.

Thus an important part of the definitions of Noun and Verb is a description of the grammatical roles they typically play in the clause. An equally important characteristic is behavior when functioning as modifier in the NP. Essentially, Nouns precede a modified head Noun, while Verbs follow. This excludes certain regular exceptions to the modifier-modified order for Nouns, involving the type of semantic relation between the parts of the expression (e.g. generic-specific; cf. Chapter 4); it also ignores the fact that postposed modifying Verbs are a special case of postposed Attributive Clause (9.2). But position when modifying a Noun remains a very useful test for distinguishing Verb and Noun.

Below is a list of the form classes of Kayah, including both distributional and 'clause-functional' definitions:

Noun occurs in the slot \_\_\_ *tə* 'X things, X's things'; can function as Subject, Direct Object, Indirect Object; precedes a modified Noun (with exceptions); can be counted by a ClfP containing a classifier that is lexically/semantically related.

Verb occurs in the slot \_\_\_ *lāĩ* *tə* 'hasn't yet X-d'; may function as Predicator, the head of the VC; follows a modified Noun. Includes all the morphemes translating English adjectives.

Preposition occurs in the slot \_\_\_ *ʔā* (e.g. *bəʔ ʔā* 'at this = here', *hú ʔā* 'like this'); introduces the PP.



Quantifier. occurs in construction with *kə* 'general classifier' (preposed or postposed according to the particular Quantifier); is one of the two essential components of the ClfP; includes the numerals.

Classifier. occurs in the slot *tə*\_\_\_ 'one X'; the other essential component of the ClfP; may be considered a special type of Noun.

Demonstrative. occurs in the slot \_\_\_Quantifier-Classifier.

Verb Particle. occurs in the slot *me jɛ́* \_\_\_ NP 'hard to do...' (e.g. *me jɛ́ lǎi hi to* 'not yet hard to build a house', *me jɛ́ iʃ* 'harder to do'); terminate the VC.

Sentence Particle. occur in the slot Verb-ClfP\_\_\_ ; a principal member is *tə*, the negative.

Class overlaps. In general the two major classes, Noun and Verb, are distinct, yet there are instances of morphemes with dual membership. Note the following:

(7) *bó se ʔo to*

rice fruit exist NEG

The rice doesn't have any grains.

(8) *bó se to*

rice fruit NEG

The rice doesn't fruit; the rice doesn't put out grains.

(9) *ʔa se lǎi to*

It hasn't fruited yet. (5/4)

Similar overlap is seen in *phō* 'flower; to flower', *chɛ́* 'thorn; be thorny'. Note also *ʔibe* 'language; to speak'. Ethnic designations may also be interpreted as having

membership in both Noun and Verb, in line with two characteristics:

1. like Nouns, they may modify a following Noun with a meaning of possession (*phrè haca* 'a Shan's clothes, the Shan's clothes'); they may also, like Verbs, modify a preceding Noun, with a meaning like 'X-type, X-style' (*haca phrè* 'Shan-style clothing')

2. they may appear inside the VC, a possibility normally open to Verbs and Verb Particles only (*ʔa ʔibe phrè cè nǒ to* 'he can't speak Shan at all'). This particular construction is probably to be classified as a Descriptive V-V (4.2.4).

There is a semi-regular pattern in which a single 'word' may function as Bound Noun and Classifier; note that, as mentioned, this may simply be an alternation between subclasses of Noun:

- |    |                                                                                                                                              |
|----|----------------------------------------------------------------------------------------------------------------------------------------------|
| mǎ | <ul style="list-style-type: none"> <li>a) plant (Bound N); e.g. bó mǎ rice plant, ʔa mǎ trunk</li> <li>b) Clf for plants</li> </ul>          |
| kū | <ul style="list-style-type: none"> <li>a) hole (Bound N)</li> <li>b) inside (Localizer)</li> <li>c) Clf for holes, nostrils, etc.</li> </ul> |

### **Morphosyntax, Part I (Verbal Constructions)**

The remainder of this grammatical sketch is divided into two sections. The first, after outlining some theoretical concepts (3), gives a fairly detailed description of the construction types and morpheme classes found in the Verb Complex (4), followed by some discussion of several possible formal approaches to the data (5). The second section contains less detailed outlines of several other areas of the grammar, including clause structure (6), syntax of the NP, PP, and Classifier Phrases (7), and nominalized and other non-autonomous clause types (8).

### 3. Theoretical Preliminaries.

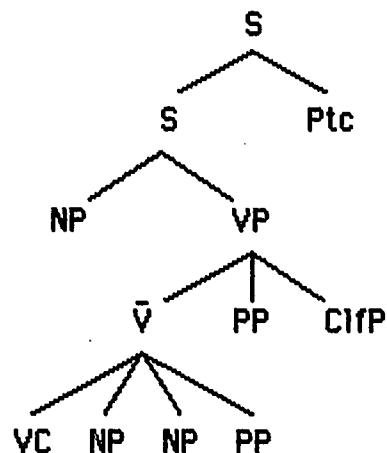
#### 3.1. Introduction; sketch of clause structure.

This grammatical sketch is primarily a work of description. As such it takes linguistic theory as a means to an end, and is not primarily concerned with advocating or improving on any particular theory. Therefore it does not include a comprehensive account of its theoretical assumptions; some of these will be discussed in various relevant places. Here it may be noted that the present grammatical sketch assumes only one level of syntactic description; there is no deep structure, and no transformations. This is largely a choice in favor of descriptive simplicity. Kayah has no movement, apparent or otherwise, of question words, and not even a limited or rudimentary passive construction; it thus has no need of the two classic movement transformations of generative grammar. Only in connection with some constructions in the VC is a multi-level transformational analysis really attractive; I discuss the possibility in the appropriate section. I do assume a distinction between syntactic structures, accountable for by phrase-structure rules, and lexical items, which are listed in the lexicon, and some of which are formed by morphological processes (in Kayah, principally compounding); the latter may be 'located' in the lexicon.

Our discussion of Kayah syntax will proceed somewhat irregularly. Although discussion of clause structure will be deferred to Chapter 6, it will perhaps be helpful to sketch it here. The linear sequence of clause constituents can be represented as follows:

NP VC NP NP PP PP ClfP Ptc

where VC is Verb Complex, ClfP is Classifier Phrase, and Ptc is Particle (more specifically, Clause Particle). I assume the following constituent structure:



The abbreviations should be self-explanatory for the most part;  $\bar{V}$  contains the subcategorized arguments of the verb, while VP contains optional 'adjuncts'. The upper NP is the Subject, while the lower two NP's may be thought of as (from left to right) approximately equivalent to the traditional Indirect Object and Direct Object of English.

A (rather unnatural) sentence illustrating all possible constituents is:

ne phō pè Phāλ dī né dīpò du hú ʔā ē  
 you cook for (name) rice Ne pot big like this QUES  
 Did you cook rice for Pha'a like this in a big pot?

Note that *pè*, glossed 'for', is not a preposition but a Verb Particle, a constituent of the VC. The two prepositions are *né* and *hú*, the former having no short English

equivalent.

I will refer to the two post-verbal NP arguments as Obj-1 and Obj-2 when two are present; a single post-verbal NP argument is designated as Obj-x. This is meant to imply a sort of neutralization of the distinction between Obj-1 and Obj-2, and will be discussed in detail in 6.4 below, but the reasoning may be sketched here. The semantic roles realized by the two post-verbal NP arguments are:

Obj-1: recipient of goods, beneficiary, causee, standard of comparison

Obj-2: patient, goal, etc. of action verb

Obj-1 is clearly more 'oblique', in an impressionistic sense, while Obj-2 is more 'direct'. But when there is only one post-verbal NP argument it may realize any of the listed semantic role types, and there seems to be little structural distinction between the Object realizing the 'oblique' roles and that realizing the 'direct' ones.

In order to discuss clause structure in detail, which includes justifying the constituent structure given and describing the various grammatical relations the structure embodies, it is first necessary to understand the rather complex inventory of construction types and morphemes making up the VC. This is because it looks as if it is the VC as a unit that takes the grammatical relations Subject, Object, and so on; thus we cannot simply talk of transitive and intransitive (or two-argument and one-argument) verbs; we must also know what happens when the VC contains, for instance, one transitive verb, one intransitive verb, and the particle *κῆ*.

We will thus proceed as follows: I will first sketch some theoretical concepts that are needed in describing the grammatical behavior of verbs, namely those of semantic roles and grammatical relations. Second, I will briefly situate the Kayah VC in its Southeast Asian context as an instance of what we will call Verb Serialization. I will then describe the types of construction that make up the VC; this will be done without much formalism, but will tacitly assume that the processes at work in the VC are more like compounding than like syntactic phrase generation. Finally I will question the last-mentioned assumption, considering also possible analyses in terms of underlying structures and transformations. Only then will I proceed to describe the other constituents of the clause, first in terms of the general structure of the clause, then in more detail in terms of the different grammatical relations.

### 3.2. Semantic Roles.

3.2.1. Preliminary remarks. Most syntactic theories assume a level of analysis at which a clause represents an action plus a number of participants having various relations to the action. These relations have been called 'deep' or 'semantic' cases (Fillmore), thematic or  $\theta$  ('theta') roles (Gruber, Jackendoff, Chomsky), semantic relations (Chafe), and perhaps other things besides; I will use the term 'semantic roles'. Typical labels for semantic roles are Agent, Patient, Instrument, and so on (Patient is sometimes known as 'Theme', at other times the two terms are used for distinct roles. Here I will make use of Patient only; see discussion below).

Much has been written on this topic, and I will not attempt a comprehensive survey here, but will outline a certain subarea that seems to be of significance in describing Kayah syntax. I assume that applying a more complete theory of semantic roles will not entail drastic changes in my analysis. This may seem simply an assertion of faith, but it does have ample precedent. The recent literature contains frequent examples of retreats from explicitness such as 'assume some version of \_\_\_\_\_ theory' and open-ended lists of roles like 'Agent, Theme, Goal...' The present discussion is thus not without company in at least some of its limitations.

I will assume that there is a small set of role types to which lexical specification of verbs may refer (possibly this set also provides the labels for relations in some level of



semantic representation, although I will not make use of such a level here). It has been pointed out that this is not a necessary assumption. For instance, ...one should not conclude that all semantic roles fall into one or more linguistically significant classes [such as Agent, Patient--DBS]. It is quite possible that the semantic roles assigned by some items are not classifiable...Perhaps the only linguistically significant piece of information an English speaker knows about the role assigned by *like* is that it is the role assigned by *like*.

Marantz 1984, 32

This terminology [Agent, Patient, etc.--DBS] implies a system of argument *types*, in that, for example, it implies that the agent arguments of two different verbs have something in common. Although this may be true,  $\theta$ -theory as outlined here is not committed to this idea.

van Riemsdijk and Williams 1986, 241

These statements are made in connection with discussions of ' $\theta$ -role assignment', which is actually a designation for certain configurational relationships between syntactic elements. In this grammatical sketch, on the other hand, I will make use of semantic roles in describing types of verbs and types of constructions they enter into; for example, I want to be able to make statements like 'the Patient role of the first verb is associated with the Patient role of the second verb' (the sense of 'associated' to be explained below). It would be quite difficult to do this without the assumption that it is legitimate to use the same role-type label for semantic roles specified by different verbs.

I will also make the assumption, shared by most writers on the topic, that a single clause contains at most one instance of any given role-type (Fillmore 1971, etc.; see the

following section for the apparent exception of Locative)

3.2.2. Selected and non-selected roles. A first distinction that must be drawn is one between roles that may occur freely with all verbs, or with a wide variety of verbs, and roles that occur only with a relatively small set of verbs; or in other words roles that (respectively) cannot be used to subcategorize verbs and those that can. Roles of the latter type will be specified in the lexical entries of those verbs with which they can appear; I will thus speak of specified or selected roles. For example, Agent is specified for *speak* and *walk* but not for *hear* or *die*; and Patient is specified for *fall* and *melt* but not for *sit*. Non-specified roles are generally an option with any verb, and typically include 'scene-setting' elements such as time and place expressions. The neat dichotomy between specified and non-specified is complicated by the fact that certain roles are specified by some verbs, but can occur freely with others; we will return to this momentarily.

Note that what we are discussing is a relation between verbs (predicates) and NP's (participants); being a relation between two types of entity, it can be viewed from the perspective of either type. The term 'specification' takes the viewpoint of the verb, which is subcategorized in terms of the roles it occurs with. It is also possible to take the viewpoint of the NP, which is or is not referred to in the verb's specifications: from this perspective, the verb is often said to 'assign' roles to NP's. I will generally treat the two as logically equivalent, although it is possible to distinguish them; see the

discussion of the theory of Marantz below (1.4).

To return to the contrast between specified and non-specified roles, in addition to the co-occurrence patterns described, there is also what may be called a conceptual distinction. Specified roles are inherent to the meaning of the verb, in the sense that they represent participants understood to be present even if they are not explicitly denoted in a sentence. For instance, the verb *buy* denotes an action that is understood to involve four participants, namely a seller, a buyer, goods and a price<sup>1</sup>. The event reported in *Tom bought a crocodile* mentions only the buyer and the goods, but a seller and a price are understood to be also involved. Let us assume a principle of discourse something like the following: in conversation, it is natural to respond to an assertion with a question that asks for a possible, but unexpressed, role to be filled in. So, of unexpressed roles, non-specified roles can be questioned with any verb, but specified roles can be questioned only with a verb that specifies them. We can now illustrate the inherent status of the seller and price roles of *buy*: natural responses to *Tom bought a crocodile* include the questions *For how much?* and *From who?* in which the first speaker is asked to fill in the unexpressed but inherent roles. In contrast, *Tom saw a crocodile* expresses all inherent semantic roles. A natural question in response might be *When?*—but that could equally well be asked of the sentence with *bought*. Most events occur in time, so the presence of a Time role is nondistinctive. The presence of the roles representing seller and price is, however, a distinctive feature of *buy*, and is one feature that

differentiates it from *see*. Thus to respond to *Tom saw a crocodile* with *From who?* or *For how much?* would be odd, because seller and price roles are not inherent to *see*<sup>2</sup>.

This inherence of semantic roles can also be put in terms of entailment: *Tom bought a crocodile* entails that there was a price (and a seller), while *Tom saw a crocodile* does not entail the presence of a price, although it does not rule it out.

There is generally a direct connection between semantic-role specification and syntactic subcategorization. A verb that specifies an Agent typically also specifies a Subject NP, specification of Agent and Patient correlates with subcategorization for Subject and Object, and so on. There do exist examples of what may be seen as subcategorized syntactic elements that bear no semantic role; once such is 'expletive' *it* in cases like *It's hot today* (roleless Subject) or *I hate it when you sing like that* (roleless Object; or perhaps appositive?). Another instance of this type may be the 'middle' construction of English, which seems to demand an adverbial, something that is usually seen as intrinsically unable to bear a semantic role: *Apricots pick easily*/\**Apricots pick*. If these are accepted the correlation between semantic roles and syntactic constituents must stand as something other than an exceptionless principle.

Generally a given role type is either always specified or never specified: Agent and Patient seem to be always specified, while Time never is. But certain role types may be specified by some verbs and also occur freely with all other verbs: Locative is a well-known example. It is an option with a wide variety of verbs, such as *eat, read, die,*

*work*, etc., but it is inherent and specified with a relatively small class of verbs including *put*, *keep*, *enter*, *live* (in the sense of 'reside'). Notice that the meaning of many of these verbs involves not just location, but a change of location. This has been referred to as a distinction between 'inner locative' (specified) and 'outer locative' (non-specified) (Fillmore 1968).

A different sort of possible distinction among semantic roles involves obligatoriness of the specified roles. This seems to apply to verbs like English *eat*, which clearly specifies a Patient, but need not appear with one, thus contrasting with *gnaw* and many others, which are ungrammatical without their specified roles:

- (1) Kim ate lunch.
- (2) Kim ate.
- (3) Kim gnawed the bones.
- (4) \*Kim gnawed.

There is disagreement over how to deal with such cases (for instance, whether the Patientless *eat* is to be considered the same morpheme as the one with Patient); however the question is not of great importance in Kayah, in which it is probably safe to say that specified roles are never optional in this sense. What may give the appearance of optionality of specified roles is the ease with which Kayah, like many Southeast Asian languages, omits (or, represents by empty pronouns) elements that are definite; that is to say, that are identifiable or recoverable by virtue of their status in preceding discourse.

This is quite different, in fact is the opposite of the function served by the optional role as in English: in *Kim ate*, the Patient is not present precisely because what was eaten is indefinite, unspecified, and probably unimportant. In the English (and probably general European) type, something is omitted because it is unspecified; in the Southeast Asian type, something is omitted because it already has been specified. The Southeast Asian equivalents of the omitted unspecified element are items referring to the most unmarked participant appropriate to the action, and the choice of unmarked Object is often identical or similar across languages:

(5) <u>English</u>	He's eating	He can write
<u>Kayah</u>	ʔa ʔe dī	ʔa rā cè liú
		he write able book
<u>Thai</u>	kháw kin khāaw	kháw khǎan naŋsǔw pen
		he write book able
<u>Mien Yao</u>	nīn nān ɲǎaŋ	nīn hāi fǐə dzàaŋ
		he able write letters
<u>Chinese</u>	tā chī fàn	tā huì xíe zì
	(all: s/he eat rice)	he able write letters

Additionally, *read* is usually *read+book*, *hunt* is *shoot+animal*, and so on. There seems, in fact, to be an inverse correlation between omission of unspecified elements (as in English) and of recoverable elements (as in the Southeast Asian languages).

3.2.3. Dependence between role types. There is thus a group of roles that are always specified, such as Agent and Patient; a group of roles that are never specified, such as Time; and the variable type represented by Locative. So far there is no particular dependence between the specified roles: of the three roles discussed, verbs may specify all possible combinations:

Agent only	talk, laugh
Patient only	melt, glow
Agent, Patient	eat, repair
Agent, Locative	enter, leave
Patient, Locative	sink, float
Agent, Patient, Locative	put, insert

There does seem to be a general dependence between Locative and a second role, of whatever type, since no verb seems to specify Locative alone<sup>3</sup>.

There is a more particular sort of dependence between role types, in which specification of a certain role seems to somehow be connected with the specification of a certain other role, or at least with the possibility of such specification. An example that has received some attention is the role Benefactive, in the sense of the participant for whose benefit an action is done (Fillmore 1971, 52; Chafe, 151). As Fillmore puts it, 'Benefactive constructions occur only in sentences with Agents'. Fillmore then alludes to (and rejects) a possible set of 'redundancy principles', which would presumably include

the information that any verb specified for Agent is also optionally specified for Benefactive. This is in fact proposed by Chafe: 'every action verb [i.e. verb that specifies Agent, D.S.] ...can optionally have a beneficiary attached'.

The role Instrument also seems to have a similar sort of dependence. Consider the familiar set of examples:

(5) Tom cut the rope

(6) A knife cut the rope

(7) Tom cut the rope with a knife

In these the same verb occurs with 1) Agent, Patient; 2) Instrument, Patient; 3) Agent, Instrument, Patient. Given that this is a fairly common pattern, and that there are very few (perhaps no) verbs that can take Instrument that cannot also take Agent, we could state a generalization to the effect that any verb that is specified for Agent may also be specified for Instrument, and when appearing with Instrument the Agent may be omitted. The latter proviso differentiates Instrument from Benefactive, which was similarly an option with any Agent-specifying verb, but the appearance of a Benefactive in a sentence does not allow the omission of Agent.

Fillmore 1972 gives another type of interpretation of Instrument as nonbasic. In the framework of a three-participant action in which an agent manipulates a Patient causing it to move into contact with a Goal, Fillmore notes that with some verbs either Patient or Goal may be realized as the syntactic (Direct) Object. Compare the following



(8a)	I beat the stick against the wall.	(Fillmore's 16a)
(8b)	I beat the wall with the stick.	16b
(9a)	I broke the hammer on the vase.	20b
(9b)	I broke the vase with the hammer.	20a
(10a)	I loaded hay onto the truck.	23c
(10b)	I loaded the truck with hay.	23a

Here in each of the (a-b) pairs, the interpretation is that the stick, hammer, and hay are Patients, and that in the b) sentences they have been pre-empted for the Object position ('set aside' in Fillmore's terms) by a Goal, which is thus brought 'into perspective'. The 'saliency conditions' that allow such pre-emption have a general reference to what might be called salience of effect. Thus if one of two affected participants is human it is more likely to be brought into perspective, since effects on humans are intrinsically more salient than effects on non-humans. With two non-human participants affected, the one that is more saliently affected will be in perspective, as in (9b), where the Goal (the vase) undergoes a drastic change of state. In (10b) the salience of effect is due to completeness (the truck is understood to be full). In (8a-b) there is little to choose between effects on the two participants, which would help explain why the two sentences seem appropriate to virtually identical situations, unlike the pairs (9a-b) and (10a-b).

Of these sentences (8b) and (9b) would certainly be said to have Instrument roles, but the approach just outlined would allow Instrument only as a 'derived notion' (Fillmore *op.cit.*, 77). In this view, no verb needs to be specified for Instrument.

3.2.4. Specification by verb; by non-verb. How does an NP 'get' its semantic role? or alternatively, how do we know what role a given NP has in its sentence? In general, an NP is shown to have a specified role by occupying a position which is linked with a role specified by the main verb of the sentence. For instance, *Tom* in *Tom ran five miles* has the Agent role because *run* specifies the single role Agent, and in English a single specified role is realized as Subject, in English marked by the structural position (say) NP dominated by S, and *Tom* occupies that position. The 'position' referred to may be such a position in the syntactic configuration, or it may be the fact of association with a morpheme (particle, adposition, affix, case); or it may be both. So in *Equus currit* 'the horse runs' *equus* is shown to be the Subject, and hence to have the Agent role, by being marked with nominative case as well as by (presumably) being immediately dominated by S. The 'link' between a position and a semantic role may be deducible from some set of principles, or it may be explicitly indicated as a lexical property of the verb; see 3.4 and 5.3.3.

An NP is shown to have a non-specified role also by occupying a 'position', but one that is not referred to by the main verb's specifications. Most often this is a position in the second sense described, in which the NP is identified by a morpheme. Thus in *Ann*

*works in Oakland*, the bearer of the Locative role, *Oakland*, is identified as such by appearing as object of the 'locative preposition' *in*. We may then, if we wish, say that morphemes like *in* specify semantic roles, just as verbs do. Note that we cannot say, of the morphemes associated with *specified* roles, that they specify those roles. They help identify the NP as having some specified role or other (i.e. they help constitute its 'position'), but the particular role is specified only by the verb: thus *equus* still has nominative case in *Equus moritur* 'the horse dies', but it is the specifications of the verb *morior* that tell us that this Subject has the Patient role.

There is also an intermediate case, namely specified roles whose identifying 'position' consists of a marker that never marks any other role; an example would be the markers of 'Inner Locative' and its subtypes. E.g. the preposition *from* invariably marks Source, whether specified as in *Dick stole a horse from the squire*, or non-specified, as in *From the roof you can see the ocean* or *The apples from that tree are the best* (assuming in the latter case that an NP in a noun-modifying PP can be said to have a semantic role). There is thus a sense in which the 'inner', specified Source is redundantly indicated: in the example, both the verb *steal* and the preposition *from* have features referring to a Source argument<sup>4</sup>. Languages that have morphemes specifically marking Agent and Patient (e.g. Lakhota, Eastern Pomo)<sup>5</sup> would perhaps be an example of the case in which all roles are redundantly indicated, thus constituting an exception to the remarks in the preceding paragraph concerning morphemes associated with specified

roles.

The notion of elements other than verbs having semantic role specifications is dealt with in several current theories:

Marantz (1984) separates assignment of semantic roles from specification of roles--or in his terms, role assignment from argument-taking. Arguments that are both specified for and assigned roles by their predicates (verbs and sometimes adjectives) are *direct arguments*; arguments that are specified by verbs but that are assigned roles by other elements are *indirect arguments*. Thus in the above example *steal* specifies Source, but the Source argument *the squire* is an indirect argument since it gets its Source role from the preposition *from*; furthermore *the squire* cannot 'have' the Source role of *steal* (or, cannot 'be' the Source argument of *steal*) until it is 'assigned' the Source role by *from*. Finally, although Marantz does not discuss this, it can be presumed that, besides direct and indirect arguments of predicates, there must also be items with roles that are not specified by verbs, but that are assigned roles by something. These would be the 'non-specified' roles Outer Locative, Time, and so on. See the feature matrix below; perhaps in this framework these are to be called non-arguments.

Items having semantic roles

	role specified by verb	role assigned by	e.g.
direct arg.	yes	verb	Agent
indirect arg.	yes	other	Source
(non-arg.?)	no	other	Time

In Lexical-Functional Grammar (LFG), the concept 'argument-taking predicate' also extends beyond verbs, so that prepositions (and other items) may be associated with information about (among other things) the semantic roles borne by their arguments.

The difference between specified and non-specified roles discussed earlier would then become a difference between roles specified by verbs (and possibly by other items) and roles specified only by non-verbs. Terminological reform is now in order: let us call the former type Core arguments, and the latter type Peripheral arguments, following Foley and Van Valin (1984).

In Kayah, non-verbs that have role specifications include the prepositions, which are at least partly similar to prepositions in English and other languages. For instance, an NP is marked as Locative by appearing as Object of one of the three prepositions *dá*, *bá*, *mú* (the three differ in evidential value); those three prepositions may then be said to have a specification for Locative. Somewhat more interesting and problematic are several roles that are marked by non-verbs, but these non-verb markers are not directly in

construction with the 'marked' NP's. For instance:

- (11) ʔa ʔe ʔé lǎ nǎ vĕ dī to  
 3 eat much more at-all is rice NEG  
 He does not eat more than me.

Here the first post-verbal NP *vĕ* 'I' has a role that can be called Standard (of comparison). Its presence is clearly licensed by the Verb Particle *lǎ* 'comparative', and it must in some sense get its role from *lǎ*. Yet *lǎ* and *vĕ* are not adjacent. One way to capture these facts is to claim that *lǎ* specifies the role Standard; that *ʔe ʔé lǎ nǎ* is a compound word, which 'inherits' the Standard specification from its constituent *lǎ*, and thus assigns the role Standard to *vĕ* (as well as assigning the role Patient to *dī*, and so on). With such an interpretation, the distinction between direct and indirect argument begins to blur. From one point of view, the main verb of the sentence, as head of the VC, is *ʔe* 'eat', which specifies only Agent and Patient; the Standard *vĕ* must then get its semantic role from the morpheme *lǎ*, making *vĕ* not only not a direct argument, but not an argument at all (in the sense of 'argument' as 'semantic role specified by the main verb'). On the other hand, if *ʔe ʔé lǎ nǎ* is a compound word, it, the verb, is specified for a Standard argument (by inheritance from its component *lǎ*) and directly assigns that role to the NP *vĕ*. So at the level of syntax *vĕ* is a direct argument, while at the level of individual morphemes it is not an argument at all.

### 3.3. Roles for Kayah.

Below I list the semantic roles that it seems useful to refer to in the analysis of Kayah. I distinguish specified roles, non-specified roles, and 'variable' roles which have both specified and non-specified versions (e.g. inner and outer locative). Each listing sketches the semantic content of the role and states its most usual grammatical realization.

#### *Specified*

Agent. The sentient, purposeful initiator or controller of an event. Not significantly different from the role of the same name in other treatments. Unmarked realization is Subject.

Patient. The entity that undergoes a process or to which a state or location is ascribed. There seems to be little need to distinguish Patient from Theme (e.g. undergoer of change of state for the former versus located or described entity for the latter). There are no general requirements of animacy or humanness. Unmarked realization is Subject if no Agent is present, Object-x if an Agent is present, Object-2 if Agent and certain other roles are present.

#### *Non-specified*

Time. Time-when expressions. These frequently occupy the 'Topic' position, in which case they are not elements of the clause strictly defined (see 4.3 below). Within the clause they appear as Prepositional Phrase; specifically, the PP bearing the grammatical

relation 'Obl-2' (see 4.2, 4.4 below).

Extent. Always represented by a numeral-classifier phrase, these denote spatial extent, temporal duration or frequency of action, or number of participants affected. Of these, temporal duration/frequency can clearly be seen as an independent role, but the other two concepts are somewhat problematic. Consider the following hypothetical examples:

(12) Phāλ hē khé te mi sō phó / sō nā  
 (name) go shoot animal three time three day  
 P. went hunting three times/for three days.

(13) Phre mē phjá khru đé thō nλ sō khri  
 (name) take firewood at over-there three packbasket  
 Phremeh got three packbaskets-full of firewood over there.

(14) phú cē cwá đé Rù sɔ lē sí sō  
 child go at (name) CLF three  
 Three children went to Rusoleh.

In each case the classifier phrase is syntactically distinct from the NP that it 'counts', but it is unclear whether it always represents a distinct semantic role. In (1) the duration and frequency expressions cannot be identified with any other participant, and so would seem to bear their own semantic roles. However in (3) the two constituents *phú cē* 'child' and *sí sō* 'three (humans)' seem to make up a single element semantically. (2) is perhaps an intermediate case, since *khri* 'packbasket' and *khru* 'firewood' designate different sorts of objects: there is no 'is-a' relation as there is between



'child', which designates an object of the same sort as, but more specific than, 'human'; *khri* can also be a noun, which may further be counted by its own classifier (*khri* *nā* *mē* 'two packbaskets'). At the level of semantic roles, however, the two seem to make up a single composite Patient.

Evidently this classifier-phrase constituent is only the realization of an independent semantic role when expressing duration or frequency. For the other cases it may be best to consider the NP and classifier phrase as distinct in the syntax but not at the semantic-role level, perhaps analogous to *Pat* and *nose* in sentences like *Terry hit Pat on the nose* (cf. Fillmore 1968).

Others. Here may be listed concepts like Standard, discussed in (3.2.4) above. These are expressed by a combination of an NP in Object-x or Object-1 position and certain particles in construction with the verb. Besides Standard, a prominent element of this type is something that can be called Comitative. As discussed, these can be said to be specified by non-verbs; their usual realization is as the grammatical relations Object-x and Object-1.

Variable

Locative. Non-specified: spatial setting of event. Specified: reference point of motion, often the place towards which moving participant moves (Goal), sometimes place from which (Source), sometimes other orientation point. The differences in the specified type are part of the lexical specifications of the accompanying verb; that is, besides

specifying the fact of occurrence of the Locative role, the lexical entry also indicates some particulars of the interpretation of that Locative. In the following examples the Locative expression *vē hi* 'my house' appears with three verbs that specify differing particular interpretations of the Locative arguments:

- (15) *ʔa cwá dʒ vē hi* He went to my house. Goal  
 (16) *ʔa the dʒ vē hi* He went/came out of my house. Source  
 (17) *ʔa cwá təlwá vē hi* He went past my house. other

I will occasionally refer to these subtypes of Locative as  $Loc_{Goal}$ ,  $Loc_{Source}$ , and so on. In (15) and (16) the Locative appears as the Object of the preposition *dʒ* 'at' (distal or non-visible). In (17) the Locative is the Object of the complex verbal construction *cwá təlwá* 'go past' (for arguments that morphemes like *təlwá* are verbal modifiers rather than prepositions, see 3.3.7). These examples also demonstrate that Locative may be realized as PP (more common) or Obj-x.

Recipient. Non-specified: receiver of good or ill effect of event. Probably restricted to animates, but not restricted to occurrence with action (i.e. [+Agent]) verbs; e.g.

- (18) *bé se pɔ kəlō du kē kū lè pè lū nʌ rʌ*  
 eye big mouth wide BEN 3OBV Nʌ Ptc  
 His eyes were [too] big and his mouth [too] wide for him. (96.7)

Thus Recipient in Kayah is unlike the type of Benefactive described above (3.2.3), in that it does not have any dependence on specification for Agent.

Specified: recipient of goods with transfer verbs (*dá* 'give', *ʔičhē* 'sell'); also perhaps causee (recipient of causation) with Directive verbs (*nō* 'command', *dá* 'let'; see 3.2.3). Realization is Obj-x or Obj-1. One bit of evidence in favor of lumping together recipient of goods and causee is the fact that when both appear in a sentence the causee is Obj-1, the goods (presumably Patient) is Obj-2, and the recipient of goods is in an oblique relation: as if recipient of goods had been pre-empted by causee for the role Recipient. E.g.

- (19) *vē nō dá Phāλ rū né ʔa pò*  
 1s command give (name) money Ne 3 YS  
 I told P. to give money to his younger sibling.

I do not posit an Instrument role for Kayah. There is no morpheme specifically marking Instrument. Many Kayah sentences that can be translated with an Instrumental *with*-phrase in English involve the preposition *nē*, but these seem to be one instance of a more general phenomenon. Constructions with *nē* are discussed in (6.5) below; here it may suffice to point out that many examples are consistent with Fillmore's backgrounded-participant analysis. E.g.

- (20) *ʔa chū li mi né tō thé nλ*  
 3 poke red fire Ne iron Nλ  
 He poked the fire red with an iron. (355.5)

Here we may construe *mi* 'fire' as the Goal and *tō thé* 'iron' as the manipulated Patient. Either role may appear as Object of *chū*.

(21) *chū bó*            to plant riceplants    (*bó* = Patient)

(22) *thé chū kənō*    Needle pricks finger    (*kənō* = Goal)

In (20) the Patient 'iron' is pre-empted for the Object position by the Goal 'fire', since the latter is more saliently affected (it becomes red). The verb *chū* 'stab, prick, poke in' may then be specified as [Agent Patient Locative<sub>Goal</sub> ], with no need to refer to an Instrument.

### 3.4. Argument Structure.

The semantic roles specified by a verb are often called its 'arguments'; the set of a verb's semantic role specifications may then be called its argument structure (as in Williams 1981); I will also occasionally refer to specified roles as 'arguments'. When a verb acts as unitary predicator ('main verb', head of VP, etc.), the roles of its argument structure are associated with various NP's in the sentence's syntactic structure, usually via a system of grammatical relations; see the following section (3.5).

When a verb is not acting as unitary predicator, argument structure is also relevant. This has been studied in connection with compounds in English: for instance, a deverbal noun or adjective may appear with other morphemes in a compound word such as *church-goer*, *slum clearance*, *hand-woven*, and so on. In all these examples there is a sense in which the preposed N is associated with one of the semantic roles specified by the verbal root of the following deverbal noun: *church* may be the Goal of *go*, *slum* is the

Patient of *clear*, *hand* could be called the instrument of *weave*, and so on.

Like most Southeast Asian languages, Kayah has little derivational morphology. But it is also true to its areal type in having extensive use of constructions that combine verbs in complex predications of the sort often referred to as verb serialization or verb concatenation (I will be claiming that some of these constructions in Kayah are more like compounds than syntactic phrases; if compounding is an aspect of morphology, this is an important sense in which Kayah--and presumably other languages of the area--cannot be said to lack morphology). For at least some of these constructions it is useful to consider how the argument structures of the verbal components of the construction interact. (I use 'construction' as a neutral term for any unit of syntactic structure, including both syntactic phrases and compound words).

Consider the hypothetical example;

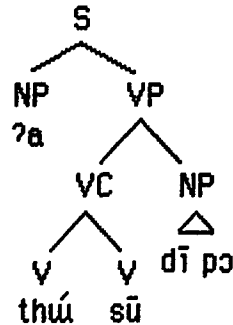
(23) ꜜa    thú    sū [ dī    pɔ̃ ]

3 wipe,rub dry rice pot

He wiped the pot dry.

Assume that (1) has the following structure:

(24)



Arguments for this sort of structure will be given in (6.2) below. The node VC (Verb Complex) may be considered as simply V, thus a compound; or it may be  $\bar{V}$ , a syntactic unit intermediate between the minimal (lexical) and maximal phrasal levels. The question of the phrase-level status of VC will be taken up in (5) below.

There are two questions that can be asked of this example: what is the argument structure of VC? and how does that argument structure relate to the argument structures of the component verbs *thú* and *sū*? Assume that the component verbs have the following sort of argument structures:

*thú* 'rub, wipe' [Agt Pat]

*sū* 'be dry, clean' [Patient]

In all, this involves one Agent specification and two Patient specifications. But the sentence as a whole contains only two argument expressions, the Agent *ʔa* 'he' and the Patient *dī pɔ* 'pot', implying that the VC also specifies just one Agent and one Patient;

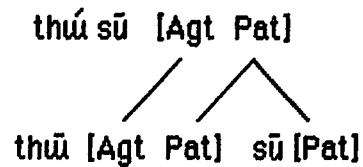
*thuí sū* 'wipe dry' [Agt Pat]

Let us refer to the above as a derived argument structure. There are a number of ways that the second question above, concerning the relation between derived argument structure and component argument structures, could be answered.

- a) Since the VC *thuí sū* has the same argument structure as its first verb *thuí*, perhaps the first verb simply transfers its argument structure to the whole VC. This could be taken as evidence that the first verb is the head of the construction.
- b) Since, of the total of three argument specifications for the two verbs, it is one Patient that is 'missing' (or redundant?) in the derived argument structure, perhaps there is some rule whereby identical or redundant role specifications cancel out.
- c) In (23) *di pɔ* has a dual role, being the participant that is wiped and also the one that becomes dry: it has the Patient role of both *thuí* and *sū*. In this view the Patient specification of the derived argument structure associates (to use an intentionally vague term) with the Patient specifications in both component argument structures.

Although these three formulations are not mutually exclusive, they apply in different ways to the different types of multi-verb construction. Accordingly I will use a fairly detailed type of representation of derived argument structure, keeping in mind that all details will not be necessary to describe all types. For *thuí sū* the structure can be diagrammed as below:

(25)



Here the upper level shows the derived argument structure, the lower level shows the component argument structures, and the slanting lines show mapping between the two--in this instance, representing the association of the two Patients of the component argument structures. Note that some of the VC components that contribute to the derived argument structure are not verbs, as was outlined in (3.2.4) above in connection with the example repeated below:

- (11) ʔa ʔe ʔé lǎ nǎ vĕ dī to  
 ʔ eat much more at-all 1 rice NEG  
 He does not eat more than me.

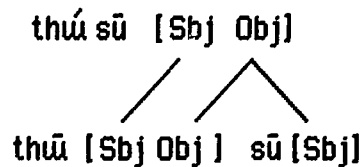
The derived argument structure for (4) will include a component represented as /ǎ [Standard], even though /ǎ is not a verb.



### 3.5. Grammatical Relations

An alternative representation to (3) of 3.4 that should be mentioned is one that refers to grammatical relations rather than semantic roles. E.g.

(27)



This is of course nearly identical to (25); it might, however, make a difference at some level that (25) maps together entities of the same type (Patients), while (27) maps together entities of different types (Sbj and Obj). For this and several other reasons, a brief discussion of grammatical relations is in order.

Recall that when a verb acts as unitary predicator the roles of its argument structure are associated with various NP's in the sentence's syntactic structure. In most theories this occurs via a system of grammatical relations (sometimes also referred to as 'grammatical functions'). I will not go into the question of the status of grammatical relations as theoretically basic versus derivative; here and in my ensuing use of such terms as 'subject' and 'object', it will not make much difference whether the reader takes the terms as abbreviations for configurational definitions like 'NP of S', or as standing for theoretical primitives. I will assume that there are generalizations concerning the realization of semantic roles as grammatical relations (e.g. Agent has the unmarked

realization Subject), and that particular verbs may lexically specify certain grammatical realizations that depart from the generalization. The latter mechanism would be needed, for example, to ensure that in *Bob hates rats* and *Rats annoy Bob*, the arguments *Bob* and *rats* will appear in the proper syntactic positions even though they have the same semantic roles in both sentences. Section 5.3.3 below will include one possible version of this approach to grammatical realization in the context of a more formal account of derived argument structure.

A distinction was made above between 'direct' and 'indirect' arguments, which we have seen can also be put in terms of arguments specified by the verb only versus arguments specified by the verb and by some non-verb. This distinction has an analog on the level of syntax in a distinction between 'pure' and 'oblique' grammatical relations (the opposite of 'oblique' would be better as 'direct', but I wish to avoid duplicating terminology between semantic and syntactic levels). The pure relations are more varied and abstract in semantic value: they realize a variety of roles, and those roles have relatively abstract meanings (Agent, Patient, 'experiencer', and so on). The oblique relations are 'impure' in that they tend to be fairly concrete and restricted semantically. They are often marked (as by English prepositions) even when the 'pure' relations have no overt marking, and they are marked by items that are restricted to marking a certain role (that specify that role, as discussed above). In Kayah it is useful to distinguish the purely configurational grammatical relations Subject, Object, and Object-2 from the oblique

relations marked by prepositions or particular categories (i.e. the Extent constituent, always consisting of a ClfP).

In describing argument structure in Kayah, I will be most concerned with the semantic roles that, in the unmarked case, are realized as these 'pure' grammatical relations. These are the direct argument roles Agent, Patient, and Recipient (and Causee, if that is distinct from Recipient, keeping in mind that Recipient also has a non-specified version); also some of the roles that are specified by non-verbal VC constituents, an example being Standard, as illustrated in (3.2.4) above. Standard is realized as Obj(-1) and so interacts with the role specifications of full verbs.

The roles that are not specified by VC constituents (whether verbs or not) will not receive as much consideration, since they are not often realized as the 'pure' grammatical relations (the fact that these relations may realize non-verb-specified roles like Standard means that 'pure' cannot be defined as 'realizing core (verb-specified arguments)'). In practice this means that the Inner Locative will receive short shrift. Inner Locative is indeed specified by verbs; however its grammatical realization is usually unproblematic, being typically identical with the realization of the non-specified (Outer) Locative, namely PP (Ob1-2), a position that is usually open and available. For instance the verb *ʔiʔhri* 'buy' specifies (let us say) Agent and Patient, and the verb *cwá* 'go' specifies Agent and Locative, the Locative being further specified as goal ('place to which') and realized as Ob1-2 (PP). The two verbs may combine in what I will call a

Sequential V-V (4.2.2), as in *vē cwá ʔiphri haca* 'I went to buy clothes'. To that sentence one may easily add a Locative expression, as *vē cwá ʔiphri haca dē klébe* 'I went to buy clothes at the market'. But it is difficult to say whether that added Locative is the Locative specified by *cwá* or the unspecified Locative allowed with any action verb; or indeed whether such a distinction can be made. There is certainly no syntactic basis for differentiation; i.e. no alternative structures equivalent to the English *I went to the market to buy clothes* (Locative is specified argument of *go*) versus *I went to buy clothes at the market* (Locative is unspecified Adjunct). The upshot is that in considering derived argument structures Locative arguments can be safely ignored in many cases.

What does most need to be sorted out in derived argument structure are the roles that may be realized as the three 'pure' grammatical relations (Subject, Object-1, Object-2). For convenience, I will sometimes refer to verbs according to the number of pure relations they take. I will use the following abbreviations:

- $V_j$             takes Subject only (Agent or Patient)
- $V_t$             takes Subject and Object (Agent-Patient,  
                  perceiver-perceived, possibly others)
- $V_d$             takes Subject, Object-1, Object-2 (Agent, Patient, Recipient)

Note that these abbreviations need not be taken as direct references to grammatical

relations. They could stand for particular interactions of argument structure and rules of grammatical relation realization, the latter including both general statements and idiosyncratic lexical features. To put it another way:

$V_i$  = a verb whose arguments include nothing that may be realized as Object

$V_t$  = a verb whose arguments include one that is realized as Object

Note that a  $V_i$  could actually specify two arguments if one of them is Locative;

similarly a  $V_t$  could specify three arguments.

Grammatical relations are marked in Kayah by a combination of structural position and syntactic category:

	order	configuration	category
Subject	1st before V	sister of VP	NP
Object-1	1st after V	sister of VC	NP
Object-2	2nd after V	"	NP
Oblique-1	3rd after V	sister of <u>VC</u> or of $\bar{V}$	PP
Oblique-2	4th after V	"	PP
Extent	5th after V	sister of $\bar{V}$	ClfP

All of these grammatical relations will be discussed below; for now we will give them some preliminary characterizations.

Subject. Generally fits the traditional use of the term..

Obj-1. Similar to English Indirect Object: recipient with transfer verbs. Also causee in certain causative constructions.

Obj-2. Similar to English Direct Object: Patient with action verbs, and so on. When only one Object is present, it will be referred to as simply Object; the question of which of the two Object positions should be identified with this single Object will be discussed in (6.4) below.

Obl-1. No good English equivalent; the closest might be the Instrumental *with*-PP.

Obl-2. Usually Locative.

Extent. No real English equivalent; describes temporal duration of action or number of participants affected, as discussed above. Since this is an instance (and the only instance) of one-to-one mapping between a semantic role and a grammatical relation, I do not bother to give separate names to the two. If separate names are desired, this could be called the Obl-3 relation.

In the following chapter (4), the descriptions of V-V types will include derived argument structures of the type outlined in 3.4, with the arguments identified as semantic roles. These representations will be used as a descriptive device only. Chapter 5 will turn to the questions of what part such representations should play in the grammar

of Kayah, whether they are best cast in terms of semantic roles or grammatical relations should include, and at which level or component of the grammar they are to be situated. The reader may consider this inductive strategy of exposition as analogous to the type of phonological analysis which begins with a list of phonetic segments and then proceeds to a phonemic 'solution' (e.g. Emeneau's (1951) analysis of Vietnamese).

## 4. The Verb Complex

### 4.1 Introduction.

#### 4.1.1. Verb Serialization; characteristics and problems.

The Kayah Verb Complex (VC) is the site of combinations of a bewildering variety of morphemes, some indisputably Verbs, some clearly non-Verbs, some less easy to classify. Since similar situations obtain in many other languages, both in and outside of the Southeast Asian area, some remarks about the general phenomenon are in order.

We will refer to the general phenomenon as Verb serialization. It includes both 1) immediate concatenation of individual verbs and 2) strings of units made up of verb plus NP argument (most typically, Verb-Object constructions). It is useful to use the terminology of  $\bar{X}$  theory in formulating the definition so as to include the latter type: the serialized items are of the category type V, and may be of the category level zero (i.e. lexical verbs) or higher (i.e. V' or V'', which may include the verb's subcategorized arguments). To this should probably also be added the feature of lacking overt marking of the relation between the components. The latter feature is added partly as a practical limit on the field of investigation: I recognize that verb serialization constructions may be closely related to structures that do have overt markings of embedding or co-ordination.

Terms previously applied to various sorts of V serialization include 'verb series'



(Vietnamese; Emeneau 1951), 'verbal expressions in series' (Chinese; Chao 1968), 'compound verbs' (Burmese; Okell 1969), and 'verb concatenation' (Lahu; Matisoff 1973). I will adopt Matisoff's term 'concatenation' for immediate serialization of verbs (without intervening arguments), distinguished from sequences of verb-plus-argument, which can be called VP-series (one might also use the  $\bar{X}$  terminology,  $V^0$  serialization and  $V''$  serialization); this then leaves 'serialization' free for use as the cover term for both. It is interesting to note that verb concatenation seems to be more characteristic of verb-final languages, while VP-series is typically associated with verb-medial languages. This is evidently not confined to the Southeast Asian area: Foley and Van Valin, citing languages of Papua New Guinea and Africa as well as of Southeast Asia, claim that

extensive nuclear junctures [i.e. V concatenation, DBS] are a widespread feature of V-final serializing languages

(Foley and Van Valin 1984, 193)

In fact it is probably better to put the generalization in the opposite way: VP-series are found overwhelmingly in verb-medial languages. It is true that V-final languages seem to make the greater use of concatenation, but V-medial languages also consistently show construction types that fit the definition of concatenation, often of the modal auxiliary type (verbs underlined):

- (1) Thai      khǎw khəaj    jàak pen thahǎan rya  
                  he   has-ever want be soldier boat  
                  He once wanted to be a sailor (Noss, 116)
- (2) Chinese    tā ài chī bīngchilin  
                  he love eat ice cream  
                  He loves to eat ice cream (Chao, 738)

Chinese also has concatenations of the 'result complement' type:

- (3)            pài    sǐ le    ge cāngying  
                  slap die-PFV clf housefly  
                  swat a fly dead      (Chao, 473)

Thus verb-medial languages have both concatenation and VP-series, but verb-final languages usually do not have VP-series (but see Wheatley 1985 for exceptions in the Yi languages of China). In Vietnamese, concatenation and V-argument serialization may alternate, according to syntactic and/or semantic conditioning:

V + Directional, conditioned by (phonological) size of NP Object

- (4)    đem    ra    ba    chai    rượu    to    túng  
          bring exit 3   bottle wine big big  
          brought out three real big bottles of wine (Nguyen, 399)
- (5)    đem    rượu    ra  
          bring wine exit  
          bring out some wine                    (Ibid.)

V + Result, conditioned by realis/irrealis

(6) hùm ăn mất trâu

tiger eat lose buffalo

a tiger ate the buffalo (Nguyen, 402)

(7) coi chừng, hùm ăn trâu mất

tiger eat buffalo lose

Be carefull the tiger may eat up the buffalo (Ibid.)

This Vietnamese data, in which concatenation and VP-series express identical propositional content, is good evidence that the two types of construction are similar enough to count as instances of a single phenomenon.

Verb serialization as I have defined it is definitely a surface phenomenon, in the non-technical sense of 'surface'. Most obviously, regarding the restriction to constructions without overt marking of relations between the verbs or VP's, it is quite possible that some such constructions are in fact unambiguous in the relations between their parts; for instance, it may be a lexical feature of certain verbs that any following verb or VP is subordinate. More importantly, to speak of sequences and strings leaves out the question of structure: a sequence verb-noun-verb may realize several possible sorts of configurational structure. For instance, the Thai expression *kwàat bāan sa-àat* 'sweep [the] house clean' might have a 'flat' structure consisting of three co-equal elements, but it may also have either of the two following bracketings:

A. *kwàat* [*bāan sa-àat*]

B. [*kwàat bāan*] *sa-àat*

One furthermore must wonder what labels belong on the brackets: in A, the bracketed material seems to be a clause with Subject and Predicate, while in B, what is bracketed looks more like a VP. We also need to know how to represent the fact that *bâan* 'house' is understood to be the participant that receives the action of sweeping and also the participant of which cleanness is predicated. This might be reflected in a third possible structure

C. kwàat bâan [X sa-àat]

One may then choose one of several possible methods of relating the items X and *bâan*; the field of choice may include methods that do not represent this information in the syntax at all. Finally, note that what has been discussed in this paragraph has been surface structure in the more technical sense: none of the possible structures mentioned is from an underlying level that must be subject to transformations (at least, not necessarily: structure C could conceivably be the result of a transformation moving the subject of the lower clause to the object position of the upper clause. 'X' would then be the trace left from that movement).

To exemplify a second genre of problem in analyzing verb serialization, consider a second Thai sentence: *chǎn kwàat bâan hāj Lék* literally 'I sweep house give Lek', actually 'I sweep [the] house for Lek'. The same questions of configurational structure can be asked as were applied to the preceding example, but now an additional question is, is *hāj* a verb? Clearly the meaning 'for' is close to 'give', but equally clearly, no concrete

object is being given in this instance. The answer(s) to this question have direct implications for the structural analysis as well. If *hāj* is a verb, the sentence can be analyzed as complex, containing two clauses whose structural relation must be described; it might also be monoclausal, but then one would need to defend a clause structure with two positions for the category V. If, however, *hāj* is not a verb but (say) a preposition, the sentence can only be monoclausal (omitting from consideration theories that would allow a preposition to be underlyingly verbal).

From one point of view, the concepts of single versus multiple clause and of verb versus non-verb can be seen as labels for the endpoints of a continuum rather than binary oppositions. Thus, when the serialized construction is more clearly analyzable as multi-clausal, the serialized morphemes tend to be more definitely full verbs, and the meaning involves more clearly separate events. In contrast, as the construction comes closer to being a single clause, its component morphemes are more like a single main V plus subsidiary non-verb elements (particles, prepositions, etc.), and the meaning describes a more unitary event. The continuum seems especially apt for use in tracing the historical development of serializing morphemes and for comparing morphemes with similar functions in different languages. As an instance of the first sort, it is well known that the Chinese words *bei* 'passive marker' and *ba* 'pretransitive, marker of preposed Object', which are not verbs in modern Mandarin (or only marginally so), were full verbs in earlier stages of Chinese, meaning 'cover' and 'hold, take' respectively. They

have moved toward the non-verb end of the continuum. As an instance of cross-linguistic comparison, the Thai equivalent of *bèi* is *thùuk*; however the passive construction introduced by *thùuk* is more specialized than its Chinese equivalent, being limited to application to adversative situations. Also *thùuk* may still occur as a full verb meaning 'strike (a mark), correct; be cheap'. It may thus be said that *bèi* and *thùuk* differ from each other in that the former is situated further toward the non-verbal end of the continuum than the latter.

It is often useful to speak of a single word having 'different meanings' in 'different uses', cf. the following statement by Chao (p. 332)

the verb 照 *zhào* 'shines upon, reflects' besides being a main verb... is often used in first position [in a VP-series] in the sense of 'following, according to'...

Here 'first position' is equivalent to Coverb. Unfortunately this manner of speaking is not strictly in accordance with current theories of the lexicon, in which a single entry may have only a single semantic representation and a single syntactic specification. Thus in the case of *zhào*, the entry for 'shine upon', with a syntactic feature [+V], will have to be distinct from the entry for 'according to', not only because of the difference in meaning, but also because the [Coverb NP] [V NP] construction is formally distinct from the [V NP] [V NP] serialization (Chao, 749-51, Li & Thompson 1974), hence the entry for 'according to' must be specified as [+Coverb] or [+Prep], or whatever feature is chosen to distinguish Coverbs from ordinary V's. The two entries are not entirely cut off from each other, however: they may be related by a lexical redundancy rule as in Jackendoff 1975,

or a lexical rule may derive one from the other<sup>1</sup>.

The loose-spoken concept of 'single V with several uses/meanings' could be replaced by something like 'family of entries related by redundancy rules'.

#### 4.1.2. Verb Serialization in Kayah.

Kayah is remarkable in that it makes more extensive use of concatenation, and more limited use of VP series, than is typical of a verb-medial language; this is true even in comparison to other Karen languages such as Sgaw and Pa-O. This has interesting typological and historical implications, but we will not go into them here. The Kayah Verb Complex (VC) is the locus of this extensive concatenation, and is the topic of this chapter. The VP series construction will be discussed in Chapter 6.

Kayah also is less rich than many other languages in its use of the equivalent of the Versatile Verbs of Lahu grammar. This is the type of verb mentioned above which occurs both as a single predicator and in serializations, and which in the latter case very often has a different (but related, more abstract) meaning. The verb 'give' is very commonly of this type, being also used with meanings of 'for, benefactive' (as in the Thai example above), 'cause, let', and/or 'passive marker'. Many of the more abstract functions that these perform in Lahu and Burmese, such as aspect-like notions, are in Kayah expressed by non-verbs. For example, in addition to use of the verb 'give' to show benefaction, the verb 'be at' is linked with durative or progressive aspect in Lahu, Burmese, Thai, and Chinese:

	give/benefactive	be at/durative
Lahu	pî	chê
Burmese	pei	nei
Thai	hâj	jùu
Chinese	gěi	zài

But in Kayah these are expressed by non-verbs, the V-particles *pe'* and *pa*, respectively<sup>2</sup>.

#### 4.1.3. Outline of Analysis of the VC.

In describing combinations of morphemes in the VC, we will make a first division between constructions combining two (or more) verbs, which we will abbreviate V-V, and constructions combining a verb with some morpheme that does not fully meet the definition of verb. The latter type of morpheme will be known as a Verb Particle (VPtc), and the construction will be known for short as V/Ptc, the slash indicating that the Ptc may either precede or follow the verb, depending on the characteristics of the Ptc in question.

It will become evident that some morphemes that we are forced to class as VPtc's are quite verb-like in certain respects; an example is the class of Bound Result Expressions (4.3.7). Conversely some other morphemes, which we treat as verbs with



special characteristics, might be argued to be better classified as Ptc's. These are morphemes like the Directives (4.2.3), which occur both as single predicator and in construction with (other) verbs; but with different meanings in the two cases. E.g. *n̄* 'use' (as single predicator), 'command, tell to do' (preceding a verb). Such morphemes are here presented as single, but polysemous, lexical items.; but if they were treated as separate (although related) lexical items, the item with the special use ('command' in the example just given) would fail the test of verb-hood, since it must occur before a (second) verb.

V-V constructions fall into five categories, distinguishable primarily on semantic grounds, but with certain identifying syntactic properties as well. To present a general idea of the types, I list them below together with a brief semantic description and a simple example. I use 'action' here to refer to the denotation of verbs in general; in most cases 'event' and even 'state' could be substituted.

Resultative. The first verb describes an action, the second verb describes a result of that action.

?a thú sū dī pò He wiped the pot dry  
 1 2 3 4 1 2 4 3

Descriptive. The first verb describes an action, the second verb gives an adverb-like modification of the action.

?a ?e phrē dī He eats (rice) fast  
 1 2 3 4 1 2 4 3

**Sequential.** The first verb describes an action, the second verb describes an action performed after and/or as the purpose of the first.

?a cwá ?iþri haca He went to buy clothes  
 1 2 3 4 1 2 3 4

**Directive.** The first verb describes an act of ordering or permitting, the second verb describes the content of that act, the action ordered or permitted.

?a nō dé vē thā He told me to get (dip up) water  
 1 2 3 4 5 1 2 4 3 5

**Modal.** The first verb denotes obligatoriness, or one of various other abstract conditions, pertaining to the action of the second verb.

?a be ?e dī He must eat (rice)  
 1 2 3 4 1 2 3 4

The above five categories may also be classified according to whether the two verb positions can be filled by just any verb, or at least by any of a large, unlistable number of verbs. The first three types listed are can be called 'open', in that fairly large numbers of verbs may occupy both first and second positions. The last two types can be called 'restricted' in that the verbs occurring in first position must belong to a certain listable class.

Several V-V types may co-occur, and multiple Verb Particles are common. VC's thus quite commonly consist of four or more morphemes; cf. the following, with three verbs and three VPtC's, from a recorded conversation:

(8) *chú kē ré kǎ lēkhē pè | vē | mi | to*  
 kindle burn good COM PTC BEN I fire NEG  
 They didn't burn it well for me. (Is.2)

*chú kē* is a Resultative V-V, literally 'kindle stg so it burns', which then forms a Descriptive with *ré*. See 4.3 for descriptions of the VPtc's.

#### 4.1.4. Versatility, another possible criterion for analysis.

One can also use versatility of the component morphemes as a criterion for classifying constructions. Versatility is of course redundant for grammatical morphemes like particles, which are versatile by definition, being non-free, of listable number, and occurring with all or large numbers of a certain class of free morphemes (indeed free-morpheme classes are often defined in terms of occurrence with certain particles). Non-free morphemes that are restricted are generally treated as bound members of compounds. For verbs, there is a partial correlation between versatility and construction type: the verbs of the listable classes occupying the first positions in the Directive and Modal V-V's are highly versatile, again virtually by definition, while the open class of verbs appearing in the other V-V constructions includes a great many fairly restricted members. But there are also verbs appearing in the 'open' V-V's that must be admitted to be highly versatile: in a Sequential V-V the verbs *cwá* 'go' and *hē* 'go (from home)' could precede practically any action verb; in a Resultative V-V, the verb *tō* 'hit a target, be correct' is probably similarly versatile as second verb. But the function of *cwá* in a Sequential like *cwá mé ho* 'go and peek at' is no different from that of the

less-versatile *bɔ̌ mɔ̌* 'open' in *bɔ̌ mɔ̌ mɛ́ ho lɔ̌* (41.6) 'open [a curtain] and peek at him'. I prefer to emphasize the commonality of function over the discrepancy in versatility.

Matisoff makes the opposite choice in his analysis of Lahu Versatile Verbs (as indicated by the term itself). This gives a system that is generally parallel to that described here for Kayah: the Lahu Versatile V's correspond closely to those Kayah verbs that fill the restricted positions in the restricted V-V's. It is worth noting that these Kayah and Lahu verbs share a feature in addition to simple versatility, namely a meaning shift between main-verb use and Versatile Verb use. The other Lahu verbal constructions include 'fortuitous concatenation', corresponding to the Kayah Sequential V-V, and verb plus Resultative Complement, which corresponds to the Kayah Resultative V-V. However the correspondence is not exact, since a single Kayah V-V type may contain more and less versatile verbs: for instance, a Kayah Resultative V-V with a versatile second verb like *tɔ̌* 'V the right one, V correctly' would probably count as a true concatenation, while the same V-V type with a less versatile second verb like *phé* 'crack' would be a verb + resultative complement. Also the Kayah Descriptive V-V does not fit so neatly. See the table below:

Kayah	Lahu
Sequential	fortuitous concatenation
Directive	concatenation
Modal	concatenation
Resultative	verb + resultative complement; concatenation (in a few cases)
Descriptive	uncertain: concatenation (few), adverbial expression

In some respects my decision against versatility as a major analytic criterion simply reflects the different character of Kayah grammar: the Kayah equivalents of Versatile Verbs are not as numerous or as complex in grammatical behavior as the Lahu Versatile Verbs. Regarding fortuitous concatenation (the equivalent of my Sequential V-V), Matisoff analyzes the Lahu construction as resulting from deletion of a connecting morpheme from a two-clause underlying structure. My analysis of the Kayah construction as just another type of V-V is required by my decision to operate without underlying syntactic structures (and hence no transformations).

Method of Exposition. The exposition will proceed in a 'bottom-up' direction, first describing two-morpheme constructions and the morphemes that may appear in them (verbs in 3.2, particles in 3.3), then some of the ways that the types of construction may interact to account for sequences of three and more morphemes (3.4). Having laid out the

facts in this manner, I will then consider a slightly more formalized account that goes rather in a 'top-down' direction. I will also consider whether that account is to be ascribed to syntax or morphology; i.e. whether the constructions in the VC are syntactic phrases or compounds.

## 4.2. Verb-Verb Constructions.

### 4.2.1 Resultative Constructions.

#### 4.2.1.1. General.

In this type of V-V construction, the second V gives the result of some action described by the first. Examples:

- (1) ʔū | mū sāl | pe  
 they beat die 1p  
 They beat us to death. (110.3)
- (2) p̄lò b̄ū r̄á | thwí  
 smear white PTC lime  
 (They) paint (it) white with lime (47.3)
- (3) ce li bē ū  
 dye red cloth  
 [he] dyed the cloth red. (5/2)
- (4) phō vā thā ku  
 boil cooked tea  
 [he] brewed the tea (till) it was done. (5/3)

(5) ʔa|kló bí| dī pɔ  
 3 cover closed rice pot  
 He covered the pot (with a lid)

(6) ʔa|jǒ kè | thō khri | tǎki  
 3 bend broken drum fragment a bit  
 He broke off a few pieces of the drum<sup>3</sup> (329.6)

Note that, for the one-argument verbs denoting states (*bū* 'white') or processes (*sā* 'die', *kə* 'break'), which are the most common fillers of the second position in Resultatives, appearance in that second position is the only way of giving them the causative meaning conveyed by the transitive English equivalents such as *break*, *shut*. The alternation between the transitive and intransitive versions of English verbs like *break* has been referred to as the causative-anticausative alternation. In Kayah, all causatives (in this sense) are complex expressions; from the Kayah point of view causatives like *break* and *shut* are vague since they describe a result but do not include a description of the causing event. The nearest Kayah equivalent is a Resultative with first verb *me* 'do, make', as in *me sā* 'kill', *me bī* 'shut'; however this is not available with all verbs. For instance with *mo* 'open', *me mō* sounds strange, the usual expression being *bǒ mō*, with first verb *bǒ* 'reach'.

Another point that should be made about the second verb in a Resultative (which may be called the 'result expression'): result expressions describe intended or expected result, but do not predicate that result as actually happening to the patient, although they may strongly imply it. Thus neither of the two following sentences is odd:

(7) ti nō |ʔa nō | to

stuff enter 3 enter NEG

(I tried to) stuff it in but it wouldn't go in (5/15)

(8) ʔa | chū s̄ | lū | né ʔiθoə | ma | ʔa | s̄ | to to

3 stab die 3OBJV Nɛ knife be-so 3 die NEG NEG

They stabbed him 'to death' with a knife, but he didn't die, either. (354.4)

The above clearly shows the 'adverbial' semantic quality of the Result construction (7) involves the Directional subtype of Resultative; see below). Note incidentally that *me s̄*, the closest Kayah equivalent of 'kill', is more accurately translated as 'do something murderous'. The second V, the Result expression, may be thought of as specifying a direction of the action: concrete physical direction as in (7), or abstract direction towards a result as in (8). Statement of this directional specification may be used to imply arrival at the intended/expected result, but coming to pass of that result is not included in the literal meaning of the construction. With this understanding, I will usually gloss Result constructions as if the result were asserted: e.g. 'stab to death' rather than 'stab so as to tend to cause death'.

Fillers of the V slots: Starting with the clearest cases, it can be said that the  $V_1$  slot is open to any  $V_t$ . The second verb can be any  $V_i$  that denotes a change of state; clear examples are *s̄* 'die', *bí* 'closed', *pjə* 'broken, ruined', and the various verbs of breaking, cracking, shattering, etc.  $V_i$  as first verb is also quite possible:



(9) belò| tā phé

glass fall crack

The glass fell and cracked. (10/31)

(10) sɔse| khrā pjá á

fruit dry ruin NS

The fruit dried out and was no good. (10/31)

(11) hō co pjá lō á

rice wet ruin use-up NS

The rice got all wet and was ruined. (10/31)

The second verb is not limited to change-of-state  $V_i$ 's, which presumably take

Patients as arguments; the second verb may also be an Agent-taking  $V_i$ , at least under

certain conditions:

(12) ʔa| me ngò | phúcè

3 do laugh child

He made the children laugh

(13) ʔa do mo du| me sí'ichē pè phú cè

3 beat gong big do afraid BEN child

He struck the big dong and frightened the children. (5/6)

(14) mū ŋo tāsí

beat weep horse

Beat a horse so it cries. (5/15)

(15) vĕ| phé kəthə cwà| ʔa

1s hu stand help 3

I helped him stand (by putting my arms around him).

(in the last example *cwà* 'help' is a Descriptive Particle, in construction with the V-V *phē kathō*). second verb can even be a  $V_2$ :

(16)  $v\bar{e}$  |  $p\bar{l}\bar{i}$   $cw\bar{i}$  |  $p\bar{u}$

1s whip pull ox

I whip the ox to make it pull (something).

(17)  $\text{ʔa}$  |  $\text{ʔij}\bar{e}$   $cw\bar{i}$  |  $t\bar{e}ch\bar{e}$

3 jiggle pull elephant

He 'jiggled' to make the elephant pull (something).

(*ʔijē* means to jiggle one's body while remaining in place; it is the action performed by a rider to get an elephant to move, the equivalent of 'Giddyap'). In these last examples, the result of the action is a second action, performed by a second agent. These are not, however, identical to the Directive V-V constructions (above, 4.1.3.2, and below, 4.2.3, where the difference will be discussed).

#### 4.2.1.2. Directional Constructions.

In these, the second verb gives a directional specification to the action of the first V

Examples:

(18)  $j\bar{e}$   $cw\bar{a}$   $r\bar{a}$  |  $s\bar{i}n\bar{e}$

carry(shouldr) go RA gun

[they] went carrying guns on their shoulders (235.5)

(19)  $\text{ʔa}$  |  $d\bar{e}$   $th\bar{e}$  |  $d\bar{x}$   $p\bar{l}\bar{o}$   $k\bar{u}$  |  $d\bar{x}$   $phr\bar{e}$   $khu$

3 put go-up at box in at shelf on

They put (it) up in a box on a shelf. (326.1)

(20) təlʌ bɔ̄ vī rál| cō the nì|lū hi  
 whirlwind PTC lift go-up get 3 house  
 The whirlwind lifted up their house. (42.7)

(21) ʔa se|khrā the á  
 fruit dry go-out NS  
 The fruit got drier. (92.6)

Notice that the direction can be metaphorical, as in (21).

Directional constructions can be considered to be a special type of Result construction. Semantically the direction can usually be understood as resulting from the action of the first V, as in (19) and (20), although in many cases it is not so clear that there is a true causal relation, as in (18) (*jé cwá* 'go carrying') and such expressions as *dé the* 'ride (vehicle, animal) up', and *lo hē* 'float away, go floating'. Syntactically, Directional expressions can co-occur with (other) Result expressions, which suggests that they form a separate type, but they may either precede or follow the other Result expressions, which suggests that the two are similar if not identical:

V Dr R

(22) ne|cwi the thū|ʔacʌ|nʌ  
 2 pull go-out long wick Nʌ  
 (if) you pull the wick out long (341.6)

V Dr R

(23) vē|me tā phra  
 1s do fall to-sound  
 I dropped (stg) so it made noise.

	V	R	Dr
(24)	bó	mō	the
	reach	open	go-out
	open (stg)		outwards

Directional constructions are also like Result constructions in that both may be embedded in a Sequence construction (4.2.2).

Directional constructions differ from Result constructions in that the V's appearing in the second position, the 'Directional' V's, form a closed class, which we will list below. This difference may, however, be a matter of degree: it may be equivalent to saying that relatively few V's specify Locative, while relatively many have whatever feature is chosen to reflect their potential for appearance as second V in the 'general' Result constructions: [+Patient], [+change of state], or some other. There are several types of Directional verbs.

#### Type A Directionals

cwá go

hē go (from home)

ka come, go (to home)

None of these has the same deictic orientation as English 'come' and 'go', namely 'motion towards/away from the speaker or other center of interest'. As main V, *cwá* usually has no deictic connotation at all; as Directional it may (but does not necessarily) mean 'away', as in *kō cwá* 'blow stg away'(of wind). *ka* and *hē* are paired, as in such



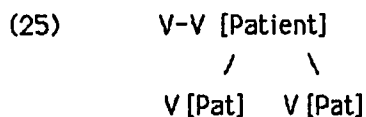
a curve; (2) go on a winding road (cf. also *kajē takā* 'dishonest ('crooked') person').

Type B Directionals may be second verb with a Type A as first verb: *cwá the* 'go up'. Most type B's may take the prefixes *kə-* 'moving Subject' and *pə-* 'orientation (no transit of space)'. For the former, note the contrast *phjá the* 'take stg out (of a container)' versus *phjá kathe* 'take stg out (of a house)'. In the former only the entity denoted by the Object moves out, in the latter both Subject and Object move out.

While the *kə-*prefixed forms are verbs, the forms with prefix *pə-* are not; see (4.3.7.2) below.

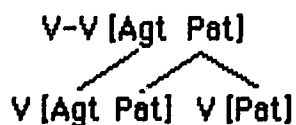
#### Derived Argument Structure: Resultatives

In cases such as (9-10), with two one-argument verbs, the single argument of the sentence has the Patient role specified by each of the verbs:



Examples (1-6) represent a canonical or prototypic pattern, with two-argument verb followed by one-argument verb; they have the derived argument structure already discussed in 3.4 above:

(26)



Examples (12-15) resemble this canonical type in consisting of two-argument verb

plus one-argument verb and in the coincidence between the Patient specified by the first verb and the single role specified by the second verb. That single role, however, is an Agent in (12-15):

- (27)           V-V [Agt   Pat]  
                   /     /   \  
                   V [Agt Pat] V [Agt]

The case exemplified by (16-17), in which both verbs specify two arguments, is problematic, and will be discussed in more detail in connection with Directive V-V's below (4.2.3).

Finally we may mention another variant of the canonical type. Examples are

- (28) vĕ ʔīchi síphrá khru  
 1s split tired firewood  
 I got tired splitting firewood;  
 I split firewood till I was tired. (8/9)
- (29) ʔa dū là síphrá hi dós kū  
 3 sweep clear tired house wall inside  
 He got tired sweeping the house;  
 he swept until he was tired. (8/9)
- (30) ne méthɫ mo [ Phēiúdu me hú ʔā ] to  
 2 see happy (name) do like this NEG  
 You are unhappy seeing P. act like that;  
 Seeing P. act like that makes you unhappy. (11/21)

- (31) *vē mé mo ne to*  
 1s look happy you NEG  
 I feel sorry for you; I pity you;  
 (literally,) I am unhappy seeing you[r condition]. (2/27)
- (32) *ʔa ʔō mu thāʔíphrè*  
 3 drink drunk whiskey  
 He got drunk on whiskey. (common expression)

The variation here is that, although these are canonical in having  $V_t-V_i$ , the result expressed by the second verb applies to the Subject of the sentence rather than the object. This is especially striking in (32), in which both Subject and Object NP's denote entities that could plausibly have *mo* 'happy, comfortable' predicated of them. But the sentence does not mean 'I see that you are unhappy'; the informant glossed *mé mo to* as Thai *sǎnsǎn* 'to pity'. The derived argument structure would be, e.g.:

- (33)
- ʔō mu [Agent Patient]  
 /                    \  
 ʔō [Agt Pat] mu [Pat]

This type is relatively scarce; see 5.3.3 for some further discussion.

Derived argument structure: Directionals. Generally the derived argument structure of Directional V-V's is like that of (other) Resultatives, in that an argument (usually the single argument) of  $V_2$  is associated with the single argument of a one-argument  $V_1$  and with the Patient of a two-argument  $V_1$ . However, in the case of the general Resultatives



the second verb is typically a one-argument verb specifying Patient. Directional verbs, in contrast, specify two roles: one participant that moves, and a second participant that is the goal (or, less commonly, source) of the motion. The second argument, which can be called (inner) Locative, is unproblematic in derived argument structure: it always is realized as a PP<sup>5</sup>, identical with both the specified Locative argument, if any, of V<sub>1</sub>, and with the non-specified (outer) Locative adjunct possible with all verbs.

The first argument, the moving entity, will be called Patient (passing over the question of possible Agent-hood when the motion is voluntarily performed by an animate participant); it generally behaves similarly to the single argument of the second verb in the 'canonical' Result V-V. That is to say, it is associated with the single argument of a one-argument first verb, as in:

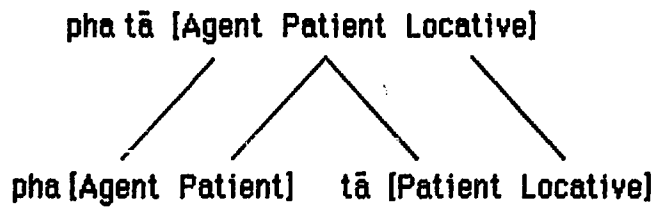
- (34) [pí pè] jo cwá rÁ  
 butterfly fly go RA  
 They [butterflies] flew away. (358.3)

With a two-argument first verb, the Patient of the second verb usually is associated with the Patient of the first verb, as in (19-20), and:

- (35) ?ū pha tā di ?iswí  
 3 drop fall food  
 They drop in food [into a pond, as an offering]. (207.2)

In (35) the food moves downwards, while the Agent 'they' does not; the derived argument structure would be:

(36)



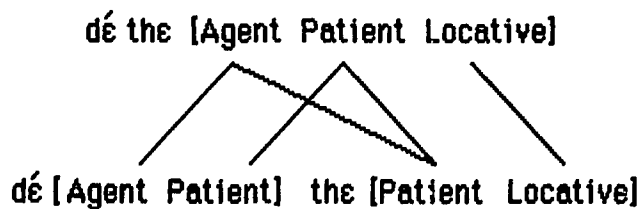
Here, with the exception of the presence of the Locative argument, the derived argument structures are the same as those for general Resultatives; compare (36) with (26). There are two classes of apparent exceptions to this similarity.

The first class of exceptions is made up of sentences with Directional V-V's in which both the Subject (Agent) and the Object (Patient) are in motion, e.g.:

- (37) ʔa d̄é the t̄así d̄é cho kh̄ā  
 3 ride ascend horse at mountain top-of  
 He rode the horse up the mountain.

Here both the rider and the horse move upwards. It would be possible to posit a derived argument structure showing this directly by associating the Patient argument of *the* with two arguments in the derived argument structure, as:

(38)



However, it is doubtful whether such a complex mapping should be allowed: note that

this is the first proposed derived argument structure with more than one line of association originating from a single argument in a component argument structure. It is perhaps best to not have the derived argument structure reflect the information that the Agent moves. That information could be assumed to be available to speakers as inference from knowledge of the meaning of certain verbs: besides *dĕ́*, only *vī́* 'drive' and *kīĒ* 'paddle' are known at this time. The lexical entries of these verbs could include some explicit notation to the effect that they represent actions performed on some conveyance while mounted on that conveyance; it would then follow that any motion attributed to the conveyance would apply to the rider as well. With this sort of treatment, this class of sentences can be analyzed as no different from others with Directional V-V's.

The second class of exceptions to the similarity of derived argument structure between Directionals and Resultatives has already been mentioned, namely those in which the direction is 'metaphoric', as (29), repeated below, and also:

- (21) ʔa se | khrā the á  
 fruit dry go-out NS  
 the fruit got drier (92.6)
- (39) khĒ bō jĕ the nλ ke  
 neck swell go-out Nλ PRH  
 Maybe his neck swelled up. (308.5)
- (40) ʔa bĕ the ʔíkhu nλ  
 3 mold go-up earth Nλ  
 He [God] molded up the earth. (337.3)

In (21) and (39) it is clear that no participant moves outwards; rather *the* is used in a metaphoric sense that could be rendered as 'increasingly'. Likewise *the* in (40) does not mean that the earth moves upwards: the sentence is from a creation-myth, and describes God creating the earth (*bē* means to mold something formless, as mud or concrete, into a form). *the* then is to be translated as 'coming into existence'.

Since no participant moves, the derived argument structure for these examples should have no line associating the Patient argument of the second verb (the moving entity) with anything in the derived argument structure. They are in fact quite similar to Descriptive V-V's (4.2.4 below, where derived argument structure is discussed more fully) and should probably be classed with them. Note that the Directional verbs we have seen in such 'metaphoric' constructions are limited to the two *the* and *the*. More research is needed to determine whether other Directional verbs can be used in this way.

#### 4.2.2. Sequential Constructions (V1-V2).

These consist of several V's (usually two, but occasionally more) that describe a sequence of actions linked by temporal order and/or purpose, usually paraphrasable as 'V1 and then V2' or 'V1 in order to V2'. Examples:

(41) ?a | ka déhā rā | ?aphā

3 return ask R<sub>A</sub> grandmother

He went back and asked his grandmother. (93.6)

- (42) hē ʔo māl klè| tλ nāl nλ  
 go sleep cut one two day(CLF)  
 [We] go and sleep (in the fields) and cut (brush)  
 for one or two days. (177.2)
- (43) ne hē nì dā mé kāl mλ  
 2 come listen look COM IMP  
 Come listen and look! (93.6)
- (44) thwi| mé thλ ʔe á| li phá| nλ  
 dog see eat NS book skin Nλ  
 The dog saw the hide book and ate it. (103.4)
- (45) ʔo nē ʔíbe chú lū ma mo mo á  
 sit speak confronting mutually be-so fun fun NS  
 Sitting and talking together is fun, too. (180.4)

Notice that (42) and (43) contain three verbs in the Sequential construction; note also that the relation between the actions they denote is better described as 'alternating action' rather than 'sequential action', since it is not the case that all instances of sleeping will precede all instances of cutting; rather, they will alternate. (45) is different again: sitting and speaking are simultaneous. The informant pointed out that the sentence (42) could also have the (nonsensical) reading of simultaneous action, 'sleep while cutting'. Just such a range of meanings is described for the Chinese 'serial Verb' by Li and Thompson (1973): purpose, sequential action, alternating actions, and simultaneous actions.

The sequential units may themselves be Resultative constructions:

- (46) *vē| pù me sāl jò khró*  
 1s catch do die rat  
 I caught and killed a rat (10/8)

- (47) *ʔa cwá kəthə phjá kwa*  
 3 go go-up take axe  
 He went up and took the axe.

Additionally, the 'sequenced' elements in (42) and (44) include two Resultative compounds: *ʔo māl* 'sleep' is made up of *ʔo* 'be at' and *māl* 'sleep, lie down', and *mé thm* 'see' is made up of *mé* 'look' and *thm* 'see'. Notice also that the sequenced element may be a Directional construction, as in (47), which supports the treatment of Directional constructions as a subtype of Resultative; cf. also:

- (48) *ʔa thə phjá tā*  
 3 go-up take go-down  
 They went up and took [it] down. (326.4)

Thus, the sequential units may be either single V's or Resultative constructions.

Although the paraphrase 'V1 and then V2' is usually associated with a bi-clausal structure, as pointed out above (4.1.2), the Kayah Sequential construction must be analyzed as part of a single, non-complex clause, for two reasons. First, the various Verb Particles precede or follow Sequential constructions just as they do simple V's, e.g.:

- (49) ʔa mò | tú hē nō dé pa | ʔa pò | thā  
 mother just-now go command dip-up DUR YS water  
 Mother just now went and told YS to draw water

Second, the Sequential V-V unit may take only one Object NP; if both verbs in the construction are transitive and can be interpreted as taking different Objects, only the Object associated with the second verb may appear:

- (50a) ʔū | bó mō mé ho | lū  
 3 open look secretly 3OBV  
 Secretly they opened it (mosquito net)  
 and peeked at him. (41.6)
- (50b) \*ʔū | bó mō mé ho | ʔikē thō  
 mosquito-net  
 Secretly they opened the mosquito net  
 and peeked (at him).
- (51a) vē | chíjā plwā | thé  
 1s untie release pig  
 I untied it (rope) and released the pig. (3/3)
- (51b) \*vē | chí jā plwā | sú plā  
 rope  
 I untied the rope and released it (pig).

Note also that *ho* 'secretly' in (50a-b) (as is clear from the context) modifies *bó mō* *mé* 'open and look' as a unit (*ho* is here functioning as a 'Descriptive Particle'; see 4.3.6 below).

Derived Argument Structure. The essential feature of Sequential V-V's is identity of Agents; the V-V in the hypothetical sentence (52) would have the derived argument structure (53):

- (52) ʔa sí ηjā ʔe dī  
 3 laugh eat cooked-rice  
 He laughs and eats.

- (53)                    sí ηjā ʔe [ Agt Pat ]  
                               /    \    \  
                              sí ηjā [ Agt] ʔe [Agt Pat]

For an actually-recorded example, consider (41), repeated below:

- (41) ʔa | ka déhā rál | ʔaphā  
 3 return ask RA grandmother  
 He went back and asked his grandmother (93.6)

This is perhaps the most typical sort of Sequential V-V, in which the first verb is a verb of motion; as such, it must specify Locative as well as Agent. The Locative is presumably simply added to the derived argument structure:

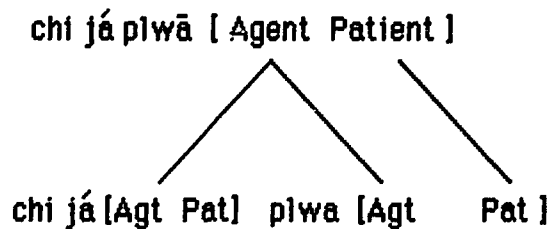
- (54)                    ka dé hā [Loc Agent X]    'go back and ask'  
                               /   /    \   \  
                              ka [Loc Agt] dé hā [Agt X] 'go back' + 'ask'

Finally, we may consider the argument structure of the case in which both verbs specify Patient, and hence potentially two Object NP's, as in (50a), repeated below:



- (50a) *vē* | *chi já plwā* | *thé*  
 I untie release pig  
 I untied it (rope) and released the pig. (3/3)

(55)



The important point is that the Patient argument of the first verb is not associated with anything in the derived argument structure, reflecting the fact that that participant cannot occur as an NP in the sentence, as discussed.

Which verbs may appear in a Sequential V-V:

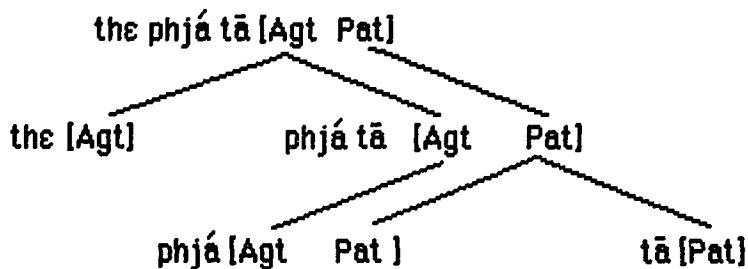
Probably the most common first verbs are verbs of motion and posture, as in (41), (42), and (47). *chwā* 'go' is very common; in fact it probably has a derived form *jā* that occurs only as  $V_1$  in Sequential V-V's (see 4.3.2.3 below). Also common is the verb *lw* 'go after, follow to overtake, go to get (usually to get some person)'. However, many other verbs may appear as first verb, and the second verb position is even less restricted. Among possible combinations,  $V_i$ - $V_i$  and  $V_i$ - $V_t$  are fairly common, while  $V_t$ - $V_t$  is perhaps somewhat less frequent.  $V_t$ - $V_i$  is also possible, as this example shows:

- (56) \*a phjá kəthe Phē tuə ʔaphē hɛ təpu  
 3 take go-up (name) pants one-clf  
 He took P's pants and went up [with them]. (313.5)

Note that the second verb, with prefix *kə-*, specifically means that the Agent went up, and that this is the reason for analyzing (56) as containing a Sequential V-V rather than a Directional V-V. The equivalent Directional construction would be *phjá the*, with unprefix second verb, and would mean that the Agent takes something which moves upwards as a result.

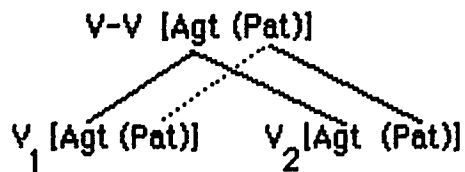
The constructions in (46-48), in which V-V's themselves consist of V-V's can be given a multi-level derived argument structure representation, as for (48) (omitting the Locative arguments of *the* and *tā*):

(57)



A generalized derived argument structure for Sequential V-V's may be posited as:

(58)



The dotted line shows optional mapping: the Patient of the  $V_1$  must be either identified with the Patient of  $V_2$  or else be given no realization at all.

#### 4.2.3. Directive V-V's.

In these, the first is one of a relatively small set of verbs denoting commanding or permitting; what Searle (1975) classifies as the 'directive' type of illocutionary acts. To paraphrase Searle, the verbs denote situations in which one person gets another person to do something, whether the second person doesn't especially want to do it (as in commanding), or whether s/he does (as in permitting). Examples:

(59) ʔa nō pā phú cè mē k lú

3 command cut child rhythm-pipe

She told the children to cut rhythm-pipes. (42.5)

(60) vē dǎ cwá ne to

1s let go 2 NEG

I won't let you go. (3/31)

(61) ʔa phē nλ ʔé cwá vē lū jò du

3 father Nλ call go dig 3OBV rat big

Her father called to her to go dig out a big rat. (36.3)

The class listed:

nō	tell to V, get sbdy to V; (as main V) use (cf. Thai chaj, with the same range of meaning)
dá	let V; (as main V) give
jo ~ jo nō	forbid
ʔe ~ ʔé nō	call to V
jū	point, order by pointing
dε ke	hire to V, employ
ʔīswá	teach to V

Note that, with the exception of the last-mentioned, these all refer to acts of communication, usually verbal.

As (61) shows, the 'second verb' in a Directive V-V may itself be a V-V: a Sequential as in (61) (*ɕwá vē* 'go and dig'), or a Resultative as in:

(62)	ʔa	nō	kú vē	the	lū	he so
	3	command	dig	go-out	3OBV	dust
	He told her to dig out the dust. (36.4)					

Here the Directive *nō* is followed by the Resultative V-V *kú ve the* 'dig out'.

As with other V-V's, the structure of clauses with Directives is fundamentally identical to that of clauses with simple predicators. There is no evidence of embedding or any other sort of multi-clause structure. The argument relations to a VC including

[Directive +  $V_i$ ] are no different from those to a VC with a simple  $V_t$ , and the same

identity holds between [Directive +  $V_t$ ] and simple  $V_{dt}$ , as can be seen by comparing the

following with (1) above:

- (63)  $v\bar{e}$  |  $d\acute{a}$  |  $\text{ʔa}$  |  $r\bar{u}$   
 Is give 3 money  
 I gave him money

In both (59) and (63) the first NP is Subject = Agent, the second post-VC NP is Obj-2 = Patient, and the first post-VC NP is Obj-1 = Recipient (subsuming recipient of goods and recipient of causation). Only one sort of negation is possible, with the Clause Ptc *to*, although this may have varying semantic outcomes. The negative version of (59) would be *\text{ʔa } n\bar{o} p\bar{a} ph\acute{u} c\bar{e} m\bar{e} k\acute{l}\acute{u} to* 'she told the children not to cut rhythm-pipes'. The difference between *d\acute{a} V to* 'not let sbdy V' and *n\bar{o} V to* 'tell sbdy not to V' is probably not to be attributed to any difference in syntactic structure, but to differing semantic interactions between the clause-level negative *to* and the various Directive verbs.

Directives have scope over any following Resultative and Sequential V-V's. A Directive V-V can also function as a unit to be the second constituent of a Sequential V-V. This means that the following structures are possible:

- (64) Drv [V V]<sub>Seq</sub>  $n\bar{o}$        $k\bar{a}l\bar{e}$        $d\acute{e}$        $th\bar{a}$   
 command go-down dip-up water  
 Tell sbdy to do down and draw water.

- (65) V [Drv V] hē nō dé | thā  
 go command dip-up water  
 Go tell sbdy to draw water.

What is not possible is a Sequential V-V with the V1 constituent made up of a Drv-V,

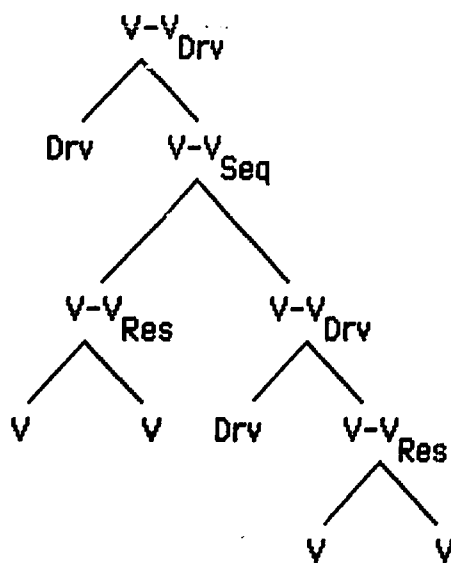
as:

- (66) [Drv V]<sub>V1</sub> V<sub>2</sub> \*vē | nō cwá mū | ?a  
 I command go hit 3  
 \*I told him<sub>i</sub> to go and (I) hit him<sub>i</sub>

(66) would probably be acceptable as 'I told him to go and (to) hit sbdy', i.e. Drv [V V]

These facts can be summed up by the following maximal structure:

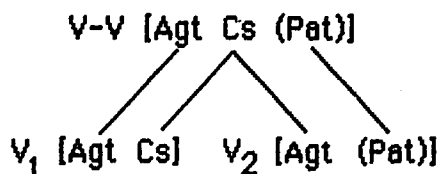
(67)



Condition: only one Directive verb may appear.

Derived Argument Structure. The distinctive feature of the derived argument structure of Directive V-V's is the identity between the Agent of the second verb and the non-Agent role of the first verb (Cs=Causee):

(68)



As was mentioned above, Causee may be identified with the Recipient role. In support of this identification, note that with a  $V_t$  as second verb, the Causee is realized as Obj-1 as is Recipient with  $V_d$ 's. In addition, when the second verb is itself a  $V_d$ , the Causee appears as Obj-1, while the Recipient originally specified by the second verb must be realized in an Oblique PP with  $n\acute{e}$ :

(69)  $v\acute{e}$  |  $n\bar{o}$        $d\acute{\alpha}$  |  $Ph\bar{\alpha}\lambda$  |  $r\bar{u}$  |  $n\acute{e}$   $\text{?ap}\bar{o}$   
 1s command give P. money Nc YS  
 I told P. to give money to [his] younger sibling.

Here  $Ph\bar{\alpha}\lambda$  is the Causee specified by the Directive verb  $n\bar{o}$  as well as the Agent of  $d\acute{\alpha}$ ;  $r\bar{u}$  is the Patient of  $d\acute{\alpha}$ ; and  $\text{?ap}\bar{o}$  is the Recipient specified by  $d\acute{\alpha}$ . Or rather, it would have been the Recipient specified by  $d\acute{\alpha}$ : it is precisely the function of  $n\acute{e}$  to indicate participants that are peripheralized; in this case  $\text{?ap}\bar{o}$  has been pre-empted for the role of Recipient and the position of Obj-1 by the Causee  $Ph\bar{\alpha}\lambda$ . This fact can be

captured by saying that both the Directive verb and the second verb specify Recipient, but the role specified by the Directive verb takes precedence and forces the second verb's role to appear as an Oblique relation (if it appears at all).

At this point we can return to the matter of Resultative V-V's whose second verb is

$V_T$  Recall (16) and (17), repeated below:

(16)  $v\bar{e} | p\bar{l}\bar{i} \quad cw\bar{i} | p\bar{u}$

1s whip pull ox

I whip the ox to make it pull (something).

(17)  $\text{ʔa} | \text{ʔi}\bar{j}\bar{e} \quad cw\bar{i} | \text{tach}\bar{e}$

3 jiggle pull elephant

He 'jiggled' to make the elephant pull (something).

In meaning these are very close to Directive constructions:  $V_1$  denotes an action that causes a sentient being, denoted by a post-verbal NP, to perform the action denoted by  $V_2$ . Although I have defined Directive verbs as denoting acts of communication, and the first verbs in (16-17) ('whip' and 'jiggle') are not obviously verbs of communication (especially the former), it might be claimed that they are in this instance given special interpretations as involving communication. Nevertheless there is a property distinguishing these from Directive V-V's, namely the realization of the arguments of the second verb. The Patient of the second verb in a Directive construction regularly appears as Obj-2. In contrast, if the Patient of the second verb in a Resultative construction



appears in the sentence, it must be in an Oblique PP with *né*:

(70) ?a me cwi panè né ?íthá  
 3 do pull buffalo Ne plow  
 He made the buffalo pull the plow. (----)

(71) ?a ?íjē cwi tēchē né sɔ  
 3 jiggle pull elephant Ne tree  
 He 'jogged' the elephant into pulling the log. (—)

By the same token, in the Directive (69) the Agent argument of the second verb comes out as Causee, in the grammatical relation Obj-1; while in the Resultative (17) the second verb's Agent clearly has the grammatical relation Obj-x, as is shown by the impossibility of a second Object relation:

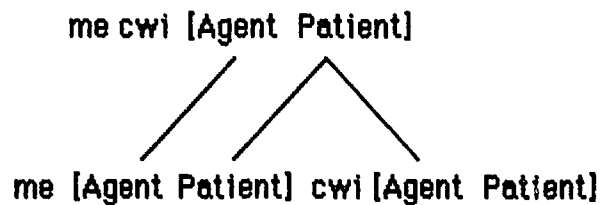
(17a) \*?a ?íjē cwi tēchē sɔ

Because of this, we cannot impute the Causee role to the participant in question: it must be a Patient, as is consistent with an interpretation of the VC as a Resultative V-V (recall that the defining characteristic of Resultatives is mapping of a role of  $V_2$  to the Patient role of  $V_1$ ). The only difficulty is the unusual mapping of Agent (of constituent) to Patient (of derived argument structure).

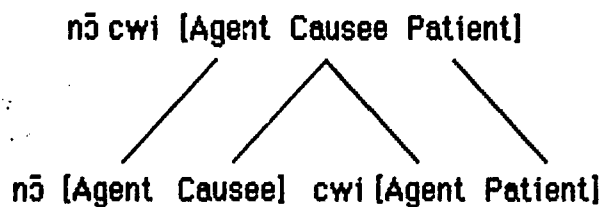
The difference in argument structure between the Directive and Resultative can be displayed by comparing (9) with a hypothetical Directive equivalent such as *pa n̄ cwi*

*Phān s̄* 'he told P. to pull the log':

(72) Resultative



(73) Directive



From the two different argument structures follow two different syntactic structures: the Directive *n̄ cwi* appears with Subject, Obj-1, and Obj-2, while the Resultative *me cwi* has Subject, Object, and Oblique PP with *n̄*

Notice that in the Resultative the derived argument structure does not have anything that associates with the Patient role of the second verb *cwi*. The participant that might otherwise have that role can indeed have a realization in the sentence in the Oblique *n̄*-PP, as in (9-10); however in such cases it can no longer be said to 'have' the Patient role. This analysis is partly prompted by theoretical considerations: neither a clause nor an argument structure may have two Patients. In addition, it seems to be the function of *n̄*

PP's to mark entities that by intrinsic quality might serve as participants in actions, but that in their particular sentences are 'backgrounded' to the extent that they do not have any role strictly speaking (see section 6.5 below). This, then, is the reason that the Patient role of *cwi* is mapped to nothing in (72).

The differences between the three types of V-V described so far can be summed up in terms of argument structure as follows:

<u>Type</u>	<u>role of V<sub>1</sub></u>		<u>role of V<sub>2</sub></u>
Directive	Recipient (Causee)	<i>mapped to</i>	Agent
Resultative	Patient	<i>mapped to</i>	single role of V <sub>i</sub> , Agent of V <sub>t</sub>
Sequential	Agent	<i>mapped to</i>	Agent

The kinship between Directives and Resultatives is shown here in their contrast with Sequential V-V's: the latter map Agent to Agent, while both of the former map non-Agent to either Agent or single non-Agent (I omit consideration of the rare Resultative type represented by *ʔō mw* 'drink-get drunk', where Agent of V<sub>1</sub> is mapped to Patient of V<sub>2</sub>).

#### 4.2.4. Descriptive V-V's.

The remaining two types of V-V, Descriptive and Modal (following section) are defined negatively, in part: for neither does a derived argument structure analysis seem appropriate. The two are distinguished from each other chiefly by relative position. In a

Descriptive construction, a verb is followed by a second verb that modifies it (in a sense to be discussed shortly); in a Modal construction a verb is preceded by a first verb with modal or quasi-modal meaning. In other words, the distinction is in terms of whether the more open slot precedes or follows the more restricted slot.

Examples of Descriptive V-V's:

(74) ʔa ʔe phrē dī

3 eat fast rice

He eats quickly.

(75) ʔa me ré rǎ lū

3 do good RA 3obv

They were good to them. (235.7)

(76) nì dā jè

listen difficult

It's hard to listen to ≈ it sounds funny

(said of unacceptable utterances)

To elucidate the semantic relation between the first and second verbs in a Descriptive V-V, consider examples of Descriptives as contrasted with Resultative V-V's. Recall that in a (canonical) Resultative, the Patient argument of  $V_2$  is associated with the Patient argument of  $V_1$ , such that a sentence  $[NP_1 V_1 V_2 NP_2]$  implies the

sentence [NP V<sub>2</sub>]; e.g. ʔū | chū sāl | ʔa 'they stabbed him to death' implies (but does not entail, as discussed in 4.2.1) ʔa sāl 'he died'. There is no such implication in a Descriptive V-V, as can be seen by comparing the sentence ʔa | ʔe phrē | dī 'he eats (rice) fast', which by no means implies dī | phrē 'rice is fast'. Another neat example involves *ré* 'good', one of the few V's that easily occurs as second verb in both Resultative and Descriptive V-V's:

(77) (Resultative) vē | me ré twà kāl rál | vē hi

1s do good beautiful COM Rλ 1s house

I (would) improve (make good and beautiful) my house (181.3)

(78) (Descriptive) ʔa | me ré rál | iū

3 do good Rλ 3OBV

They were good to (did well towards) them (235.7)

In (77), *ré* 'good' describes the result of the action of *me* 'do, make', the result applying to the Object *hi* 'house'. In (78), *ré* describes a quality of the action of *me*, and it does not apply to *iū*: 'they' did not become good as a result of the action<sup>6</sup>.

Ambiguous in the same way is *ʔibe jé* 'speak + difficult', which may mean either 'speak with difficulty, hard to say' or 'speak difficulties → make trouble [by speaking]'. It is plausible to construe the first reading as the Descriptive construction (speak in a difficult manner) and the second as the Resultative (speak so that something becomes difficult).

The 'metaphoric' use of the Directional verbs *the* 'go out; increasingly' and *the* 'go up; coming into existence', discussed at the end of (4.2.1.2) above, probably qualifies them as further examples of verbs that can occur in Descriptive V-V's as well as Resultatives.

Below we will list the verbs appearing as V<sub>2</sub> in Descriptives. The list is almost certainly not exhaustive; there are also a great many non-verbs that resemble these verbs in meaning and syntactic position (Descriptive Particles, 4.3.6).

sú	wrongly, to V the wrong one.
phrē	fast, quickly.
jō	slow to cover distance (cf. pajè 'slow to accomplish')
sē	new, anew, over again.
ré	good, well, carefully.
mo	enjoyably, enjoy V-ing; (as main V) feel good, healthy, have fun.
ηjā	for a long time, last a long time.
lō	exhaustively, including all of a set; (as main V) use up, spend; similar to Thai <i>mòt</i> .
ro	early in the morning.
p ē	late in the morning.
cha phoə	early in the evening, before dusk.
hé	in the evening, late (as compared to afternoon)
ʔé	many, affecting many things; (as main V) be many, much.

bé                    have the wherewithal to V; (as main V) be rich,  
well-endowed.

jè                    difficult

jū                    easy

and probably:

the                    increasingly; (as main verb) go out.

the                    coming into existence; (as main verb) go up.

Two morphemes that frequently follow verbs in a Descriptive-like construction are not easily classifiable as verb or Particle: *rà* 'beforehand' and *no* 'afterward'. Examples:

(79) ?ibe    cè    rà,        cwá cè    no  
      speak able beforeh. go    able afterw.  
      [He's] able to talk first, able to walk later. (c)

The meaning of these morphemes in such constructions is like that of  $V_2$  in a Descriptive V-V, in that it is purely adverbial, and does not have an attached grammatical position which can express the temporal point of reference (i.e. which can answer the question 'before/after what?'). They are thus unlike prepositions/subordinating conjunctions such as English 'before' and 'after', with which the point of reference can be expressed as a complement clause or object of preposition. The most natural equivalent of English *I left after he came* does not subordinate the clause *he came* to the equivalent of *after*, but places it first with a perfective marker:

(80) ʔū hē thō á vē cwá no

3 come finish NS 1s go afterw.

He having come, I went afterwards=After he came, I left. (c)

The problem with the two Kayah morphemes is that they do not satisfy the primary criteria for full V's, being unable to function as complete predicators or to appear on their own in construction with the V Ptc's. They do, however, appear in two positions otherwise restricted to or typical of V's: embedded as nominalized objects of Prepositions, and modifying a preceding noun. For the first, we may cite the two very common expressions *dé ʔa rā* 'recently', with the general-purpose Preposition *dé*, and *chá nō pā* 'later on', with the future-time Preposition *chá* and the Clause Ptc *pā* 'unrealized situation'. It is of course possible that in such cases *rā* and *nō* are N's, since that is the other form class that may be Object of a Preposition. For the second, there is at least one example of one of them modifying a preceding noun: *phú rā* 'previous child' (i.e. child by a previous marriage) e.g. *ʔa phú rā ʔo sí nā* 's/he has two previous children. This is a characteristic of verbs (and of nouns in certain construction types) but not of Particles. The case for treating them as V's is slightly better, since

- i. ability to appear in the post-verbal Descriptive-like construction is not a characteristic of nouns;
- ii. ability to modify nouns is not a characteristic of particles;
- iii. ability to be followed by the Clause Particle *pā* (as in *chá nō pā*) is not a characteristic of nouns or particles.



The difficulty of giving a derived argument structure analysis of Descriptive V-V's can now be seen: for example in (74), repeated below

(74) ?a ?e phrē dī            He eats quickly.

there are two verbs, which presumably have argument structures like this:

?e 'eat' [Agent Patient]

phrē 'be fast' [Patient]

The problem is, of course, that in a sentence like (1) there is no element that can be said to have the Patient role specified by *phrē*. If *phrē* is a predicate, it predicates quickness, not of any participant in the action, but of the action as a whole. If it has an argument, that argument must be either the whole proposition 'he eats (rice)' or perhaps just the action 'eat'. But it is far from obvious that such facts are best captured by saying that the linguistic elements *?a ?e dī* or *?e* bear the Patient role of *phrē*. We will return to this topic in 5.5 below.

As to the fillers of the first and second slots, the only clear restriction is that the second verb must be a one-argument verb, and that it cannot specify Agent. The first verb slot seems completely open. The interesting question here is how to reflect the difference between Resultative and Descriptive constructions. In practice, the sets of verbs occurring as second verb in the two types seem to be largely disjoint. We may ask,

(i) is there a principle, representable by a feature, capable of determining which

Verbs may appear as  $V_2$  in Descriptives?

(ii) is that principle broad enough to make the class of such verbs an open class?

One way to characterize the difference could be based on the observation that the  $V_1$ 's that appear as second constituents in Resultative V-V's tend to denote processes, or states that result from processes; e.g. *phé* could be rendered either as 'to crack' (process of cracking) or as 'be cracked' (result of process of cracking). Those  $V_1$  that appear as second constituents of Descriptives, in contrast, tend to denote states only, or rather states that are typically not thought of as resulting from processes; e.g. *phrē* 'fast' (state), but could probably not often be rendered 'become fast, speed up' (result of process). This is essentially the distinction drawn by Chafe (1970, 124) between intrinsic processes ('break' cited as example) and intrinsic states ('wide' and 'be open' cited as examples). It also seems possible for there to be V's that are fairly neutral between state and result-of-process; *ré* 'good', which appears naturally as both Ds and Result expression (4-5 above), would be an example.

#### 4.2.5. Modal V-V's.

In these the first verb has a meaning relating either to modality, in the sense of obligation, or to various notions that can be rather vaguely characterized as having to do

with the emotional setting of the action, often in terms of the mental attitude of the Agent participant. The term 'modal' is thus used as a convenient label rather than in the more usual technical sense; 'quasi-modal' might be better for the mental-attitude type.

The first verb is one of a closed class, known as Modal verbs; some of these have little meaning shift from their single-predicator use:

sí plɔ̄ tō	like; like to V
dé sí plɔ̄	decide; decide to V
təne	think; think to V, intend to V
kha	promise
do	abstain, swear off
je	make as if to V, threaten to V

The first two are frozen compound forms, containing sí plɔ̄ 'heart', tō 'strike a target' (cf. Thai thùuk caj), and dé 'put'.

Others have notable shift (single-predicator meaning in parentheses):

me	(do, make) try to do something undesirable or prohibited
be	(impinge, affect, become manifest) must, should
tō	(be correct, strike accurately) should, time is right to V
ré	(good) should be the case that, you'd have thought that, counterfactual

Examples:

- (81) ʔíre nʌ ma | vĕ̄ kha | ʔíre du á  
 work Nʌ be-so 1s promise work PTC NS  
 As to work, I promise to work myself. (84.1)

- (82) ʔa do ʔō thā ʔiphɾè  
 ʔ abstain drink whiskey  
 He gave up drinking. (10/13)

*do* as single predicator:

- (83) bó ré to ma ʔū do  
 rice good NEG be-so ʔ abstain  
 I'm abstaining because the rice [crop] isn't good. (175.1)

(note use of remote third-person pronoun *ʔū* for humilific self-reference). *kha* as single predicator may be found in *kha no* 'go into debt', but the category of *no* is uncertain.

#### Argument structure and constituency.

As with the Descriptive construction, positing a derived argument structure seems not to be the most natural way to analyze this type of V-V. In both cases, the semantic relation between the two verbs seems to be more like modification than like cause-effect or temporal (or other) sequence. This is reflected by the fact that the argument structure of a Modal V-V is in all cases identical with the argument structure of the second verb: the modal does not add to or otherwise alter the argument structure of its co-constituent. For some of the Modal verbs, the modification is reminiscent of the

relation between operators and predicates in logic.

Use of the name 'Modal' raises the possibility that these verbs are like modal auxiliaries in ways that go beyond their meaning: for instance, auxiliary verbs are often analyzed as being in constituency with a following VP, and so as being subcategorized [NP \_\_\_ VP] or something similar. There is evidence against this analysis for the Kayah Modals, in the form of certain morphemes that, although not adjacent to the Modal, seem to have scope over it. Consider

- (84) *do cwá bæche á*  
 abstain go bored NS  
 be tired of abstaining from going (10/13)

The meaning suggests that *bæche* relates to *do cwá* as a unit; it has a kind of scope over it, which can be taken as evidence that the structure is [*do cwá*] *bæché*]. If the Modal *do* were in construction with the entire remainder of the sentence (or at least with the remainder of the VP), the structure would be [*do*[*cwá bæche*]], which would give *do* scope over *bæche*, thus meaning 'abstain from being tired of going'. A similar example is:

- (85) *de síplɔ cwá cè to*  
 decide go able NEG (10/20)  
 be unable to decide to go

Here the structure must be [[[*de síplɔ cwá cè*] *to*]; if *de síplɔ* 'decide' were in construction with the following items as a unit, as [*de síplɔ* [*cwá cè to*]] one would expect the meaning to be 'decide to be unable to go'<sup>7</sup>. Again we see that the Kayah VC is a

close-knit unit, in which verbs interact first with each other and then, as a unit, with the rest of the clause.

#### 4.2.6. Summary of V-V's.

Here we may summarize the types of V-V, in terms of the various derived argument structures that have been described.

	<u>argument-mapping</u>		<u>valence</u> (*of arg's)
	<u>arg of V<sub>1</sub></u>	<u>mapped to arg of V<sub>2</sub></u>	
Resultative	Pat	Pat	$v=V_1$
Sequential	Agt	Agt	$v=v\text{-highest}$
Directive	Causee	Agt	$v=V_2 + 1$
Descriptive	not applicable		$v=V_1$
Modal	not applicable		$v=V_2$

The *argument-mapping* column shows which argument of the first verb is mapped to which argument of the second verb. In the *valence* column, the valence  $v$  (here, the number of arguments only) is expressed in terms of the valence of the component verbs: thus the valence of a Resultative is the same as the valence of the first verb, the valence of a Directive is equal to one more than the valence of the second verb, and so on.

The argument mapping could also be shown in terms of grammatical relations, as follows, using Abs (absolute) as an abbreviation for 'Subject of intransitive or Object of transitive':

	<u>arg of V</u>	<i>mapped to</i>	<u>arg of V<sub>2</sub></u>
Resultative	Abs		Sbj
Sequential	Sbj		Sbj
Directive	Obj		Sbj
Descriptive		not applicable	
Modal		not applicable	

### 4.3. Aux/V Constructions.

#### 4.3.1. Introduction: types of verb particle

We distinguish verbs, which may function as unitary predicator, from Verb Particles, which appear in the VC but cannot function as unitary predicator (there are also clause particles and sentence, or illocutionary force particles [see 4.6 and 7 below]; in this section I will often use the term 'particle' to refer to Verb Particles only). The two categories are distinct enough to be useful, but they blur at the edges in certain ways. One prominent aspect of this are the classes of particles which combine with verbs to form constructions that parallel types of V-V construction. Compare:

#### (86) Modal V + V

ʔa be cwá            he must go

Ptc + V

ʔa lò cwá            he ought to go; it's his duty to go

#### (87) V-V (Descriptive)

ʔa ʔíbe jè            (jè: verb, 'difficult')

he speaks with difficulty; it's hard for him to speak

V + Ptc

ʔa ʔíbe cè            (cè: particle, 'know how, good at')

he speaks skillfully; he knows how to speak

Note that the resemblance in each case extends to both semantic value and syntactic



positioning: both the Modal verb *be* and the particle *lo* precede a co-occurring verb, and both the verb *je* and the particle *ce* follow verbs. In such cases we will take advantage of the resemblance by naming the class of particles after the similar verbal construction type: thus *lo* is a Modal Particle (4.3.3), and *ce* is a (type of) Descriptive Particle (4.3.7). This may also be taken to support the hypothesis that some or all of these particles descend from one-time full verbs. *ce*, for instance, may be cognate to Sgaw *be*, a verb meaning 'able'.

There are also classes of particles that have no obviously associated type of full-verb construction; unfortunately they also lack any easily-described common semantic value that could serve as a name for the class. In such cases I fall back on naming the classes after an arbitrarily chosen class member: the *khwe*-class particles, the *si*-class particles (4.3.2, 4.3.6)

Finally, there are some non-verbs that seem to function like second verbs in Resultative V-V's, of both General and Directional types. These are the most verb-like particles, and in recognition of this I have called them 'bound result expressions' (4.3.8).

Verb Particles may be divided into pre-verbal particles, which precede verbs and V-V constructions; and post-verbal particles, which follow.

### 4.3.2. Pre-Verbal Particles: khwe- class particles.

These express aspect and aspect-like notions.

tarē	almost V'd
lèkīō	have ever V'd, experiential
khwe	in the midst of V-ing, continuous
tε	about to V, incipient
tānē ~ kənē	about to V, incipient
tú	just now V'd

Both *tú* and *khwe* usually co-occur with the *ɾí*-class Post-Verbal Particle *pá* 'durative, etc.' (see 4.3.6). *lèkīō* may possibly consist of *lè* 'purpose nominalizer, thing for V-ing' and *kīō* 'speech, language': place for speaking, occasion for language; i.e. one has done something and hence knows enough to speak about it.

There is a symmetry between the *khwe*-class particles, which appear first in the VC, and the *ɾí*-class particles, which come last in the VC. Neither relates to any V-V construction type, and both have meanings relating to aspect (although the *ɾí*-class, much more numerous, includes many with non-aspectual meanings).

### 4.3.3. Modal and quasi-modal particles.

Like the Modal verbs, these include some meanings that are truly modal in referring to obligation or probability, and some meanings having to do with attitude of the Agent.

#### Modal Particles

lò                      ought (by duty)

klé                     should be the case (epistemic)

(possibly related to a word meaning 'to require as ingredients')

thú                     probably is the case [analysis uncertain]

#### quasi-modal Particles

sí                        want to V

lā                        V intrusively, rashly

*lā hé* to interrupt (hé 'say'); *lā kəʔɔ́ hé* to guess; *phrè nɛ lā li mē nɛ* the Shans were smart [to our detriment] (in a story concerning the Shans' expropriating Kayah land).

tarú                    go ahead and V, feel free to V

cà phrā                keep on V-ing (in spite of stg)

Possibly also to be listed here is *ré* 'should be the case, you'd have thought that' if not related to the V *ré* 'good'.

### 4.3.4. *já*, a 'Sequential' particle.

*já* is like a worn-down, grammaticalized first verb in a Sequential V-V. It has lost the literal sense of motion, perhaps similar to *go* in colloquial English *Now you've gone*

*and done it.* E.g.

- (88)  $v\bar{e}$  ke ja s $\bar{a}$  b $\bar{o}$ λ | ka i $\bar{o}$  b $\acute{s}$  ? $\bar{a}$   
 1s if die and-then come bury at here  
 If I ('go and') die, come bury me here. (217.5)

It is most likely cognate to the full verb *cwá* 'go', but is distinct enough

grammatically that the two may co-occur:

- (89) ?ajjá cwá t $\acute{e}$ pa | w $\bar{a}$   
 3 go exclamatory DCL  
 He (went ahead and) went (even though I told him not to) (10/29)

#### 4.3.5. Post-verbal particles: *rá*-class.

The V Ptc's occupy the final positions in the VC, and have quite a variety of meanings.

Some of these have to do with aspect or aspect-like notions, some relate to the presence of certain participant roles in the clause (e.g. an NP with a Benefactive role), but many have meanings that cannot be easily gathered under a single rubric. This class resembles the *khwe*-class Particles in its inclusion of aspectual meanings; the two classes could in fact be called VC-initial Particles and VC-final Particles.

In the list that follows, the *rá*-class Particles are listed in the general order in which they occur. Some of them are grouped together in an order-class, which means either that they are mutually exclusive in that position or that they may occur in any order with respect to each other.

khjā            back again; commonly occurs together with sé (listed below). E.g. dá khjā

sé 'give stg back'. Related morphemes: khjā 'back' (location noun), kəkhjā 'backwards'

(Directional Bound Result Expression).

kóΔ    temporarily, V instead for a bit. This is obviously related to the Descriptive

Particle kó 'temporarily'; but both may occur in a single VC: ʔo kó rə kóΔ lw 'rest a

bit first' (for rə see Descriptives; for lw see below).

pò        additionally; meaning similar to Thai ìik. Tends not to occur with V<sub>i</sub>'s denoting

states ('adjectival V's'). Compare ʔə below.

thō       finish, perfective aspect; nearly always followed by á 'new situation' (listed

below).

pé        not after all, new negative situation; must co-occur with the negative Clause Ptc

to. E.g. ʔa | pù cè ʔe | to 'it turned out that he didn't know how to catch (fish) after

all' (pù 'catch', cè 'know how to V').

sé back in reaction, completing one half of a cycle. Examples: ʔu | sípɔ du sɛ | pe  
 'they'll get mad back at us' (if they hear you talking that way about them); hē pò sɛ |  
 kúkiē, ||ka tò sɛ 'went to work the swidden (again), and returned again' (pò listed  
 above, tò 'arrive').

nó emphatic or unexpected negative; must co-occur with negative to. E.g. n̄  
 ɲ | hóhé | to ē 'aren't you going to school?'; v̄ē | cwá ɲ | to 'I'm NOT going'.  
 Somewhat similar to Thai māj-dāj.

kā comitative participant involved, interested person involved (COM in  
 glosses). Has the effect of adding an argument, as in cwá k̄ | v̄ē 'go with me'. cwá,  
 being a V<sub>1</sub>, could not take a following argument like v̄ē without the presence of k̄ (or  
 some other valence-increasing morpheme). Evidence that the extra argument does not  
 relate directly to k̄ as in a PP: hē ʔíchi k̄ lāí phē khru to 'haven't yet gone out  
 with Father to cut firewood' (5/7). Very often, however, k̄ seems to add virtually  
 nothing to the meaning of the sentence, either because (i) the additional argument has  
 already been made possible by some other valence-increaser such as the Descriptive  
 Particles cwá 'help' and bēbū 'show the way to', or (ii) k̄ only indicates a vague  
 relevance of the event to either some NP participant in the discourse or to some  
 human participant in the speech event. As an example of the latter sort, cf. the

common utterance (in my presence, at least!) sí ngē k̄ā | to '(you) don't understand ("with" us)'. This and the following morpheme are extremely common, usually omissible, and particularly difficult to analyze.

rá participant obliquely involved. Indicates that a following NP argument has a role such as instrument, object of emotional-state verb, or some other vague relation. Similar to k̄ā but seems not to be restricted to allusion to human participants; common, optional, further analysis needed. Examples:

béswá rá p̄anè	be companions with a buffalo (19.3)
síplɔ du rá ʔamòʔaphē	be angry with his parents
ʔā ma nō ʔe rá ʔitē	what's this used for?
ʔo rá lè donē	to be a matter for legend (51.6)
ʔase de rá ʔaklwī ko	save the fruit for seeds (91)

dítw̄ keep on V-ing, of an undesirable event/action:

kè	le sé	pī	<u>dítw̄</u>	mi
AMB	wind-blows	go-out		fire
The wind keeps blowing out the fire. (1/24)				
bó	kā	<u>dítw̄</u>	bja thé	i hú ʔā n̄  ma hé  hú tē
weave	COM	PTC	cloth like this	N̄ be-so say like what
[if she] went on weaving like this what would he say? (286.2)				
(of a woman whose husband had told her not to work too hard)				

mā imperative:

thō mā tɛ kle    come over, Kle! (14.4)

já mé mā khē | Klémè lè mā | nɔ ho    go look at Klémeh's bed. (268.3)

phé only, just:

vē ?o phé tɛ ?ā ηē ba to: I have only (as much as) five baht.

vē cwá nó to, vē ?o phé ?a bā    I didn't go here, I live here! (169.10)

May be repeated at the end of the clause:

?e phé taklé phé    ate only half. The second *phé* can be considered to belong to the class of clause particles (4.6).

tamō            constantly:

Doɔ chwá tamō    Do'a keeps getting fevers (133.1)

chílw̄ ~ sílw̄    too, excessively, very:

ké ku tɔlwá chí\_lw phé    It's just too hot! (129.5)

lēkhē           plural action, often used in imperatives:

?a |ka kəthe tò lēkhē | bɛ ?ā |nɔ hō    They all came back up here, and... (247.3)

chwá mé lēkhē tɛ | me ní    Don't you (pl.) go look at it, now! (326.2)



khрэ plural questioned entity, 'what-all':

ʔā tenā |me khрэ|ʔitē What [various things] did you do today? Thai equivalent: wan  
nīi tham araj bâaŋ? (c)

bja V because of a feeling of obligation:

cū kā bja|lū te accord with his (things=) words → heed him [as you ought] (c) Note:  
this is probably related to the first syllable of *bja phé* 'for free, gratis', itself of  
uncertain analysis, but probably a Descriptive Particle.

lāte on the other hand

lāteá instead

síŋē ʔé á|| pe | síŋē kā lāte|kulā nò : taphré : to You understand a lot, on the other  
hand not one of us understands English. (220.6)

ʔa dā pa li thē to, ʔa dā lāteá li phá he didn't give a gold book, he gave a hide book  
instead. (100.4)

já [meaning very uncertain]

vē ʔo mā mè | já méthλ ʔé cè já pa | rū

I s sleep dream go-and see much really DUR money

I dreamed [that I] went and saw a great deal of money. (121.2)

ləkε                    V instead of somebody less able

cwá ləkε | dǝ ne    You go instead. (247.7)

ɗw on one's own, of one's own accord

ʔa | hó ló ɗw | ʔa te    She is pregnant on her own (i.e. immaculate conception) (78.3)

chā | ʔe ɗw    The chicken ate it of its own accord (i.e. it's not my fault). (104.30)

súplɔ cō ɗw    a self-tying [magical] rope (360.1)

...ma ɗw | pe phú    ...is still our own child. (52.6)

Note: reflexive meaning is usually indicated by the NP X nè 'X's body'; however *ɗw*

may occur as well: mé rá ɗw | ʔa nè    looks at himself in it [a mirror]. (c)

á                            new situation (NS in glosses)

Note that although á very often occurs with the Particle thō 'finish, perfective'

(above), the two are distinct in meaning. Compare:

ké cw á    It's raining (said on noticing the rain, whether or not it has been raining  
prior to the time of noticing).

ké cw thō á    The rain has stopped ('it's finished raining').

lāi (not) yet: must co-occur with the clause Particle tɔ 'negative' vĕ | ʔe lāi | dī | to |  
haven't eaten yet.

lw (meaning uncertain) ? just, a bit. Often occurs following kól (see above) and/or preceding the ClfP taki 'a little'. nì dā kól lw taki (listen=)wait a minute. (used in conversation)

pa durative, still (DUR in glosses):

mé thλ nó k̄ | 'a | to | ?a | ka pa | d̄ khjā [I] haven't seen him, he's coming along behind. (197.2)

ké | cw pa | to Rain is no longer falling. (c)

?a | ?o pa (1) there's more; (2) it's still there

du l̄x pa pa still bigger=even bigger (c). For the reduplication, see Chapter 2.

thwā NP pa (still is an NP=)NP is alive. Note: thwā is an 'incorporating' verb (see 4.5). The construction V pa ... to, literally 'not V any more', can have the force of 'don't have to V'. More investigation is needed for this morpheme; the translation 'durative' is only a tag, and is not meant to exclude related notions like progressive, continuous, and so on.

nē high time to V, must V because hasn't V'd in so long:

pá rò ma ké cw nē dw á Tomorrow it's just got to rain (9/29)

?a ká khjā lāi to ŋjā lō á, pá rò ?a klé dá khjā sé nē taphé He hasn't paid [me] back for ages, tomorrow he's got to (9/29)

tapa exclamatory; usually co-occurs with declarative Clause Particle wā: ʔa

kleə sɪŋē pwā cɛ tapa wā

should know every sort DCL

You *should* understand everything! (221.6)

tè shouldn't; often occurs with Sentence Particle *me* 'don't' (section 8.2):

chá tè lū me Don't fight! (9/22)

cè maybe not, tentative negative:

cwá pé pa cè pā

go after-all DUR IRR

May not go after all (c)

Note that this negative, like all other *ŋ̄*-class Particles, has a fixed position

following the last verb in the VC, and so cannot be used to pick out the main verb (or

any particular member of a V-V), resembling in this respect the Clause Particle *to*.

lū each other, together

Distinct from the homophonous obviative pronoun (but probably related); cf. *pe me*

cwà lū We help each other (not 'we help him')

ne kũ when...then

ʔa | thō mé sú né.kũ | rʌ || hé tò já | pe mē ʔ  
 he go-over look wrong PTC say arrive PTC we excl.

When he went over and couldn't find it, then he went and (talked=) complained to us!  
 (251.4)

pè, pjà benefactive/malesfactive, dative of interest, to sbdy's benefit or

detriment (cf. 3.3); adds a Recipient argument

ché pè phúcè ca to sew clothes for children (c)

me bī pè ʔa kadā Shut the door to keep him out (9/26)

ʔa sō lō pè lū It all rotted on them, to their detriment (155.6).

The difference between the two forms pè and pjà seems to be that pjà is confined to contexts in which the Recipient is third-person. Note the possible contrast: ʔa me pè mē He did it for you, OS (mē as a term of address); ʔa me pjà mē He did it for OS (mē

as a term of reference). pjà may be a fusion of pè and the third person pronoun ʔa.

Order classes including more than one member:

khjā	sé	phé	á	tè	pè
kóʌ	nó	tamō	lāí	cè	pjà
		chílū	lʌ	təpa	
		lēkhē	pa		
		khré			
		bjá			
		láte			
		rʌ			

#### 4.3.6. Descriptive Particles.

This is a fairly large and heterogeneous class; its unifying characteristics are the broad syntactic features of a) following co-occurring verbs; and b) preceding the *ni-* class particles. The class is not an analytically satisfying one, and it seems likely that further investigation may show it to be an amalgam of several distinguishable classes, or that some of it may turn out to be part of a class already described. As it stands, the class of Descriptives is more heterogeneous than the other Verb particles, with meanings ranging from quite concrete to highly abstract. Also, the class contains a great many members, although not so many as to make an unlistable number (i.e. it is probably a closed class); and the members show hard-to-discern properties of ordering, both with respect to each other and with respect to other classes.

It can be divided into two subclasses.

##### 4.3.6.1. Movable.

These may permute with each other, with other Descriptive particles, or with  $V_2$ 's in Descriptive V-V's, with corresponding shifts in meaning. The first four in the following list may also be grouped together under the rubric 'Potentials':

<i>cè</i>	able, know how to V, good at V-ing
<i>pè</i>	physically able, strong enough to V; possibly related to full V <i>pè</i> 'win'
<i>bé</i>	have the wherewithal to V (money, raw materials)
<i>bu</i>	dare to V

-----

có	insist on V-ing, stubbornly V
rò	many (people) V, plural action by humans
lǎ	more, -er (comparative degree), very, than; 'CMP' in glosses
lū	mutually, each other
khru	equally, as (much) as
hā	often (possibly a verb, more data needed)
rà	beforehand (possibly a verb, see [4.2.4])
no	afterward "

Examples of permutation:

cwá rǎ jè going beforehand is hard

cwá jè rǎ going was hard before

ʒírē phrē lǎ jè hard to work faster

ʒírē phrē jè lǎ harder to work fast

cwá jè hā often hard to go

cwá hā jè hard to go often

These are subtle distinctions, at least when deprived of context. In the last pair, *cwá* *jè* *hā* might be said in reference to a road that is frequently washed out, while *cwá* *hā* *jè* could contrast with 'but to go once or twice would be easy'. Cf. also

ʔíbe cè phrē learn rapidly to speak, e.g. a precocious child

ʔíbe phrē cè good at speaking fast.

kədā kū | me bí jè lǝ all windows are hard to shut (5/23)

kədā kū | me bí lǝ jè hard to shut all the windows "

#### 4.3.6.2 General Descriptive Particles.

Again, there is little in the way of a semantic common denominator to this class; like the second verb in a Descriptive V-V, the sense is vaguely 'adverbial'.

ho secretly, sneakily.

kó temporarily, for a while; e.g. ʔo kó 'to rest' (ʔo 'be at, dwell), phjá kó 'borrow' (phjá 'take').

cwà help to V; adds an argument. E.g. me 'do, make' is a V(t), but in combination with cwà in can appear with an additional (post-verbal) argument, as in ʔa | me cwà | Mía | Thīm hi 'he helps Mía build Tīm's house' (here the name Mía is the added argument).

bébw show the way to V, take sbdy V-ing; e.g. cwá bébw 'lead sbdy'. This also adds an argument, in a fashion similar to cwà.

phe supplanting, appropriating; e.g. Phrè | ka ʔo phe 'the Shans came and lived (on our land, supplanting us) (200.1).



- re unrestrainedly and to an undesirable extent; e.g. cwá re 'go all over, go just anywhere', ?íbe re 'talk loosely, talk wildly'; perhaps related to the following.
- tare cū regularly, all the time.
- təple over again, a turn; e.g. sē təple | hɬ ca 'change one's clothes' (contrast 'put on clothes again', which would use the V Ptc's khjā sé, sē khjā sé hɬ ca); probably originally a Numeral-Classifier construction (tə- 'one').
- tè wrongly, V the wrong one; cf. ?íbe ta tè 'make a slip of the tongue', also the Elaborate Expression ta tī ta té 'unclear, halting' (of speech).
- ?one against, defensively; usually co-occurs with V ptc pè benefactive/malefactive, dative of interest; e.g. me bī ?one pè | kajē | kadā 'close the door (kadā) against people, close the door so no one gets in'.
- lé keep up with, to V overtaking somebody
- lè hurry and V

The Descriptive particles, including both the General and Movable subgroups, can intervene in Descriptive V-V's: another property linking them to verbs. E.g.:

*cwá ho jè*      hard to go secretly

*me cwà l̄s*      all help to do

*k̄s tho tare cū r̄ò*    get up early regularly

In particular cases it may be that there is only one possible order; in others there are several possibilities. It remains to be discovered what principles are at work in such cases.

#### 4.3.7. Bound Result Expressions.

We have defined the class of verbs as consisting of any morpheme that can function on its own as predicator; a distributional definition would be any morpheme that can occur in the slot \_\_\_\_\_ *l̄ā í to* 'hasn't V'd yet'. A secondary feature of verbs is that they follow a modified noun in an NP; however this is not a criterial feature since there are certain types of noun-noun construction in which the modifying noun follows the modified.

There are certain morphemes that do not satisfy the primary criterion for verbhood (occurrence as unitary predicator), but that occur following verbs in constructions that are semantically and syntactically indistinguishable from Resultative V-V's.

This is demonstrated in the following examples, in which *chwí* 'cool' and *le* 'warm'

each appear in four contexts: as unitary predicator, following the colorless verb *me* 'do, make', following a verb with more semantic color (*dε* 'put' and *thɔ́* 'be covered, wrap oneself up'), and modifying a preceding noun:

(90a) *dī chwí á*

rice cool NS

the rice is cold

(b) *ʔa me chwí dī*

3 do cool rice

He cools the rice.

(c) *ʔa dε chwí dī*

3 put cool rice

He set out the rice to cool.

(d) [<sub>NP</sub> *dī chwí* ]

cold [cooked] rice

(91a) *\*[<sub>S</sub> dī le ]*

rice warm

the rice is warm

cf. (91a') *dī ku le*)

rice hot warm

(b) *ʔa me le dī*

3 do warm rice

He warms up the rice.

(c) *ʔa thó le ʔikē*

3 cover warm blanket

He wrapped himself up warm in a blanket.

(d) [<sub>NP</sub> *dī le* ]

warmed-up rice

The (a) sentences demonstrate that *chwí* is a verb while *le* is not. The sentence (91a') gives the grammatical equivalent of the ungrammatical (91a): as a predicator, *le* must be in construction with some preceding verb. The (b) and (c) sentences show the close similarity of the two morphemes. The similarity is semantic, in that both relate to the preceding verb (and hence to the rest of the sentence) in the manner that we have discussed as the derived argument structure for Resultative V-V's: the participant bearing the Patient role of  $V_1$  (in this case *ǎ* 'rice') also is naturally interpreted as being associated with the Patient role of the second morpheme: the rice becomes warm (for the concept of roles specified by a non-verb cf. 3.2.4). The similarity is also syntactic, in that nothing can intervene between the two morphemes in either case, and the two-morpheme constructions act as a unit in relation to other VC constituents; e.g. they may act as  $V_2$  in a Directive construction:

(92) *nō me chwí*    tell [somebody] to cool off [something]

(93) *nō me le*      tell [somebody] to warm up [something]

Also, both morphemes can modify a preceding noun; it is particularly striking that the non-verb *le* can do so on its own, as in (91d).

To reinforce the point, let us examine a parallel set of examples with another pair of morphemes, one verb and one non-verb (unfortunately I do not have data on the possibility of this non-verb modifying a noun):

(94a) ʔa ka á

3 return NS

He's going home.

(b) phjá ka

take return

get stg (and bring it home)

(c) ʔīphri ka

buy return

buy (and bring home) stg

(95a) \*ʔa kē

3 New-Location

He's staying temporarily.

(b) ʔa ʔo kē

3 be-at NL

He's staying temporarily.

(c) kɪw kē

shave NL

shave off (e.g. a beard)

Note: kē is best glossed as 'so as to be in a new location' (abbreviated 'N.L.').

Since morphemes like /e and kē act like V<sub>2</sub> in Resultative V-V's in all but verbhood, let us refer to them as Bound Result Expressions (BRE). Furthermore, like the Resultative V-V's, these particles can be divided into General Bound Result Expressions like /e and Directional BRE's like kē (see also the following section).

Following the logic of this chapter, the BRE's might better be called Resultative

Particles. This would further increase the parallelism between V-V construction types and particle classes: we already have Descriptive Particles, Modal Particles, and one candidate for the title Sequence Particle (4.3.2.3). There are, however, two reasons to distinguish the BRE's from (other) Verb Particles.

First the Bound Result Expressions are closely related to certain other non-verbs that also follow verbs, often in combinations that are easily construable as having Resultative-like meaning. These morphemes, however, are restricted to occurrence after only one particular verb: they are of the type that has been called 'cranberry' morpheme, or 'morphan' (Matisoff). In other words, these morphemes are bound and restricted, while the BRE's described above are bound and versatile. *ꞵ* 'be at, exist' is especially important as first verb with these: *ꞵ ꞵḡ* 'lie down, sleep', *ꞵ ꞵḡ* 'sit', *ꞵ ꞵḡ* 'do for fun, visit' (like Thai *thiaw*). In each case, the second morpheme is not known to occur after any other verb. These constructions are perhaps not unambiguously resultative in meaning; it is at least possible, though, to construe them as 'be located with the result that one sits/lies/enjoys'. Another example with more clearly resultative meaning is *thiaw* 'see', only after *mḡ* 'look' (*mḡ thiaw* look with the result that one sees). These constructions are best considered to be compounds, although they are semantically rather transparent for compounds. The second elements can be referred to as Restricted BRE's, or Resultative Morphans. If these are considered to be related to BRE's, the latter then differ from the Verb Particles, which are all highly versatile.

A second factor dividing the BRE's from the Verb Particles is the distinct possibility that at least some of the BRE's are verbs after all, their seeming non-verbhood being a result of idiosyncracies that happened to cause them to fail particular tests of their possible use as predicators. For instance, with *mō* 'open', *\*pa mō* 'it's open' is not acceptable, seemingly indicating non-verbhood, but *kadā mō í* 'the door is open, the door has been opened' is acceptable. It is not known whether the difference is caused by the *í*-class particle *í* 'new situation', the full lexical Subject NP *kadā* 'door', or something else.

Below are some other Bound Result Expressions that have been found:

- |            |                                                                                                                            |
|------------|----------------------------------------------------------------------------------------------------------------------------|
| <i>plú</i> | to the end of an expanse, all the way through                                                                              |
| <i>dwá</i> | away, available for future reference (like Thai <i>wáʔ</i> ); possibly also occurs in the compound <i>dwá sé</i> 'thrifty' |
| <i>bja</i> | having a damaging effect on; very similar to the verb <i>be</i> (see 4.2.1)                                                |

Bound Directionals. These are quite similar to the verbs that appear as  $V_2$  in Directional V-V's; some, in fact, are derived from Directional verbs by prefixation. As with the general BRE's just discussed, these Bound Directionals could well be called Directional Particles; in fact with even more justification, since they form a closed class of versatile morphemes, without any restricted related class analogous to the Resultative Morphans.

Three subtypes can be distinguished.

(A) Orientational directionals.

The Type B Directional verbs (4.2.2 above) with added prefix *pə*, plus several forms with other prefixes, make up a class (these have not actually been tested for verb-hood; they are placed in this section by hypothesis). This class describes the orientation of some participant in the action, usually that denoted by the Object NP, if present.

Semantically these differ from the Directional verbs in two ways. First, the oriented participant does not necessarily move; e.g. *mé pəhē* 'look ahead' (*mé* 'look'), *jò pathe khe* 'raise one's leg, have the leg raised' (*jò* 'extend, hold extended'). Second, if the situation does involve motion, the *pə*-prefixed Directionals give the orientation of the moving entity, while the Directional verbs refer to the path it describes through space. In Gestalt terms we might say that *pə*-prefixed Directionals describe orientation of the Figure while the Directional verbs describe motion through the Ground.

*pə*-prefixed Directionals:

*pəhē* ahead, forward

*kəkhjā* backwards

*pathe* upwards

*tālē* downwards

*pathe* outwards

*pəñō* inwards

*pārē* across



Like the Directional verbs, the *pə*-prefixed Directionals may refer to non-literal direction:

(96) phú cè | li mē pəhē á

child smart ahead NS

Children are getting smarter (these days). (15.6)

(B) *talwá*-class directionals.

təva around, encircling

və around a obstacle, circumventing

talwá past

phā ~ taphā out of the way

bé ~ kabé across, crossing, omitting

rwá along

Cf. also:

taka around, in a curving path (full verb)

These have several special characteristics. a) they consistently follow Directional verbs; in fact they favor an intervening Directional verb if the first verb is not a Directional; b) they add an argument to the clause, the added argument denoting the entity serving as reference point for the spatial configuration (i.e. having the same semantic function as the object of a preposition in English); c) the added argument, which can be identified as Locative, appears as Obj-x rather than in a PP. The full verb *taka* is also listed here because it shares these two characteristics, differing from the other

morphemes listed only by being a full verb--yet another example of the closeness of these Bound Result Expressions to verbs.

To illustrate the argument-adding property, consider the verb *cwá* 'go', which is a  $V_f$ , taking no Object NP. The combination *cwá talwá* 'go past', in contrast, is effectively a  $V_f$ , as in *ʔa cwá talwá vē hi* 'he went past my house'. Note that in this sentence *vē hi* 'my house' functions as Object of the construction *cwá talwá*; the constituency is [[*cwá talwá*] *vē hi*], not [*cwá* [*talwá* [*vē hi*]]]. The *talwá*-class directionals are constituents of the VC; they do not form a unit with the following NP. There are many additional VC constituents that could intervene between the Bound Directional and the NP, including a verb in a Descriptive construction, and/or any of the *ra*-class Particles; e.g. *ʔa | cwá talwá lāi | vē hi | to* 'he hasn't gone past my house yet'.

#### Abstract Bound Directionals.

Finally, we will list four morphemes that are fairly common following verbs in Resultative-like constructions. These are all non-verbs, and their meanings are directional only in a fairly abstract sense; still they co-occur with the (other) Directionals, often with both possible relative orders.

*kē* new location, to V so that something ends up in some place. E.g. *dē kē* 'put down, put somewhere'; *wī kē* 'throw away'; *kʷ kē* 'shave off' (the beard ends up in a new location); *ʔo kē* (1)'stay, dwell temporarily', (2)'separate, split up (as a couple)'. The

same morpheme probably appears in the compound *kē dē* 'go to waste, be wasted' (*dē* is a Descriptive Particle 'to V in vain').

*pē* transfer of possession. E.g. *ʔīche pē* 'sell to'; *dá pē* 'give to'; *khē ʔo pē* 'rent to'.

In the last example *pē* follows a Resultative construction consisting of *khē* 'rent' and *ʔo* 'be at, dwell', literally 'rent so that (sbdy) lives at'. This Bound Result Expression is distinct from, but related to, the *nī*-class verb particle *pē* 'benefactive'. Abbreviated 'TRN' in glosses.

*ʔó* be hidden, into hiding. E.g. *kīé ʔó* 'escape' (*kīé* 'run'); *dē ʔó* 'hide stg' (*dē* put').

*chw* confronting, facing, coming from opposite directions; often appears together with the *nī*-class Particle *lū* 'mutually, each other'. E.g. *chwá chwí lū* 'meet, approaching from opposite directions' (*chwá* 'go'); *ʔíbe chw lū* 'converse' (*ʔíbe* 'speak').

#### 4.4. Interactions.

Thus far we have described the VC in terms of binary combinations of morphemes; these may be called *simple V-V's*. This section will consider some of the ways in which constructions of more than two members are formed in the VC. The division is between constructions involving verbs only and constructions involving both verbs and particles.

##### 4.4.1. V-V and V-V

Constructions of more than two verbs are formed in two ways. First, one type of V-V construction may have more than two members; this may be called a *compound V-V*. We

have already seen examples of this for Resultatives and Sequentials; cf., respectively,

(22) and (42), repeated below:

- (22) ne | cwī the thū | ʔacɿ | nɿ  
 2 pull go-out long wick Nɿ  
 (if) you pull the wick out long... (341.6)

Here both *the* and *thū* describe results of the action.

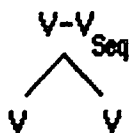
- (42) hē ʔo māl klē | tə-nāl nāl  
 go sleep cut one two day(cif)  
 (we) go and sleep (in the fields) and cut (brush)  
 for one or two days (177.2)

Here *hē*, *ʔo māl*, and *klē* describe sequential actions<sup>B</sup>.

At present I have nothing to add to what was said about such constructions in the relevant sections above.

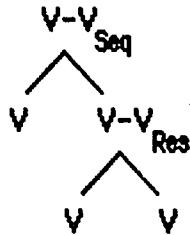
Secondly, different types of V-V construction may combine; that is, in many cases one or both 'V's' may themselves be V-V's. This can be called a *complex* V-V.

To facilitate discussion of complex V-V's, let us adopt a type of constituent-structure notation in which nodes or brackets are labeled with the name of the V-V construction type; thus a Sequential V-V could be [<sub>Seq</sub> V V] or



Similarly, a Sequential V-V whose second member is a Resultative V-V would be

represented as [<sub>Seq</sub> V [<sub>Res</sub> V V ]], or



The abbreviated node labels are Mod = Modal, Dsc = Descriptive, Drv = Directive, Seq = Sequential, and Res = Resultative.

Examples of complex V-V's which have already been cited include those combining Sequential with Resultative and Sequential with Directive; two are repeated below:

(46) [<sub>Seq</sub> V [<sub>Res</sub> V V ]] vĕ | pù me sĀ | jò khró

1s catch do die rat

I caught and killed a rat. (10/8)

(64)=3.2.3(6) [<sub>Drv</sub> V [<sub>Seq</sub> V V] nā kəɛ dé thā

command go-down dip-up water

tell sbdy to do down and draw water

Of course not every type of V-V can expand both of its V constituents as any other type of V-V. For instance, Modal and Directive V-V's are defined by having first verbs that belong to certain closed classes; since many (indeed, most) individual verbs are barred from acting as that first constituent, *a fortiori* V-V's are also barred. This still

leaves a great many possible combinations, not all of which actually occur. For instance,

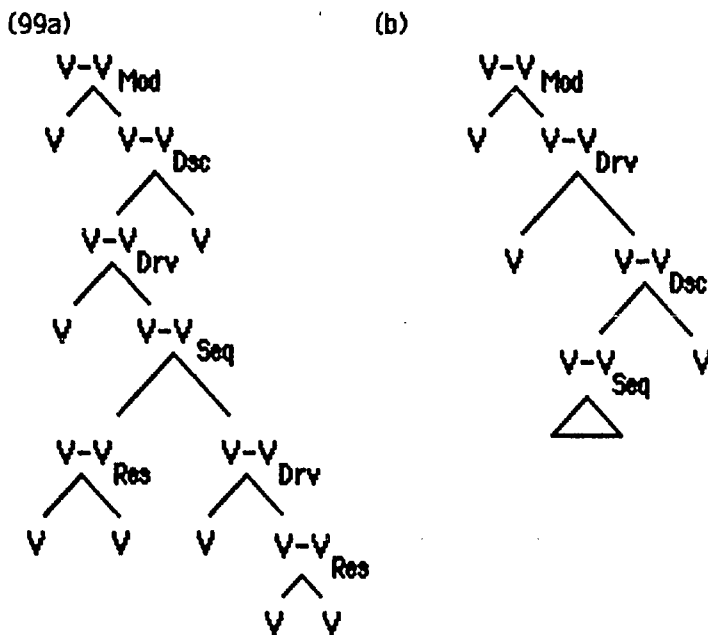
[<sub>Drv</sub> V [<sub>Seq</sub> V V]] is common, as in (64) above, and:

(97) *nō cwá ʔilò* tell sbdy to go plant (from 20.8)

but [<sub>Res</sub> V [<sub>Mod</sub> V V]] seems to be impossible:

(98) \**mūi sí nò* beat sbdy so that they want to weep

The sum of possible combinations in complex V-V's can be displayed as follows, using a tree diagram; the maximal pattern has two possible forms:



(the structure below the Seq node, abbreviated in [b], is the same as in [a]). Note that (99a) and (b) differ only in the relations between the upper Drv node and the Dsc node: in (99a), Dsc dominates Drv, while in (99b) Drv dominates Dsc. This can also be phrased in

terms of the notion c-command:  $\alpha$  c-commands  $\beta$  if the first branching node dominating  $\alpha$  also dominates  $\beta$  (and  $\alpha$  does not itself dominate  $\beta$ ). In (99a), V(Drv) c-commands V(Dsc), while in (99b), V(Dsc) c-commands V(Drv). These two patterns can be generated by a fairly simple set of constraints referring to c-command relations:

- (a)  $V_1$  in a Modal V-V c-commands all following V(-V)'s.
- (b)  $V_2$  in a Descriptive V-V c-commands all preceding V-V's, except Modals.
- (c)  $V_1$  in a Directive V-V c-commands all following V-V's.

The latter two conditions overlap to give both possible c-command relations between  $V_1$  of Directive and  $V_2$  of Descriptive. A further condition needs to be stated: a complex V-V can include at most one Directive verb, as discussed in (4.2.3) above.

Examples of various complex V-V's follow; they illustrate the 'semantic scope' relations that are the evidence for the constituent structure I have posited. This semantic scope has already been cited in 4.2.5 above; for an additional example, (108) below is *nō ʔíre phrē* 'tell somebody to work fast'. In other words 'fast' is part of the content of the command, which is taken to indicate that the structure is [*nō [ʔíre phrē]*]. If the structure were [[*nō ʔíre*] *phrē*] we would expect it to mean rather 'quickly tell somebody to work'.

Examples of complex V-V'sSequential with Resultative

- (100) pù [me sã] catch and kill stg
- (101) the [phjá tã] go up and get down
- (102) [cõ ma] mû tie up and beat stg
- (103) [bó mō] mé open ('reach'+ 'be-open') stg and look

Sequential with Directive

- (104) hē [nō dé] go and tell to draw [water] (Seq+Drv)
- (105) nō [kale dé] tell to go and draw [water] (Drv+Seq)

Directive with Resultative

- (106) nō [kúvē the] tell to dig out

Directive with Descriptive

- (107) [nō ʔíre] je have difficulty telling sbdy to work=sbdy is hard  
to get to work (10/11)
- (108) nō [ʔíre phrē] tell sbdy to work fast
- (109) [nō ʔíre] ré tell sbdy nicely to work
- (110) nō [ʔíre ré] tell sbdy to work well

Descriptive with Resultative

- (111) [me pjá] ho ruin stg secretly (c)
- (112) [takō mu] l̩ chop stg all up fine



Descriptive with Sequential

- (113) [kèkjá kəthɛ] nì                      manage to fold up stg and go up, fold up stg and go  
up taking it (317.3)

Modal with Descriptive

- (114) tō [cwá jè]                              should go with difficulty
- (115) sí [cwá hā]                              want to go often (10/15)

Note: the examples cited in 4.2.5, such as (84) [[do cwá] beche á] 'tired of abstaining from going', involve Descriptive Particles rather than verbs.

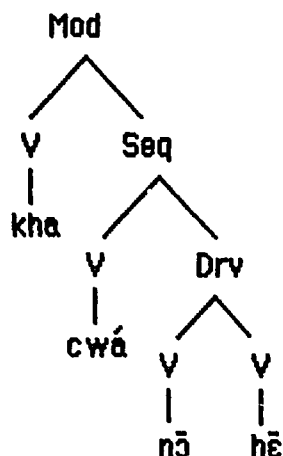
Modal with Directive

- (116) tō [dál cwá]                              should let sbdy go (5/26)
- (117) sí [nō ʔíre]                              want to tell sbdy to work (10/15)

Multiple: Mod + Seq + Drv

- (118) ʔa kha      cwá   nō      hē vē pò kúklē  
he promise go command go I YS field  
He promised to go tell my younger brother  
to go out to [work] the field. (10/13)

The structure is given below:



#### 4.4.2. Particles.

The *khwe*-class and *ná*-class particles are unproblematic in their interactions with verbs and with other particles: they are always the outermost constituents, occupying first and last positions respectively in the VC, and they seem to have the whole of their co-constituent in their semantic scope. E.g.

- (119) tarē cwá jè á  
 almost go difficult NS  
 almost hard to go (not hard to almost go) (10/15)

It is consistent with this that a particle may apply to one part of its co-constituent rather than to another:

- (120) nō cwá pò sé  
 command go additionally back-again  
 again order him to go or order him to go again (10/13)

Note: I do not take the second reading as evidence for a structure like [nō [cwá pò sé]], the primary reason being that [cwá pò sé] does not otherwise act like a constituent

in V-V constructions. In (108) above 'tell somebody to work fast' is analyzed as [nō [ʔírɪ phrē]] both because of its meaning and because [ʔírɪ phrē] can act as a constituent of a V-V in other contexts; e.g. *ʔírɪ phrē jè* 'hard to work fast', *bɛ ʔírɪ phrē* 'must work fast'. But *\*cwá pò sɛ̀ jè* is impossible (the correct version is *cwá jè pò sɛ̀*).

It is not at all clear which of these two classes should be said to c-command the other; i.e. whether [tú [cwá thō á]] or [[tú cwá] thō á] is the correct analysis for 'just finished going'. This is largely because it is nearly impossible to imagine a context in which the relative scope of such abstract meaning could make a difference: in the example just given, is the finishing-going recently applicable, or is the recent going finished?

**Descriptive Particles:** the variable positioning of these and its semantic effects have already been described (4.3.3.2); essentially that would come under the heading of interaction with V-V's. There are also some noteworthy examples of interaction with

**Modal verbs:**

- (84) [dɔ cwá] bæche á  
 abstain go bored NS  
 be tired of abstaining from going (10/13)

- (85) [dɛ sípɔ cwá] cè to  
 decide go able NEG  
 be unable to decide to go (10/20)

Worthy of note here is the fact that these Descriptive particles have scope over the

modal verbs, while second verbs in Descriptive V-V's usually have *lower* scope than Modal verbs, as shown by (114-115) above. This is interesting because in other respects the Descriptive particles are like second verbs in Descriptive V-V's.

**Modal Particles:** little is known at this stage concerning regularities of interaction of this class. There are hints that the Modal Particles have lower scope than at least some V-V types, as in the following example including a Descriptive:

(121) [lā cwá] jè                      be hard to plunge-in and go (see 4.3.3.2 for *lā*)

**Bound Result Expressions:** as already mentioned, apart from the fact that the BRE's fail the main test of verbhood, the verb + BRE construction behaves exactly like a Resultative V-V. The similarity extends also to possible interactions with other constructions.

#### 4.4.3. Excursus: noun-incorporating verbs and particles.

A few verbs and particles have the property of incorporating classifiers or nouns into the VC. Since classifiers in other respects are considered a special type of noun, I use the term 'incorporation' to imply a similarity to the type of construction by that name in some North American languages. The similarity is not exact, however; noun incorporation as usually understood incorporates specified arguments of a verbal stem into the complex verb, while the classifier cannot be considered a specified argument.

Examples:

Classifier-incorporater *pé* 'by the \_\_\_ unit, so as to be \_\_\_ units'

(122) *ʔíchē pé kīlō lāi to*

sell kilo yet NEG

They don't yet sell it by the kilo. (10/29)

(123) *dē pé plu thō á*

put pile finish NS

They finished piling it up ('putting it in piles'). (10/29)

(124) *vē me tā phé pé khri belò*

1s do fall crack shard drinking-glass

I dropped the glass and it broke to bits. (10/29)

Noun-incorporater *thwā* 'be, become'

(125) *tēú nʌ lɛ thwā tēú sé nʌ*

fish Nʌ go-down be fish again Nʌ

The fish went down and came back to life.

(Cf. *ʔa thwā N pa* 'N is still alive', literally 'it is still an N')

(126) *sílù thwā kúpè lāi to*

caterpillar become butterfly yet NEG

The caterpillar hasn't yet become a butterfly. (10/29)

(127) *taré cu thwā thā lō á*

wax melt become water use-up NS

The wax has all melted into liquid. (10/29)

In (122-3), [*pé*+classifier] precedes *ʔí*-class particles, as does [*thwā*+noun] in

(125-6); in (127) [*thwā*+noun] precedes a verb in what is probably a Descriptive V-V, cf.

the additional possibilities *taré cu thwā thā jū* 'the wax melts easily' (*jū* 'easy'), *taré cu*

*thwā thā rā* 'the wax melted beforehand' (for *rā* see 4.2.4). This is all evidence that these incorporating constructions are inside the VC. The incorporator+noun construction also precedes Descriptive Particles:

- (128) *vē ʔíchē pé kīlō cè to*  
 1s sell by-the kilo able NEG  
 I can't sell it by the kilo.

Note also:

- (129a) *ʔa ʔíchē sú pé kīlō*  
 3 sell wrong by-the kilo  
 He sold it by the kilo, incorrectly [i.e. it should be sold by the piece].
- (b) *ʔa ʔíchē pé kīlō sú*  
 id.

Here *pé kīlō* may either precede or follow the verb *sū* 'wrong', which is functioning as second verb in a Descriptive V-V; this suggests that the incorporator+N unit functions either as  $V_2$  of a Descriptive V-V itself or as Descriptive Particle.

Only these two incorporating morphemes are known thus far. Note that *pé* 'by the \_\_\_ unit' is probably a loan from Thai/Shan *pen*, a copula that can function in a similar manner; cf. Thai *man khāaj pen kīlao* 'it sells by the kilo'.

## 5. Formalizing the Analysis: Syntax or Compounding?

### 5.1. Introduction: many propositions, one clause.

The compounding analysis I have given the VC has thus far been a descriptive convenience only. In this section I will consider some alternatives to treating the VC as a compound. I will also make the compounding analysis more explicit, and will mention a few implications it has for morphological theory.

The crux of the problem presented by the Kayah VC is its property of apparently enabling sentences with simple, single-clause syntactic structure to express complex meanings consisting of more than one predicate. In the next section I will review the arguments for monoclausal structure that have been made in connection with various V-V types. I will then situate the Kayah V-V's in the context of constructions in other languages that also combine monoclausal structure with multi-predicate semantics, and will identify two complementary approaches that have been made to analysis of these constructions, one with a syntactic, transformational focus, the other with a morphological, lexical focus. In subsequent sections I will explore each type of approach as it might apply to the Kayah V-V's.

5.1.1. Review of arguments for monoclausality.

Compare the following sentences, containing respectively a single verb, a Directive, and a Sequential:

(1) vē dā Phāλ thā sō pē

Is give (name) water three bottle

I gave P. three bottles of water.

(2) vē nō phjá Phāλ thā sō pē

Is order take P. water three bottle

I told P. to take three bottles of water.

(3) vē cwá dā Phāλ thā sō pē

Is go give P. water three bottle

I went and gave P. three bottles of water.

All sentences have Subject, VC, 'Indirect Object', Object, and Extent expression, with no indication of embedding of a clause or second VP.

For Sequentials, the *khwa*-class and *ra*-class Particles, occur on either side of the Sequential unit, and cannot intervene:

(4a) ʔa mō | tú hē nō dé pa|ʔapò | thā

mother just-now go command dip-up DUR YS water

Mother just now went and told YS to draw water.

(b) \* ʔa mō tú hē pa nō dé ʔapò thā



Other form classes likewise combine with the Sequential as a unit:

(5a) ?ũ | b́ mō mé ho | lū

3 open look secretly 3OBV

Secretly they opened it (mosquito net) and peeked at him. (41.6)

Here the Descriptive Particle *ho* 'secretly' modifies *b́ mō mé* 'open and look' as a unit, and could not intervene between *b́ mō* and *mé*. Also, only the Object associated with the second verb may appear<sup>1</sup>:

(5b) \*?ũ | b́ mō mé ho | ?ikē thō

mosquito-net

Secretly they opened the mosquito net and peeked (at him).

Unfortunately I have no reliable data using clause-bounded phenomena, such as reference of anaphors, to demonstrate monoclausality of V-V sentences. The closest Kayah equivalents to English anaphors are the *rá*-class particles *áw* 'self, of one's own accord' and *lū* 'each other', and the morpheme *né* 'body, -self' (which is usually preceded by a personal pronoun); their properties, including boundedness of reference, are obscure pending further research. The following, however, is suggestive:

(6) ?a nō chū rá ?a nè

3 command stab RA 3 body

He<sub>i</sub> told [somebody<sub>j</sub>] to stab him(self)<sub>i~j</sub> (4/19)

where *?a né* as Patient/Object of the 'lower' verb *chū* can corefer with either the matrix Subject *?a* or the Causee (unexpressed in this example). This suggests either that

*ne* is not clause-bound, or that the object of the 'lower' verb *chū* is not in a separate clause, being accessible to the matrix subject.

There is one syntactic phenomenon in Kayah that is clearly clause-bounded, involving an alternation between an 'unmarked' and an 'obviative' form of the third person pronoun. Briefly, if a third-person pronoun is preceded by a non-coreferential third-person NP (pronominal or otherwise) in the same clause, it must be the obviative form *lū* (note the homophony with the reciprocal morpheme); otherwise it is the unmarked form *ʔa*. The Causee in a Directive V-V is susceptible to this effect:

- (7) ʔa ʔé ʔo jwā lū  
 3 call wait 3OBV  
 He<sub>i</sub> called to him<sub>j</sub> to wait. (157.1)

The interpretation of this fact hinges on one's assumptions about the nature of reference relations in general and of this obviative relation in particular. Given only the description in terms of 'the same clause' as just stated, it could be concluded that if the Object NP in (7) belonged to an embedded clause it could not be the obviative form, since the preceding noncoreferential third-person NP would not be in the same clause. On the other hand reference relations of this sort are not seen simply in terms of same-clause-hood in current theories like Government-Binding (the 'binding' of the title being a term for these reference relations); rather a notion of 'governing category' is used, which allows elements like *him* in *He expects him to win* to be both a constituent

of an embedded clause and at the same time accessible to the reference relation determining that *he* and *him* must be disjoint in reference. We may then wonder whether a) the Kayah structure is of the sort that allows this (exceptional) relation; and b) whether the obviation relation is subject to the same principles as those operating in the English sentence. Finally, the structure may still be biclausal even if *lū* is in the same clause as *ʔə*, with *lū* as Object of the matrix clause and an embedded clause including *ʔə jwā* 'wait for' and a zero Subject NP that is coindexed with the matrix Object *lū*<sup>2</sup>.

Let us return to the question of how, and in what component of grammar, the VC and its components are formed. Linguistic theory offers at most two methods for accounting for the combination of morphemes, namely base generation and, in some theories, transformation. Base generation may further be divided into generation by the syntax and combination by morphology (the latter understood to cover formation of compound words as well as of root+affix structures).

We may remark at the outset that it is to be expected that the proper analysis of the VC will be non-obvious: Kayah sentences with V-V constructions are (seemingly) unitary structures that (seemingly) contain more than one verb; this is bound to be problematic for any linguistic theory that takes the single-verb clause to be a (or the) basic unit of syntax.

To gain a broader perspective on the question, let us consider the V-V's in terms of the general semantic relations holding between the two constituent verbs in the five

types.

### 5.1.2. V-V as related events.

5.1.2.1. Predication vs. Modification: 1. A first division can be made between Sequential, Resultative and Directive on one hand and the Modal and Descriptive V-V's on the other. In the former, both verbs predicate (rather than modify), and the relation between them is one of related events (using 'event' in its broadest sense, to cover actions, processes, and states). In the latter, the relation between the two constituents is more like modification, with one verb applying its meaning to the other verb, and the other verb serving as predicator for the NP constituents of the clause. This division corresponds, of course, to the division between those V-V's for which derived argument structure is an appropriate analysis and those for which it is not. If my characterization is accurate, the Modal and Descriptive V-V's are instances of adverbial modification. The semantics of adverbs is a complex and problematic topic, and I will not pursue it here; suffice it to say that when a verb is modifying, rather than predicating, its argument structure does not interact with that of its head in the manner I have been calling derived argument structure.

5.1.2.2. Causation vs. Sequence. Of the predicating types, a further division can be made according to whether the notion of causation is involved:

- I. Modifying
- II. Predicating
  - A. Causative
    - 1. Resultative
    - 2. Directive
  - B. Sequential

This can be viewed in terms of different types of relations between events. In the Causatives, event 1 is the cause of event 2. In the Sequential, event 1 precedes event 2 in time, and may at the same time have event 2 as its purpose; event 1 may also alternate with or be simultaneous with event 2. Note that we are discussing semantic values of linguistic constructions, not a theory of action or causation. An actual series of events could easily combinemost or all of the characteristics just listed, as when event 1 causes event 2 (like the Causative), and event 1 temporally precedes and is intended to cause event 2 (like the Sequential). The difference between Causative and Sequential is rather a matter of emphasis, with the former emphasizing causation and the latter emphasizing purpose and temporal sequence.

5.1.2.3. Common Participant. The Causative-Sequential distinction of course by no means exhausts the possible relations between events (or even the possible linguistically-encodable relations). The event-relations depicted by the (predicating)

V-V's are only several instances of a general class of relations that is distinguished by being relatively closely linked. Less closely linked events can be shown in Kayah by sequences of full clauses, often with various connecting morphemes (*bā* 'and then', *ma* 'marker of topic-comment, antecedent-consequent, etc.', and others, cf. 9.4 below).

Worthy of mention at this point is a type of relation between events that is fairly close, but is not expressed by V-V's: this is the quotative type, which can be said to involve a relation between an individual (a talker or thinker) and a proposition (the content of the speech or thought). The type is represented in English, in addition to *think say, know*, by verbs like *expect* ( — *that he will come*, — *him to come*). In Kayah these are limited to *hē* 'say' and *ne* 'think', which take the 'content' clause as a preposed Subject or Topic (analysis not completely certain; see 9.3.3 below); e.g.

(8) [ʔa me pā hú tē ] dā dā vē ne nʌ  
 3 do IRR like what worry 1s think Nʌ

'If you do that, how would it be?' I thought worriedly. (299.5)

Returning to the predicating V-V's, one of their distinguishing features is the fact that the linked events have at least one participant in common; this can be seen as a corollary to the close link between the events. The roles of this common participant provide a second way of characterizing the difference between the Causatives and the Sequential. In the Causatives, the common participant is the recipient of action in event 1 and the agent in event 2; in Sequentials, the common participant is agent in both events. These are of course the facts that I have been attempting to describe in terms of derived

argument structure.

The two Causative V-V's can be differentiated according to features of their common participant: Directives have a human participant who retains some degree of control over the caused event; Resultatives do not specify humanness of the common participant, and the common participant retains little or no control over the caused event.

### 5.1.3. Skewing of syntax and propositional semantics.

5.1.3.1. the Problem of skewing. Now, prototypically at least, there is a one-to-one relation holding between syntax, predicate/role semantics, and possibly also logical representation if present: a single clause represents a single predicate with its set of roles, which in turn embodies a single proposition. In some theories, this is not just a prototype but an exceptionless principle, e.g. the Projection Principle of Government-Binding theory and others ('projection' because syntactic structures can then be simply projected from lexical features). The Kayah V-V's show the close link between the events they depict by representing them with (apparently) monoclausal structure, i.e. by casting a multi-propositional meaning in (apparently) mono-propositional form.

This skewing between propositional meaning and syntactic form is characteristic of similar constructions in numerous other languages, causatives being one such type that has achieved prominence in the literature. The Sequential V-V makes it clear, however, that causatives are only one instance of a more general phenomenon. It is to be expected that constructions with this type of skewing will present analytical problems to an

Linguistic theory taking the single-verb clause as a basic unit of syntax, even if it is not elevated to *the* unit, as by the Projection Principle.

It should be pointed out that the term 'causative' is often used to refer to a particular type of construction, also called the morphological causative. This is the sort found in Turkish, Japanese, and many other languages, in which a morpheme with the abstract meaning 'cause' attaches to a verb root to produce a derivated causative verb. E.g.

(Turkish):

- (9) dişçi mektub-u müdür-e imzala - t - tı  
 dentist letter-ACC director-DAT sign - CAUS- Past  
 The dentist got the director to sign the letter. (Comrie, 169, #14)

But this is not the only type of construction that involves the notion of causation in a prominent way. In particular, given the concept of causation as a relation between two events, it is evident that the Turkish-style morphological causative suppresses the causing event. In the above example it is simply left unclear how the dentist brought about the result; a logical translation of the sentence in terms of two propositions might represent the causing event as the equivalent of 'the dentist did something'. The Kayah Resultative and Directive V-V's, in contrast, do include the causing event, as discussed. We have also seen that the inclusion of both causing and caused events further involves the presence of the participant common to both, which may be called the 'pivot' (the term is borrowed from Chao). The pivot adds a third problematic characteristic to those of monoclasality and multipropositionality: some way must be found of representing the



overlapping semantic roles borne by the pivot.

5.1.3.2. Multi-level syntax: Transformational and control approaches. The problem of analysis, then, is to deal with the conjunction of multi-propositional semantics, represented by multiple verbs, and mono-clausal syntax. The multiple propositions must be telescoped into a single unit at some level of representation; depending on characteristics of the particular construction and language, various approaches will be more or less appropriate. One way is to 'deconstruct' the unitary syntax, to claim that what appears to be a unitary syntactic structure is actually complex, which amounts to representing multi-propositionality at a deeper level of syntax, then having the telescoping operation take place in the relation of that level to the shallower level(s).

In the case of causatives this is commonly done by analyzing the causing predicate as a matrix clause with the result predicate as a lower clause embedded in it. The overlapping semantic roles of the pivot participant are then easily shown by having the pivot occupy both the object position of the matrix clause and the subject position of the embedded clause. For the morphological causatives this involves deconstruction of not only of syntax, but also of lexical units: the causative affix is analyzed as being underlyingly a verb. This approach for causatives is embodied in the Clause Union transformation of Relational Grammar treatments. Again, causatives are not the only construction type in question: for the Sequential type, the prior event could be taken to be the matrix clause, with the subsequent event embedded in a manner similar to a

purpose clause.

The same general strategy is represented in the earlier generative literature by the transformations Object Raising and (object-controlled) Equi-NP deletion. The English constructions that these account for are not strikingly monoclausal in the manner of morphological causatives; they are however, less than absolute in their biclausality.

Comparing canonical examples like:

(10a) I believe that John ate the pie

(b) I believe John to have eaten the pie

we see that the complementizer *that* in (10a) clearly marks *John ate the pie* as an autonomous embedded clause, while in (10b) the boundary between the matrix and embedded clauses is more vague: *John* seems to have ties to both clauses at once, in that it has the Agent role in relation to the embedded verb *ate*, and it also has syntactic characteristics consistent with it being the Object in the matrix clause (accusative case as shown by the fact that the pronominal equivalent would be *him*, and passivizability, as shown by *John was believed to have eaten the pie*). Additionally, the embedded clause now contains an infinitive verb. This example is of the Raising type: in underlying structure, the NP *John* has the Subject position in the embedded clause, allowing for its semantic role in that clause; it is then raised to the position of matrix Object, allowing it to acquire the syntactic characteristics of that position. More closely related to the present topic is the Equi type, which is where the English causatives are found: matrix

verbs like *force*, *persuade*, *compel*, and so on. In

(11) I forced John to eat the pie

*John* again has ties to both clauses, this time semantic in both instances: it is the Agent of *eat* and the Patient of *forced*. In other words *John* is the common participant, the causee, and in the Equi analysis this is shown by having *John* occupy two positions: I forced John [<sub>S</sub> John eat the pie ]. The lower-sentence *John* is then deleted under identity with the upper-sentence *John*. The strategy here is less radical than in the case of morphological causatives, since the surface form is already partially biclausal: one need only posit an underlying form in which the biclausality is unambiguous, the transformations then blurring the biclausality, but not effacing it entirely.

Since the analyses described so far require a transformation to mediate between the underlying multi-clause structure and the monoclausal surface structure, this may be dubbed the transformational approach.

A second approach can be described which can be placed under the heading of 'syntactic' along with the transformational approach; it is in fact the replacement of Equi in later generative theory. It can be called the dependency or control approach, and posits a biclausal structure, biclausal at the surface as well, but in which the embedded clause has special features that result in its subject position being in a sense defective. In one version, associated with Government-Binding theory, the subject position is present but filled by the phonologically empty element PRO. A set of principles, control theory,

determines the coreference between PRO and some NP in the matrix clause. In a second version the structure is not strictly speaking *bi-clausal*, but the embedded element is a VP (the subject position thus being absent entirely), functioning as a predicate whose subject, located elsewhere in the clause, is determined by predication theory<sup>3</sup>.

The control approach is something of a compromise: a complex structure, different from that of a simple clause, is retained at the surface, but its subsidiary part (the erstwhile embedded clause) is something less than a full clause.

5.1.3.3. Multi-level lexical structure: Lexicalist approach. The mirror image of the transformational approach is to accept monoclausal syntax, but to show multi-propositionality and the causative relation within the structure of the complex lexical item. For instance the Turkish form *imzala-t-ti* 'caused to sign' combines the lexical structures of *imzala*'sign', a two-argument predicate, and the causative *-t-*, a one-argument predicate, to produce a new lexical item that has added a) a causer, to produce a three-argument predicate, and b) the meaning 'cause'. The syntactic structure into which this complex verb is inserted then has no need to show multiple clauses.

Even when the causative morpheme is an autonomous verb, a lexicalist analysis may amalgamate the verbs, claiming that what appears to be several verbs is actually a unit, perhaps of verb plus affix(es), perhaps a compound. A well-known instance of this is several treatments of Romance causatives, which appear to consist of a verb preceded by the verb 'do' or 'let' (e.g. French *faire, laisser*). E.g. Zubizarreta (1985) argues that

Romance causatives [i.e. the verbs 'do', 'let', etc.--DBS], although morphophonologically words...function morphosyntactically as bound morphemes. (p. 247)

For a different instance of the same general approach, dealing with constructions more closely akin to the Kayah V-V's, cf. Thompson (1973) on Mandarin resultatives. The amalgamated verbs, although unitary at the syntactic level, will still have a multi-propositional semantics. This, which can be called the lexicalist approach (since current theory assigns the function of word formation to the lexicon), amounts to representing multi-propositionality in the morphology/lexicon, since the telescoping is done by the process of word-formation, 'prior' to lexical insertion.

#### 5.1.4. Afterword.

This is perhaps the appropriate spot to point out that, as the reader may have already noticed, the difference between a biclausal and a monoclausal treatment of the Kayah V-V's is not simply a matter of reanalysis. This is unlike the Turkish-type causative, for which the difference between the two can be shown by changing the bracketing only, as follows (bracket labels for illustration only):

biclausal [<sub>S</sub> disçi [<sub>VP</sub> [<sub>S</sub> mektub-u müdür-e imzala <sub>S</sub>] -t-tı <sub>VP</sub>] <sub>S</sub>]

monoclausal [<sub>S</sub> disçi [<sub>VP</sub> mektub-u müdür-e imzala-t-tı <sub>VP</sub>]<sub>S</sub>]

dentist letter-ACC director-DAT sign-'CAUSE'-Past

The dentist got the director to sign the letter.

(One would also have to account for the dative case of the embedded subject *müdüre*).

The point is that no movement needs to be posited, the surface constituent order being compatible with either biclausal or monoclausal structure.

In contrast, the Kayah constructions will require a movement operation. This means that if a biclausal underlying form is to be converted into a monoclausal surface structure, two transformations will be necessary: a) reanalysis, to remove the clause boundaries of the embedded clause; b) movement, to make the two verbs adjacent. Note that b) will be needed even if a) is not (i.e. if a monoclausal analysis of the surface structure is not accepted). Ancillary to these two would be c) some means of dealing with the linking of the pivot NP (the causee) to both the Object position of the matrix verb and the Subject position of the embedded verb. To illustrate (I use X for the embedded subject, to equivocate between PRO and a full NP identical with the matrix object) with the sentence 'I told Pha'a to get water':

underlying: [<sub>S</sub> vĕ nō Phāλ [<sub>S</sub> X dé thā ]]

I tell P. dip-up water

reanalysis: [<sub>S</sub> vĕ nō Phāλ X dé thā ]

movement: [<sub>S</sub> vĕ nō (X) dé Phāλ thā ]

This omits consideration of a) whether to move the matrix object *Phāλ* rightward or the embedded verb *dé* leftward; b) what to do with the embedded subject X (delete it, keep it as phonologically null, etc.).

Note that movement and reanalysis are logically separate; in section 5.2.2 below we will consider a specific transformational analysis that involves the former but not the latter.

## 5.2. Multi-level syntax: Transformationalist approach.

Since Kayah has neither *wh*-movement nor passive (nor any other construction for which NP-movement could be invoked), the V-V's are the only area of the grammar which might call for an analysis involving movement, or some other way of relating multiple levels of syntactic representation. However their combination of multi-propositional meaning with monoclausal surface form, as described above, makes some sort of multi-level treatment virtually a necessity. The V-V's are to a large extent amenable to either of the two main approaches described, the syntactic and the morphological. We will consider the syntactic approach first.

Accounting for the V-V's in the syntax seems attractive in several ways. Multi-propositionality aside, the VC constructions have the 'feel' of syntactic structures, for two reasons. First, they are highly productive, and clearly different from lexicalized compound verbs such as the group containing *sí* 'heart' (e.g. *síne* 'wake up', *sí tarē* 'embarrassed', *sí ŋē* 'know', etc.). Second, they are semantically transparent: to understand them one only needs to know the meanings of the component morphemes, and perhaps also the meaning (if we may put it that way) of the construction. It is difficult

to come up with examples of morphemes that have a specialized meaning in one particular combination (but see 5.3.1). (Note that this is not the same thing as those verbs that show a meaning shift when occurring in a particular position in a V-V; *dā* 'give' means 'let' when it is the first verb of a Directive V-V, but it always means 'let' in that position, regardless of what verb it precedes).

### 5.2.1. Syntactic Base-Generation.

Before exploring a transformationalist approach, let us first consider the possibility that the constructions in the VC can be accounted for by the phrase structure rules of the base. Setting aside the multi-propositional semantics for the moment, it might be taken as the null hypothesis that the V-V's are directly generated, without reference to transformations.

The focus of a base-generation account of the VC would be on how to formulate the rules expanding VC. Recalling that the syntax of sentences with complex VC's is no different (apart from the complexity of the VC) from the syntax of sentences with simple, even monomorphemic VC's, it may be concluded that the rules accounting for complex VC's should introduce no new S or VP categories.

The VC might be of the category  $\bar{V}$ , intermediate in level between the lexical (V) and the maximal (VP). It could be formed by a rule like  $\bar{V} \rightarrow V \bar{V}$ , which would give the right-branching structure that seems to be predominant, as discussed in 4.4; recall



[Seq V [Res VV]] *vē puì me sã jò khró* 'I caught and killed a rat'; and [Drv V [Seq VV]] *nš kə̀lɛ dɛ̀ thã* 'tell sbdy to go down and draw water. However left-branching structures are also needed: [Seq [Res V V] V] *bš mō mɛ̀* 'open ('reach'+ 'be-open') stg and look'. It is clear that whatever analysis is adopted will have to cope with the complexity somehow. A possibility for a syntactic analysis is to have a maximally simple phrase-structure rule  $V \rightarrow V V$ , in conjunction with constraints on configurational relations between forms or form classes; the c-command constraints stated in 4.4. are an example of this.

The aspect of the VC that creates the most problems for a syntactic-generation approach is that of subcategorization. In the standard approach, the number and type of NP arguments present or possible in the sentence is reflected in the subcategorization features of the sentence's main verb. However in a Kayah sentence it is not sufficient to refer only to the main verb. In the Directive V-V, for instance, the number and type of arguments occurring in the sentence is a function of the argument structures of both verbs involved. There are also Particles that have an effect on argument number and type by allowing an additional argument in the clause. These have been described in several sections above; they include the Bound Directionals like *talwá* 'past' and others (3.3.7); the Descriptive Particles *cwá* 'help', *bébw* 'guide, show the way to' (3.3.6), *lɛ̀* 'more, than', and *khrw* 'equally, as \_\_\_ as'; and the *kã*-class Particles *kã* 'with, comitative' and *pɛ̀* 'for, ethical dative' (3.3.5). I have argued in favor of analyzing these morphemes as

co-constituents of the verb rather than prepositions directly in constituency with 'their' arguments (the arguments that they add to the VC). The reason is the same in all cases, and is simply that these morphemes need not be adjacent to their arguments (with the exception of *pē*, which happens to occupy the final position in the VC). Here we may add that there are no alternative ordering possibilities; e.g. the sentence 'he has eaten more rice than me' includes the morphemes *l̄x̄* 'more than', *thō̄* 'perfective', and *vē̄* 'I', and they must be in that order:

(10) ʔa ʔe ʔé l̄x̄ thō̄ vē̄ dī  
 3 eat much more-than PFV 1s rice

(11) \*ʔa ʔe ʔé thō̄ l̄x̄ vē̄ dī

(12) \*ʔa ʔe ʔé thō̄ dī l̄x̄ vē̄

It would thus be implausible to claim that *l̄x̄* has been moved from a position next to *vē̄* in a structure like (11) or (12). All of this means that the number of possible arguments (and, in some cases, their interpretations) is not simply a matter of features of the main V, but is a function of those features as modified by other VC constituents. It might seem simplest to reflect this fact by allowing that function to assign features to the VC as a whole. But in standard generative theory these subcategorization features are usually considered to be properties only of lexical items, not of phrasal categories<sup>4</sup>; to be consistent with this the VC should be considered a compound V.

Consider how the subcategorization facts would need to be stated in a generative

grammar in which the VC was produced by phrase-structure rules. The lexical entries for verbs would include subcategorization features, as usual. A verb such as *cwá* 'go' would be subcategorized as [+NP \_\_\_] (ignoring for the moment the Locative argument that should probably be included), stating that it takes a single preposed NP argument, a Subject. But if *cwá* appears in a V-V preceding a transitive verb like *phjá* 'take', the sentence will include an Object NP as well; i.e. *cwá* can also appear in the frame

$$\begin{array}{c} [\text{NP} \quad \_ \quad \text{V} \quad \text{NP}] \\ \quad \quad \quad [+NP\_NP] \end{array}$$

*cwá* may also be followed by an argument-adding morpheme like *kā* 'with, etc.', with an NP Object again possible, as in *cwá kā vē* 'go with me'. The subcategorization must then include a feature like [+NP \_\_\_\_kā...NP] (dots indicate that other elements may precede and follow *kā*), and must add a similar feature for each of the six or so other valence-affecting morphemes. But this is also a violation of standard generative practice, since subcategorization usually refers to syntactic categories only, not lexical items; this cannot be done here since only certain members of each form class affect valence: e.g. *kā* is a *kā*-class Particle, but most of the *kā*-class Particles have no effect on valence, so a feature like [+NP \_\_\_ Ptc NP] would also be incorrect. A subcategorization frame that included *kā* would be equivalent in English to subcategorizing intransitive verbs as [+NP out-\_\_\_ NP], to show that the prefix *out-* produces words like *outrun*, *outlast*, *outwait* that take Direct Objects. Also, subcategorization of V's usually refers to possible arguments of the V, and *kā* and the

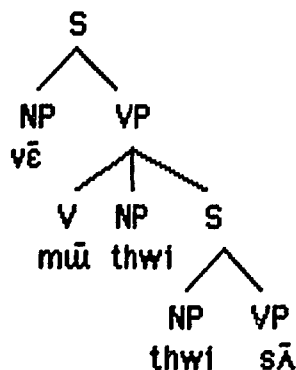
like are not arguments of the V in any usual sense of the term.

### 5.2.2. A Transformationalist analysis: Verb Incorporation.

We will now consider a particular transformational analysis of the V-V's. First we may briefly list the advantages of any such analysis in general, as well as the form it would take.

In the Resultative V-V, for instance, the overlapping roles of the pivot participant (the 'Patient-to-Patient' mapping in the derived argument structure) can be shown by positing the type of underlying structure already mentioned, with embedding:

(13) underlying structure of *vĕ mū sĀ thwĭ* 'I beat the dog to death'



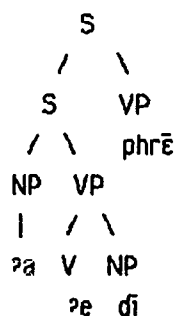
(the subject of the lower clause could also be PRO, in the control approach.) This underlying structure provides a natural way to display the information that the Object *thwĭ* has the Patient role of both the upper and the lower verbs. Transformations are then required to accomplish movement and reanalysis, as discussed above. A further advantage of this approach is that it would facilitate typological comparison: in the case

of Resultatives, it provides an easy way of relating the Kayah sentence to its Thai equivalent *chǎn tii mǎa tǎaj* (literally 'I hit dog die'), which could be said to differ only in not having undergone the movement transformation. In certain contexts, Vietnamese has both types, e.g. *tôi làm đơ quần áo* ~ *tôi làm quần áo đơ*, both 'I made the clothes dirty'; the transformation would then be said to be optional<sup>5</sup>.

Finally, other types of underlying structure can be imagined: for instance, instead of embedding one sentence in the other, the two sentences could be arguments of an abstract predicate CAUSE; e.g. [vĕ mū thwi] CAUSE [thwi sǎ].

Similar underlying structures can be easily imagined for the other types of V-V: in a Sequence V-V the Subjects would be either identical or coindexed, in a Descriptive V-V the second verb would be the main verb to which the remainder of the sentence functions as sentential Subject:

(14) ?a ?e phrĕ dī 'he eats rice fast'



It is easy enough to imagine underlying structures and complex transformations modifying them. To seriously evaluate a transformational analysis, it must be situated in some theory that is sufficiently constrained to be realistic. The specific proposal we

will examine is one of the most recent in the transformationalist vein, Baker's (1985)

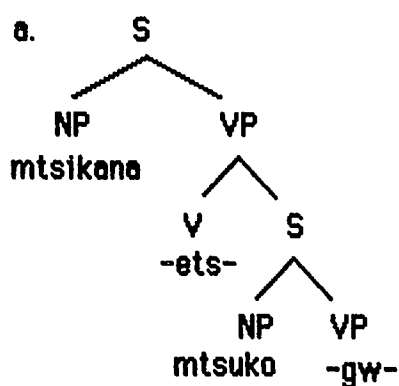
### Verb Incorporation.

Baker considers only morphological causative constructions, those in which the causative morpheme is a verbal affix. He analyzes this affix as underlyingly a verb that takes a sentential complement, the verb of the sentential complement being moved up to adjoin to the causative affix. As a concrete instance we may cite Baker's example from Chichewa: (his 3a, p.205)

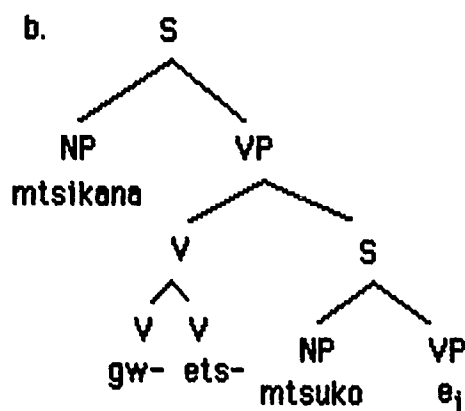
- (15) mtsikana anau-gw-ets-a mtsuko  
 girl agr-fall-made waterpot  
 The girl made the waterpot fall.

This is analyzed in terms of the two structures below ('make' represents the causative affix *ets-*):

underlying structure



derived structure



The lower verb *-gw-*'fall' moves to adjoin to the causative morpheme *-e's-*.

The name 'Verb incorporation' is meant to reflect a similarity to constructions found in many American Indian languages, known as Noun Incorporation, in which a complex verb 'incorporates' a nominal stem (the interpretation is usually that the nominal relates to the verb in a way that would be translated as Object in more familiar languages). Baker's treatment of Verb Incorporation is only one section of a general theory of incorporation.

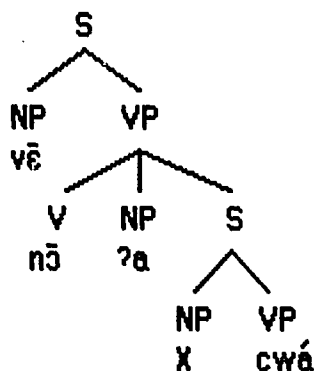
As Baker notes, Verb Incorporation is only the latest in a series of similar proposals about such constructions, including the Predicate Raising of Generative Semantics and the Clause Union of Relational Grammar. A major difference (and an advantage, in Baker's view) is that Baker operates in the framework of Government-Binding theory, which sets various sorts of constraints on the possible structures and processes involved. In particular, the movement of the lower verb is treated as an instance of the generalized transformation 'move alpha'; the movement therefore leaves a trace, and certain conditions must be fulfilled regarding the configurational relation between the moved verb and its trace.

We may first consider some apparent differences between the constructions treated by Baker and the Kayah V-V's; not all of these are critical. First, the Kayah constructions involve only full (i.e. lexical) morphemes, while Baker's causatives involve combination of an affix with a full morpheme. If anything, combination of full morphemes should be more congenial to an Incorporation analysis, since it obviates whatever difficulties there

may be in having the causative morpheme change its category from verb to affix when it moves. Baker in fact states that (p.232) 'it should also be possible for syntactic Verb Incorporation to correspond to morphological compounding in some languages'. The Kayah V-V's might simply provide the predicted filler of a gap in the set of possibilities.

Secondly, Baker's causative morpheme, being simply an affix with the abstract meaning 'cause', is fittingly analyzed as taking a sentential complement; that is to say, it makes sense to view the predicate 'cause' as a relation between an individual (the causer) and a proposition (the caused event). The Kayah equivalents of this morpheme are those verbs appearing in first position in the V-V's with causative meaning, namely the Resultative and the Directive. But the Kayah verbs have more content than the abstract 'cause', since they also describe the causing action<sup>6</sup>. This is what produces the effect of the overlapping semantic role of the causee (or pivot), which not only performs the caused action, but also receives the effect of the causing action. This difference could presumably be accommodated in Baker's analysis by saying that the Kayah verbs take both an NP and a sentential complement, as in the 'Equi' structure suggested above; for instance, the underlying structure of *vē nō cwā ʔə* 'I told him to go' would be:





In Baker's framework the lower subject would presumably be the abstract element PRO, which would relate to the upper-clause object according to control.

A third discrepancy is that the Kayah V-V's include not only those with causative meanings, but also the Sequential and Descriptive types. However Baker states that the causatives he treats are 'part of a somewhat more general phenomenon of Verb Incorporation', citing examples that include semantic equivalents of Sequentials (translation: 'I am going to beg maize'), Descriptives ('The man broke the boat easily') and Modals ('The man wants to see the woman'). It is not too difficult to construct analyses of these additional cases with a matrix verb plus sentential complement.

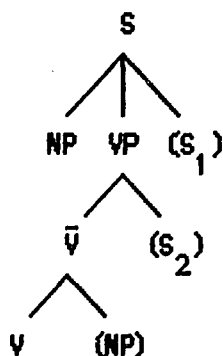
There is no grave objection to subcategorizing the Directive verbs for a sentential complement; they form a small class with special characteristics that are appropriately captured by some such description. Also those with meaning shift, like *dh'* 'give; let' can be said to require a second verb (in this analysis, the sentential complement) in their directive meaning (=the Drv verb of the related pair). But difficulties arise when one

tries extending the analysis to the Resultatives. The first position in a Resultative V-V can be occupied by virtually any verb, or at least by a very large subset of all verbs, and these verbs are completely indifferent to whether they occur in a Resultative construction; i.e. whether they are followed by a second verb. It thus makes no sense to posit a subcategorized sentential complement as underlying a Resultative V-V, unless we are willing to have virtually every transitive verb in the language be subcategorized in that manner--and the subcategorization would have to be optional. The expression denoting result is not conceptually intrinsic to the transitive verbs in the manner of the complement of a causative verb like *persuade*. It is true that the complement of *persuade* can be omitted: *I persuaded John* is grammatical; however it invites a question asking for the complement to be filled in, e.g. *To do what?* But a Kayah transitive verb like *mū* 'hit' does not omit anything: *vē mū thwi* 'I hit the dog' does not invite a question like 'with what result?' (I am frankly not sure how that would be asked).

Similar arguments would apply to the Sequential and Descriptive V-V's. To account for Sequentials a great many verbs would need optional subcategorization for a sentential complement, and one whose subject is somehow guaranteed identity or coreference with the upper subject. Descriptives would presumably take sentential subjects<sup>7</sup> as an alternative to NP subjects; e.g. 'he eats rice fast' [<sub>S</sub> a ?e dī] phrē → [<sub>S</sub>?a ?e phrē<sub>i</sub> dī] t<sub>i</sub>.

Clearly the lower sentence in all of these types must be an optional constituent. But it could not be a  $\bar{V}$ -external constituent like a purpose clause (although purpose is indeed one of the readings of the Sequential construction). If it were, the lower sentence would not be a sister to the matrix verb but to  $\bar{V}$  (as is  $S_2$  below), or possibly a sister of VP (as

$S_1$  below):



But then the lower verb will be ineligible for movement into the VP, in Baker's theory 'an X- $\alpha$  may only move into the Y- $\alpha$  that properly governs it' (208, citing Travis 1984). But X governs Y only if a) Y is contained in the maximal projection of X and b) X c-commands Y (van Riemsdijk and Williams, 291). Here X is the matrix verb, its maximal projection is VP, and thus if the lower verb Y is outside of the VP, as in  $S_1$ , it cannot be moved into it. If the lower verb is in the position  $S_2$ , it will be in the maximal projection VP, but it will not be c-commanded by the matrix verb (at least under the standard definition of c-command), since the branching category  $\bar{V}$  intervenes. The lower clause

must then be in  $\bar{V}$  for movement to be possible out of it. This may or may not be a serious difficulty, depending on whether one allows for non-subcategorized constituents of  $\bar{V}$ , which may also be required for 'predication structures' (like *eat the meat raw*, mentioned above)<sup>8</sup>.

A more serious obstacle is the fact that the putative biclausal structure never occurs on the surface; in other words the incorporation transformation must be obligatory. It is also obligatory for the morphological causatives, but there is a natural explanation: since the matrix 'verb' is really an affix, it must attach to some lexical morpheme (in Baker's terms, it is subcategorized for attachment to a verb), hence the transformation must apply to fulfill the subcategorization of the matrix verb. But the Kayah morphemes are all full verbs and have no need to attach to any particular lexical category.

Thus far it seems that Incorporation, although problematic for Resultatives, Sequentials and Descriptives, might still be appropriate for the Directives. Returning now to the Directives, we have said above that their sentential complements would have PRO as subject. To expand slightly on the reasons for this assumption: their sentential complements would have to contain a subject that is referentially dependent on the object NP of the higher clause. But that dependence could not arise by movement, since the two NP's have different semantic roles--that is indeed the motivation for positing two NP's in the first place. In GB terminology, it is a violation of principle (the 'theta criterion') to have a moved NP and its trace bearing different thematic roles. So the

lower subject cannot be 'NP-trace'. What is said to occupy this position is the base-generated empty NP PRO; this is the analysis of the English equivalents of the Directives, namely sentences with verbs like *tell, persuade, invite* followed by infinitival complements. PRO, in fact, can only appear as the subject of an infinitive. The problem is that Kayah verbs, of course, do not distinguish finite from infinitive forms. One could claim that the second verb in a V-V is indeed an infinitive, but one which happens not to have overt marking of its infinitival character. It would of course be desirable to avoid making a claim that is to such a degree ad hoc and theory-bound; a more promising line would be to discard occurrence in nonfinite clauses as a criterial feature of PRO.

A second difficulty for an Incorporation analysis of Directive V-V's arises in connection with complex V-V's. Recall that a Directive can, among other possibilities, act as  $V_2$  of a Sequential; e.g.

(16)  $v\bar{e} h\bar{e} n\bar{o}$   $\text{ʔir}\bar{e}$   $Ph\bar{a}\lambda$       I went and told Pha'a to work

But we have already seen that Sequentials at least cannot be produced by transformation; they must then be base-generated, whether syntactically or by compounding. But we have also seen that they are not easily accounted for by syntactic base generation (cf. the discussion of *cwá phjá* 'go and take' in the preceding section). Thus the construction  $h\bar{e} n\bar{o}$  'go and order' must be a compound, and would have to be formed first. The Directive would then have to be 'embedded' in the Sequential by

adjoining of the lower verb *ʔírē* 'work' to the Directive verb *n̄* 'command'. This would be a clear violation of the principle of Lexical Integrity, which states that syntactic processes, especially referential indexing, cannot pick out parts of a word (Jackendoff 1972, Selkirk 1982, Huang 1984). In this case the Incorporated verb *ʔírē* would be coindexed with its trace in the lower sentence, thus requiring the trace to corefer with a subpart of the complex word *hē n̄ ʔírē*.

I would remark, finally, that perhaps the operative difference between the causatives treated by Baker and the Kayah V-V's is that the morphological causatives are at once more particular and more abstract. More particular in that only one morpheme is involved (or perhaps a few, since some languages seem to have one for 'cause' and one for 'let'); more abstract in that the morpheme has no content other than 'cause'. The corresponding Kayah morphemes, in contrast, are vastly more general in their numbers and at the same time more particular in their meanings, which makes difficulties for the positing of an underlying structure with sentential complement.

### 5.3. Multilevel lexical units: lexicalist (Compounding) analysis.

Having considered movement and syntactic base-generation as analyses of the VC, we proceed to explore the remaining alternative, generation by the morphology, which in this case amounts to compounding.

Note that we have already implicitly made a case for some sort of word-forming rule, by showing that the argument-adding Particles in the VC cannot simply be base-generated in constituency with verbs (5.2.1 above). Presumably forms like *cwá ká* 'go with', effectively a transitive verb, are then best accounted for by lexical operations. This section will be devoted to applying such an analysis to the V-V's.

#### 5.3.1. Characteristics of compounds vs. phrases.

Intuitively the constructions of the VC do not look like typical compounds; as already discussed, they have the feel of syntactic phrases, being highly productive and semantically transparent. Also there is no phonological evidence, such as stress patterning or tone sandhi, that would indicate that the V-V's are words. In contrast with the productivity and transparency of syntactic constructions, compounds are traditionally expected to be semantically idiosyncratic, with meanings not necessarily understandable as the sum of their parts (e.g. *six-shooter, whitewall*), and syntactically irregular, being often of limited productivity ( e.g. the verb-object pattern exemplified by *pickpocket, lickspittle, killjoy*, and not many more). However, it is uncontroversial that these are generalizations only. First, syntactic does not necessarily equal transparent:

there are numerous examples of phrasal idioms like *kick the bucket*, *get X's goat*, *pull X's leg*, which have idiosyncratic meanings but behave to varying degrees like syntactic phrases. Second, morphological does not necessarily equal idiosyncratic: there are highly productive word-forming processes, involving both affixes and full morphemes, whose output additionally may be quite transparent in meaning. Consider the pattern NP V-er: *cheese-eater*, *cat-hater*, *leaf-collector*, and so on, all of whose meanings can be accurately rendered as 'one who/that V's NP' (one who eats cheese, one who hates cats, one who/that collects leaves, etc.). There is clearly no need to list these separately in the lexicon; they can easily be accounted for by a single process, whether that process is described as a word-formation rule or as something else.

On this account there is then no reason to deny that the VC constructions are compounds on the basis of their productivity and transparency. It is, however, still slightly incongruous that there seem to be so few exceptions to transparency. More typically, even the most productive compound-forms have non-transparent instances: in addition to the transparent words of the *cheese-eater* type listed above there are also words of identical form but non-transparent meaning: *skyscraper*, *cow-catcher*, *jawbreaker*, *windbreaker*, *bee-eater* (kind of bird). These will require their own lexical entries, although they still bear a relation in form to the transparent counterparts.

There are two points that can be made concerning the lack of idiomatic V-V constructions.



First, there are some possible examples, e.g.:

**kle ʔi** 'run + defecate' → 'have diarrhea'; clearly a Sequential, having the typical motion-verb as  $V_1$  and the meaning classifiable as 'alternating action'.

**chō t̄apa** 'forget'; first element restricted to this compound, second element probably identifiable with the second elements of *ʔo mā t̄apa* 'asleep' (*ʔo mā* 'lie down, sleep') and *ʔō mu t̄apa* 'dead drunk' (*ʔō* 'drink', *mu* 'drunk, poisoned'); Resultative in meaning ('forget so as to be unaware')

**t̄o ma** 'remember'; first element restricted to this compound; the second component probably *mā* 'be fixed, firm, steady'; Resultative, presumably with an original meaning like 'remember stg so that it stays [in the memory]' (cf. Chinese *jì-zhu* 'remember+stay', *jì bù-zhu* 'remember+not+stay→can't remember').

Other possible examples are rendered unclear by uncertainty of morpheme identity (as exhibited in the second and third examples, but worse).

Secondly, a more fundamental reason for the transparency of the V-V constructions is that they lie very close to the border between syntax and morphology. They thus, like morphological causatives, share in the features of canonical syntactic phrases and of canonical lexical formations, as described in (5.1.3), and a decision as to which side of the syntax-morphology border to locate them on requires detailed examination and must

be at least partly a theoretical one.

### 5.3.2. some principles of Lexicalist morphology.

To address these questions I will make use of some concepts associated with the theoretical approach often known as lexicalist morphology. This approach assumes that at least some (in some versions, all) word formation is accounted for in the lexicon rather than by syntactic transformations. In particular, I will assume that the Kayah V-V constructions are compounds that are formed by one or more compounding rules.

#### 5.3.2.1. headedness and percolation.

One of the assumptions of lexicalist morphology is that complex words, like syntactic phrases, have heads, and that the head determines the features of the structure that it heads (this is often put as an instance of the more general conformity of word structure to the  $\bar{X}$  theory of syntactic structure). This is formalized in Percolation Conventions (Lieber 1980, 1983), which state that features of the head 'percolate' up to the dominating category. This provides an easy way of showing, for example, that the compound word *blackboard* is a noun: the word as a whole has the feature [+N], because that feature has percolated up from the head of the word, the stem *board* which, being a noun, has that feature as one of its lexical specifications. A second percolation convention that will concern us here states that if a head is unspecified for some feature, that feature may be filled in by non-head members of the word.

### 5.3.2.2. formation rules or not?

In one version of lexicalist morphology, word-formation rules operate in the lexicon; the operations they perform include combining of morphemes (affixation, compounding), alterations of argument structure (e.g. causativization) , and/or alterations in realization of argument structure (e.g. passivization). See Aronoff 1976, Williams 1981, Selkirk 1982. In another version, there are no word-formation rules as such; rather

affixes as well as roots have their own lexical entries and all morphemes are inserted into unlabeled binary branching word structure trees. Marantz, 121

...the locus of all idiosyncratic information in the lexicon is in lexical entries; morphemes may have diacritics, or bear peculiar restrictions on their subcategorization, but the actual principles by which morphemes are combined into words are exceptionless. Lieber 1983, 254

It is not always evident how much difference this choice of formalism makes at an empirical level, especially given a sufficiently rich set of possible lexical features. For instance, Williams (1981) discusses a rule 'internalize an argument'<sup>9</sup> that produces, among others, verbs by suffixing *-ize* to adjectives; clearly *-ize* could be given a lexical entry including the feature [+Internalize] which would give the proper effects when *-ize* appears in a word structure.

When one turns from affixation to compounding the equivalence becomes less obvious. The differing kinds of derived argument structure mappings I have assigned to the V-V's are an obvious candidate for being the properties of differing word-formation rules, since

they are not likely to be features of the constituent verbs themselves. On the other hand it is not out of the question that the mapping facts may be derivable from independent principles or conditions, and I will suggest such an analysis in 5.3.3 below.

### 5.3.2.3. Argument linking.

Directly related to the status of specified arguments as intrinsic is the concept of argument linking, which states that all arguments specified by a verb (or other argument-taking item) must be 'linked' or 'satisfied' for the verb to legitimately appear in a syntactic structure. Several lexicalist treatments of English compounds assume that some version of this applies within complex words (Selkirk 1981, 1982; Lieber 1981, 1983). Cf.

...a particular syntactic (or morphological) structure containing a lexical item with a particular argument structure is ruled as well-formed only if there is, in essence, a 'match' between the grammatical functions assigned to the syntactic structure and the grammatical functions associated with the lexical item's arguments.

Selkirk 1981, 255, cited in Botha, p.60

(Selkirk operates with the theory of LFG, in which it is grammatical functions rather than semantic roles that are primary in grammatical description)

In the configuration  $[ ]_{\{V,P\}} [ ]_{\alpha}$  or  $[ ]_{\alpha} [ ]_{\{V,P\}}$ , where  $\alpha$  ranges over all categories,  $\{V,P\}$  must be able to link all internal arguments.

Lieber 1983, 258 (symbols V & P originally vertically aligned)

Lieber (1983) places particular reliance on the principle of argument linking in accounting for acceptability of compounds in English, arguing that this 'independently needed principle of syntax suffices to account for the range of possible compounds' (ibid, 251), with no need for lexical transformations (as Roeper and Siegel 1978) or other word-formation rules (presumably including those proposed by Selkirk 1982, although they are not specifically mentioned in the work cited).

For example, in Lieber's analysis of the compound *drawbridge*, the verb *draw* has its Patient argument specification satisfied by (or, its Patient argument links with) *bridge*. The Agent argument of *draw* is not linked in Lieber's theory because it is the verb's external argument--more or less equivalent to deep-structure Subject (but not exactly, see Williams 1981); and external arguments may not be linked inside compounds. Thus *draw* links all internal arguments (in this case, its Patient) and so is in accord with the principle quoted above. Furthermore the fact that *draw*, the nonhead, links its argument within the compound, is in accord with an additional principle posited by Lieber (258-9) which states that nonheads link their internal arguments inside the word. Heads, in contrast, link their arguments outside the word (but with certain exceptions: 258, 263). (In the case of *drawbridge*, the head is *bridge*, which is not an argument-taker).

The motivation behind this type of analysis is clear: there is a non-accidental connection between the meaning of compound words including a verb (or deverbal form) and phrases in which the verb takes the nominal constituent as a syntactic argument:

<i>drawbridge</i>	<i>draw a bridge</i>
<i>twist-drill</i>	<i>twist the drill</i>
<i>rowboat</i>	<i>row a boat</i>
<i>cheese-eater</i>	<i>eat cheese</i>
<i>flute-playing</i>	<i>play the flute</i>

Conversely, unacceptable compound words correspond to unacceptable phrases:

<i>*draw-river</i>	<i>*draw a river</i>
<i>*speed-eater</i>	<i>*eat speed</i>
<i>*book-playing</i>	<i>*play a book</i>

In a transformationalist account, this intuition would be captured by having a transformation produce the word from the phrase: the syntactic relation between the verb and the nominal is taken as basic, and the word-level modifier-head relation is derived from it. The lexicalist account generalizes the argument-predicate relation so that it may hold equally within morphological and syntactic structure, neither being derived from the other.

## 2) External vs. internal linking (Predication versus Modification, 2).

Let us return to the principle which states that heads satisfy their argument structure outside the compound word, while nonheads satisfy their argument structure inside the word (here and in the following discussion I refer to argument-specifying morphemes only so 'head' is short for 'argument-specifying morpheme as head', and so on). As Lieber states

it, this distinction simply follows from the fact that a head 'pass[es] its argument structure on to the compound' (p.258); thus it is actually the compound as a whole that satisfies its argument structure, necessarily doing so external to itself. A nonhead, however, 'does not pass any of its features on to the compound as a whole' (259), and so must link internally. But this leaves out the situation provided for by the second percolation convention, in which features of a nonhead do percolate if the head is unspecified for those features. Lieber assumes that the features comprising an argument structure are an opaque unit, which either percolates or does not, but cannot acquire additional features. Suppose however, that argument structures are not opaque in this way, but can be affected by the second percolation convention: e.g. if a head specifies Agent and a nonhead specifies Agent and Patient, the Patient may be added to the argument structure of the compound. If this is possible, then the nonhead arguments that percolate will indeed be satisfied outside the compound. I will argue in 5.3.3.2 that this is in fact the case with V-V's.

### 5.3.3. A Lexicalist Treatment of the V-V's.

In this section I will outline a way to account for the syntax and semantics of the V-V's by treating them as compounds. They are considered to be produced by a single lexical/morphological rule whose output is subject to several well-formedness conditions. After consideration of the simple V-V's, I will make a few suggestions for extending the analysis to complex V-V's.

In this discussion it will be most useful to distinguish three general types of V-V, selected from the terms put forward in 5.1.2.2, namely Causative, Sequential, and Modifying. I will take Descriptives to be representative of the Modifying type; at a later point I will mention some possible difficulties in including the Modals.

What sorts of facts should be accounted for in this analysis? Let us take as an example a sentence with Resultative V-V: *Klēmē thú sū dīpə* 'Klemeh wiped the pot dry'.

Several things can be said about it:

- a) the V-V as a whole has the argument structure [Agent Patient].
- b) the V-V also takes the grammatical relations Subject and Object.
- c) the semantic relation between the two verbs in the V-V (or rather, between the events denoted by the verbs) is that of causation.
- d) the Object *dīpə* has the Patient role associated with both verbs, being the thing that is wiped and the thing that becomes dry; the Subject *Klēmē* is the wiper, the Agent associated with the first verb only.



The problem of predicting the fact in a) can be called that of the *derived argument-structure*. b) states the V-V's *syntactic valence*. The question of how to state the information in c) is the problem of *interverbal semantics*. The facts in d) are those that I have been representing by lines of mapping connecting the constituent argument structures with the derived argument structure; the principles dealing with them may therefore be called those of *mapping*. Finally, there is also the problem of predicting possible and impossible combinations of verbs; since this can be thought of in terms of possible inputs to the compounding rule it can be called the question of *input constraints*.

#### 5.3.3.1. The Proposed Analysis.

(1) Input Constraints. To account for the possible binary combinations of verb types in simple V-V's, I will use a simple classification of verbs in terms of a mixture of syntactic valence and argument types:

$V_p$	verb specifying Patient only
$V_a$	verb specifying Agent only
$V_t$	verb specifying any two roles that are realized as Subject and Object

The classification of intransitive verbs into  $V_p$  and  $V_a$  is familiar from the Relational Grammar literature, where  $V_p$  is known as *unaccusative* and  $V_a$  as *unergative* (cf.

Perlmutter 1978). It is also convenient to refer to a broader class  $V_{a/t}$  including both the second and third type. The possible and impossible combinations of these classes can be charted as below:

Chart 1

		second verb	
		$V_p$	$V_{a/t}$
first verb	$V_p$	+	-
	$V_a$	+	+
	$V_t$	+	+

In other words, all combinations are possible except  $V_p - V_a$  and  $V_p - V_t$ : a  $V_p$  can only be followed by another  $V_p$ . These possibilities can be covered by a compounding rule  $V+V \rightarrow V$  in conjunction with a well-formedness condition ruling out concatenations  $V_p - V_a$  and  $V_p - V_t$ .

I will assume for now that  $V_d$ 's behave in the same way as  $V_t$ 's for purposes of input constraints.

In most cases, a given string that is well-formed according to the condition just stated may embody more than one type of V-V. For example, a string  $V_t - V_p$  could be either a Resultative, as *ʔa ʔe lɔ̄ aŋ* 'he ate all the rice' (*ʔe* 'eat' + *lɔ̄* 'be used up'), or it

could be a Descriptive, as *ʔa ʔe phrē dī* 'he ate (rice) quickly' (*ʔe* 'eat' + *phrē* 'fast').

Even a single  $V_t$ - $V_p$  may be ambiguous, as in the case discussed above (4.2.4), *me rē*

'improve stg' (Resultative) ~ 'be good to sbdy' (Descriptive). Now a fundamental

difference between a Resultative and a Descriptive is that a mapping analysis applies to

the former but not the latter (cf. 4.2.4). This fact can be represented by saying that a

compound made up of  $V_t$ - $V_p$  is specified as [ $\pm$ mapping], abbreviated [ $\pm$ mp]:  $V_t$ - $V_p$  [+mp] is

a Resultative,  $V_t$ - $V_p$  [-mp] is a Descriptive.

Again, there is no need for one rule producing [+mp] V-V's and a second rule producing [-mp]; the value of the feature can be freely assigned to the output of the compounding rule. Although this free assignment is subject to certain constraints, to be stated momentarily, such an analysis still predicts a great deal of ambiguity. Ambiguous examples do exist, as *me rē* (above), but the amount of actual ambiguity is probably less than would be predicted. This is to be attributed to pragmatic considerations; for instance the V-V in *ʔa ʔe phrē dī* 'he ate (rice) fast' will be generated both as [+mp], a Resultative, and [-mp], a Descriptive. The fact that it seems impossible as a Resultative (the translation would be something like 'he ate himself fast' or 'he ate rice and [as a result] it became fast') is because it is difficult to imagine any real-world scenario in which an action of rice-eating could result in either the rice or the eater becoming fast. However in this analysis we have to say that the Resultative interpretation is

nevertheless made available by the grammar<sup>10</sup>.

Note that we have not yet looked into the principle determining what is mapped to what; this will be discussed below. Assume for now a list of instructions based on the types of the constituent verbs; e.g. for  $V_t$ - $V_p$  map the Patient of  $V_1$  with the Patient of  $V_2$ , for  $V_a$ - $V_t$  map the two Agents together, and so on. For many of the possible combinations of verb type the feature [mapping] can determine the V-V type:

Chart 2

Csv=Causative, Dsc=Descriptive, Seq=Sequential

		$V_2$	
		$V_p$	$V_{a/t}$
$V_1$	$V_p$	[+mp]=Csv [-mp]=Dsc	*does not occur
	$V_a$	[+mp]=Csv [-mp]=Dsc	[+mp]=Seq (no contrast with [-mp])
	$V_t$	[+mp]=Csv [-mp]=Dsc	[+mp]=Csv or Seq (no contrast with [-mp])

Note that some combinations allow [+mp] only, cutting down on possible ambiguities; as before, this can be taken care of by a condition stating that  $*V_{a/t}$ - $V_{a/t}$  [-mp] is ill-formed. Also, wherever [-mp] is allowed it simply entails the Descriptive type, while

[+mp] varies between Resultative and Sequential. In addition, further discussion is required for the [+mp] V-V's in the bottom row. Finally, this chart simply presents the interaction of verb types, mapping, and V-V types as an aggregation of possibly-accidental facts. Ultimately, one would wish for a more principled explanation of why just these interactions and no others are found, although I will have little to say in that vein in the present work.

At this point, a shift of theoretical and terminological gear will be made.

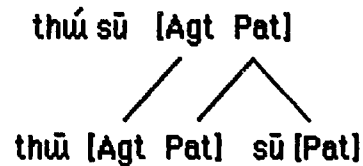
(2) Mapping as control. As proposed in 5.1.3.1 above, I am using 'pivot' to designate the participant associated with overlapping roles in derived argument structure (e.g. the Obj-x of *thwí sū* 'wipe dry', which has the Patient role assigned to both verbs). A glance over the derived argument structures proposed in the previous chapter shows that while the pivot participant can be mapped to various arguments in the argument structure of  $V_1$ , in all cases it is mapped to the Subject of  $V_2$ , taking 'Subject' to mean 'the argument that would appear as Subject in a sentence with the verb in question as simple main verb'.  
Supposing a compound  $V_1$  [Agt Pat] +  $V_2$  [Agt Pat], there are no examples where the Agent of  $V_1$  and the Patient of  $V_2$  are mapped to the same role in derived argument structure (the meaning would be something like *éKim hit Leslie with the result that Chris smacked Kim*), or where the Patients overlap but the Agents do not (meaning like *éKim hit, and Pat poked, Chris*).

This fact, that the mapped argument of  $V_2$  must be its Subject (in the sense described), constitutes a strong resemblance between what I have been calling mapping to the pivot and the concept generally known as 'control' (see above, 5.1.3.2). In particular, control is by definition concerned with the Subject argument of complement verbs: the task of a theory of control is to predict which constituent of a matrix sentence is coindexed with (=controls) the empty subject PRO of the complement clause (in GB theory) or is otherwise identified with/as the complement subject (in LFG, also Culicover and Wilkins 1986); but what is 'controlled' is always the subject of the complement clause, never any other constituent or relation. This resemblance is unlikely to be accidental; consider, for example, that the English translations of Kayah V-V's are often control structures.

I will therefore replace the term 'mapping' with 'control', and the feature [ $\pm$ mapping] with [ $\pm$ control], abbreviated [ $\pm$ cn]. This is admittedly an extended use of the term. Control has hitherto been considered a relation holding in syntactic structure, while I am proposing to use the same term to refer to a relationship holding between argument specifications in different argument structures; i.e. in lexical structure. Perhaps this may be taken as a generalization of a notion originally conceived to apply only in syntax, somewhat in the same vein as the generalization described in 5.3.2.3, in which argument linking is allowed to apply within morphological as well as syntactic structure.

I will also replace the multilevel derived argument structure diagrams with

single-line displays in which the mapping relationships are shown by coindexing. For example the argument structure of *thú sū* 'wipe dry', formerly



will be rendered as

$$\text{thú [Agt Pat}_i\text{]} \quad \text{sū [Pat}_i\text{]}$$

I will also show the syntactic realization of the argument structure by means of dotted vertical lines connecting arguments with grammatical-relation specifications, e.g

$$\begin{array}{cc} \text{Sbj} & \text{Obj} \\ | & | \\ \text{thú [Agt Pat}_i\text{]} & \text{sū [Pat}_i\text{]} \end{array}$$

This notation is basically that of Carrier-Duncan (1985, see below), the major difference being that her argument structures may include abstract predicates such as *CAUSE*. There are also some differences in terminology. Carrier-Duncan uses 'binding' for the coindexing relation that I am calling control, and 'linking' for the process of pairing semantic roles with grammatical-relation specifications.

(3) Derived argument structure and syntactic valence. Since the concept of control refers by definition to the grammatical relation Subject, adopting the term implies also a shift in emphasis from semantic role structure to grammatical relations. In fact it will

not be necessary to radically alter the semantic–role-based theory of argument structure set out the Chapter 3, but it will be useful to be more explicit about the link between semantic roles and grammatical relations. In this section I will outline a way of accounting for both the argument structure and the syntactic valence of V-V's. I will follow the approach put forth most recently in Carrier-Duncan 1985, but in essence dating back at least to Fillmore 1968. The basic idea is that both semantic roles and grammatical relations are arranged on hierarchies, and that the realization of semantic roles in a given verb's argument structure is usually determined by pairing the argument that is highest on the semantic–role hierarchy with the grammatical relation that is highest on the grammatical relation hierarchy (usually Subject), and then proceeding downwards on both hierarchies (for a related concept see Foley & Van Valin's 'Actor/Undergoer hierarchy').

For Kayah, the two hierarchies would be approximately the following:

<u>Sem. roles</u>	<u>Gram. relations</u>
Agent	Subject
Patient	Obj-2
Recipient	Obj-1
Locative (Goal/Source)	Oblique (PP)

'Obj-x' now becomes a neutral designation for a single unmarked postverbal NP.



Exceptions to the hierarchical realization are dealt with by specifying the realization in the verb's lexical entry ('pre-linking' in Carrier-Duncan's terms), but in general the hierarchies serve to free the argument structure from the need to determine syntactic realization.

To illustrate, we may take the verb *æ* 'eat', which specifies Agent and Patient (with no particular ordering relation between the two). The two roles are first ranked on the semantic role hierarchy, then hooked up with the correspondingly-ranked grammatical relations: Agent to Subject, Patient to Object. This may seem trivial, but it proves useful in dealing with compound argument structures: for instance, one is then able to refer to 'highest argument on the semantic-role hierarchy' rather than Subject or Agent. The main point is that the verb's lexical entry needs to list only the specified semantic roles, not their grammatical realizations--at least in the unmarked case: particular verbs may specify the grammatical realizations for certain arguments. It might seem that given this apparatus the lexical entry will not need to include a direct specification of syntactic valence: e.g. any two semantic roles, with no idiosyncratic grammatical specifications, automatically come out as Subject and Object. However, there is one aspect of Kayah syntax for which it is advantageous to assign syntactic-valence features to V-V's independent of their argument structures (see the discussion of derived  $V_d$ 's with Bound Directionals in 6.4 below) so we may as well allow reference to such features<sup>11</sup> in other cases as well. Therefore the full lexical entry for *æ* would be:

ʔe 'eat' [Agent Patient]

[Sbj Obj-2]

Let us now look at how the apparatus sketched so far might deal with the V-V *cwá* *ʔichi* 'go and split (e.g. firewood)'. I will assume for the moment that *cwá* 'go' specifies Agent only, although in reality it probably also specifies Locative (cf. the discussion in 3.5 on the difficulty of distinguishing specified from unspecified Locative with Sequentials).

1. *cwá* is  $V_a$  and *ʔichi* is  $V_t$ , a well-formed combination.

2.  $V_a$ - $V_t$  must be [+cn] according to Chart 2; the derived argument structure will then

be:

*cwá* [Agent<sub>i</sub>] *ʔichi* [Agent<sub>i</sub> Patient]

Since control by definition coindexes the highest-ranking argument of  $V_2$ , and since  $V_1$  has only one argument, the choice of coindexed arguments is automatic.

3. Coindexed arguments are realized as a single grammatical relation (this is equivalent to the converging mapping lines in previous derived argument structure diagrams). In hierarchical ranking the coindexed pair of Agents must outrank the Patient, so the grammatical realization may be added to the argument structure as follows:

Subject	Object
cwá [Agent <sub>i</sub> ]	ʔíchi [Agent <sub>i</sub> Patient]

Here the tag 'Subject' is attached to the Agent argument of the first verb rather than the second. This might appear arbitrary, but consider what happens when the coindexed arguments are of different types, as in the Directive *vĕ nĕ ʔíchi Phān khru* 'I told Pha'a to split firewood'. The argument structure is:

Subject	Obj-1	Obj-2
nĕ [Agent Recipient <sub>i</sub> ]	ʔíchi [Agent <sub>i</sub> Patient]	

Here the coindexed pair (realized as the Obj-1 *Phān* in the example) must count as a Recipient, not an Agent, since it is outranked by both the (other, uncoindexed) Agent and the Patient. This is best explained in terms of headedness:  $V_1$  would be defined as the head of the compound verb, so its arguments take precedence in determining the type of a coindexed pair.

4. Chart 2 also states that a combination  $V_a-V_t$  must be a Sequential, which is to say that the relation between the two events denoted by the verbs is of the type 'and-then, in-order-to, and-alternately...' (etc., see 4.2.2). This information may be represented in a rule of interpretation, in a decompositional semantics using an abstract predicate, or possibly in some other way; I will consider 'Sequential', 'Causative' and 'Modifying' to be abbreviations for whichever of these tactics is ultimately selected. The point being

combination  $V_a-V_t$  and the feature [+cn] (although in this instance the two are redundant:

there is no  $V_a-V_t$  [-cn]).

We thus have (the beginnings of) an account of derived argument structure (2), syntactic valence (3), and inter-verbal semantics (4). The principles of input constraints and mapping (now control) are stipulated in Chart 2. Note that I have in effect reversed the order in which these were presented in Chapter 4, where the inter-verbal semantic types (Resultative, Sequential etc.) were taken as starting points to which were added the data on derived argument structure, input constraints etc. We are now starting with input constraints and mapping, and from them deducing derived argument structure and inter-verbal semantics.

We now turn to the more complex case in which  $V_1=V_t$ . Apart from the possibility of [-cn], available only when  $V_2=V_p$ , at least five possibilities exist:

- (17) Phāλ thú sū dīpɔ  
 P. wipe dry pot  
 Pha'a wiped the pot dry.
- (18) Phāλ ʔichi síphrá khru  
 P. split tired firewood  
 Pha'a chopped firewood till he was tired; Pha'a got tired chopping firewood.
- (19) vē plí cwi pù  
 1s whip pull ox  
 I whipped the ox to make it pull.
- (20) vē nō cwi Phāλ súpλ  
 1s command pull P. rope  
 I told Pha'a to pull the rope.
- (21) vē chijá plwā thé  
 1s untie release pig  
 I untied [it] and released the pig.

(17-19) are Resultatives, (20) is a Directive and (21) is a Sequential. In the terms of Chart 2, (17-18) are  $V_t-V_p$  [+cn], and (19-21) are  $V_t-V_t$  [+cn]. Some additional means is definitely needed to distinguish (19) (20) and (21) from each other, and possibly also (17) from (18) as well. To begin with (20) and (21), there is a difference in control: the argument of  $V_1$  associated with the Agent of  $V_2$  is the Recipient/Causee in (20), but the Agent in (21). This would be represented in derived argument structure as follows:

Arg. Struc for (20) nō [Agent Recipient<sub>i</sub>] cwi [Agent<sub>i</sub> Patient]

Arg. Struc. for (21)    *chijá* [Agent<sub>i</sub> Patient] *plwā* [Agent<sub>i</sub> Patient]

Evidently [+cn] is not enough; what is needed is a further specification of [+cn] as control by Subject [Scn] or control by Object [Ocn]. More accurately, since the hierarchical pairing of arguments with grammatical relations need not yet have applied, we may refer to the semantic-role hierarchy and render 'Subject control' as 'control by highest-ranking argument of  $V_1$ ', and 'Object' control as 'control by second-ranking argument of  $V_1$ '. This way of defining control is needed in any case for Sequentials: although many Sequentials do indeed map Agent to Agent, some do not, as in the following (also appears as example 44, 4.2.2):

(22) *thwi mé thá ?e á li phá ná*  
       dog    see    eat NS book skin NA

The dog saw the hide book and ate it. (103.4)

*mé thá* 'see' does not specify an Agent. There is not complete agreement on how to classify the perceiver argument of perception verbs; labels that have been used include Experiencer (Fillmore 1971) and Locative (Foley & Van Valin 1984). Whatever label is chosen for the perceiver argument, it is clearly higher on the hierarchy than the argument denoting the thing perceived, at least in Kayah, English, and many other languages in which the perceiver is Subject and the perceived is Object. Therefore the mapping in the derived argument structure of Sequentials is best described with reference not to semantic roles, but to relative position on the hierarchy: e.g. 'coindex the

highest-ranking role of each argument structure'.

These two control types will also serve to distinguish (17) and (18): since in (18)  $V_2$  expresses a result applying to the Agent of  $V_1$ , (18) is [Scn] while (17) (and (19)) are [Ocn]. The bottom row of Chart 2 may now be revised:

$V_t-V_p$ [Ocn] = Res (ex.17)	$V_t-V_t$ [Ocn] = Res (ex.19) or Drv (ex.20)
[Scn] = Res (ex.18)	[Scn] = Seq (ex.21)
[-cn] = Dsc	

Finally, there is the difference between the Resultative (19) and the Directive (20). This is not a matter of control type, since in both it is the Object (in the sense just described) of  $V_1$  that is associated with the Subject of  $V_2$ : *pu'ox* in (19) and *Phān* in (20). The difference can be described as one of syntactic valence: the V-V of (19) takes only Subject and Obj-x, while that of (20) takes Subject, Obj-1 and Obj-2 (cf. also the description in 4.2.3, where the difference is described in terms of differing semantic roles).

Consider the two derived argument structures, with syntactic realizations included:

Arg. Struct. for (19)

Sbj	Obj-x	
pí	[Agent Patient <sub>i</sub> ]	cwi [Agent <sub>i</sub> Patient]

Arg. Struct. for (20)

Sbj	Obj-1	Obj-2
b) n̄	[Agent Recipient <sub>i</sub> ]	cwi [Agent <sub>i</sub> Patient]

The structure for (20) ranks the arguments as Agent > Patient > Recipient<sub>i</sub>-Agent<sub>i</sub>.

This illustrates the principles stated above: a) that a coindexed set of arguments counts as a single argument in syntactic realization; b) that coindexed arguments of nonidentical type count as an instance of the type of the argument of V<sub>1</sub> for purposes of ranking<sup>12</sup>.

Why is the Patient argument of V<sub>2</sub> in a) not realized? First, to allow syntactic realization of two Patients would violate the principle of at most one instance per semantic role type per clause (3.2.1). Secondly, we can also say that the V-V *pí cwi* has the syntactic valence [Sbj Obj-x] and so has no syntactic position available for a third argument. Directive V-V's like *n̄ cwi*, in contrast, have the valence [Sbj Obj-1 Obj-2] and so allow both the Recipient<sub>i</sub>-Agent<sub>i</sub> and the patient of V<sub>2</sub> to be realized.

It is probably best to say that the syntactic valence of a V-V is inherited from that of the head member; in this case, the Directive verb *n̄* is plausibly subcategorized as V<sub>D</sub>



(i.e. [Sbj Obj-1 Obj-2])<sup>13</sup>. Several of the directive verbs are in fact identical with or related to  $V_d$ 's: *dá* 'let', homophonous with 'give', and *āśwā* 'teach'.

The proposal thus far may be summarized in the following revision of Chart 2:

	$V_p$	$V_{a/t}$
$V_p$	[+cn]=Csv [-cn]=Dsc	*does not occur
$V_a$	[+cn]=Csv [-cn]=Dsc	[+cn]=Seq (no contrast with [-cn])
$V_t$	[+cn] { [Scn]=Csv [Ocn]=Csv [-cn]=Dsc	[+cn] { [Scn]=Seq [Ocn]=Csv (no contrast with [-cn])

The only differences between Resultatives and Directives, in this view, are due to the lexical features of the Directive verbs, particularly in the subcategorization of Directive verbs as  $V_d$  which allows the Causee argument to appear as Obj-1.

Note that '[Scn] Csv' stands for the relatively rare type of Resultative represented by the following (repeated from 4.2.1.1):

- (23) *vē āichi śiphrā khru*  
 Is split tired firewood  
 I got tired splitting firewood.

- (24)  $v\bar{e}$   $m\bar{e}$   $mo$   $ne$   $to$   
 Is look happy you NEG  
 I feel sorry for you; I pity you;  
 (literally,) I am unhappy seeing you[r condition]. (2/27)
- (25)  $\text{ʔa}$   $\text{ʔ}\bar{o}$   $m\omega$   $th\bar{\lambda}'\acute{i}phr\bar{e}$   
 3 drink drunk whiskey  
 He got drunk on whiskey. (common expression)

It is worth noting that there is something like complementary distribution of [Scn] and [Ocn]. With  $V_t-V_p$  [Ocn] is the norm and [Scn] is rare; with  $V_t-V_{a/t}$  [Scn] is common, while [Ocn] is either associated with a special class of verbs, the Directives, or is of the rare type represented by  $\text{p}\acute{i}\acute{i}$   $\text{cwi}$  'whip stg so it pulls (stg)'.

### 5.3.3.2 Remarks on Headedness and Argument Linking.

(1) Headedness and Percolation. Headedness does not seem to play a very significant role in the V-V's, which is also to say that it is difficult to identify the head constituent with great certainty.

The Descriptives are the exception: since  $V_1$  conclusively determines the argument structure and syntactic valence of the V-V, and since the argument structure of  $V_2$  is not syntactically realized at all, it is obvious that  $V_1$  is the head. This also fits with the generalization that verbal modifiers follow their heads, while nominal modifiers precede

Causatives (Resultatives and Directives) are also best analyzed as head-initial. The evidence for this is two sorts of prominence that are given to the first verb. First is the

fact that while the predicate expressed by  $V_1$  is asserted, that expressed by  $V_2$  is only implied (4.2.1). Secondly, there is the ranking of a coindexed pair of arguments according to the argument type of the argument of  $V_1$  rather than that of  $V_2$ .

In Sequentials it would be possible to claim that the second verb is the head, which would explain the fact that, if the two verbs have conceptually different Patients, the Patient of  $V_1$  cannot be realized but that of  $V_2$  can (4.2.2); i.e. a Sequential cannot have an Object that is not an argument of the second verb.

A head determines the features of its compound by the percolation conventions quoted above (5.3.2.1). Since argument specifications are features, it seems that they would percolate. They certainly do in the case of the Descriptives, although trivially, the number and type of arguments being completely determined by the head, as was just stated. Consider next the Directive. The head verb is specified for Agent and Recipient(Causee). The second verb must include an Agent specification. If it includes nothing further, the Agent is mapped to the head's Causee; the compound then takes Agent and Causee, and thus has an argument structure identical with that of the head. If the second verb also takes a Patient, that argument must be added to the derived argument structure:

$$V_1 [\text{Agt Rec}_i] \quad V_2 [\text{Agt}_i \text{ Pat}]$$

This is reminiscent of the second percolation convention, which states that features

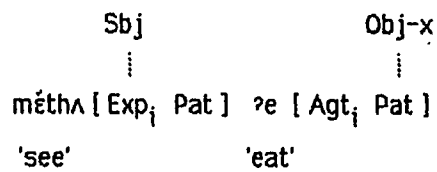
unspecified by the head may be contributed by nonheads: if the nonhead specifies a role different from those specified by the head, that specification percolates and is added to the derived argument structure. As was mentioned in 5.3.2.1 above, it is sometimes assumed that semantic role specifications are not the type of feature that can be 'filled in' by the nonhead in this manner, but the facts just cited argue otherwise. The same point is made by Marantz:

Semantic role assigning properties, like argument structures but unlike grammatical features such as [ $\pm$  log sub], should not be seen as something for which a lexical item is either specified or not, such that if the item is specified for the features, it cannot inherit them...a constituent with semantic role assigning properties may inherit other semantic role assigning properties as long as this inheritance does not violate independent constraints, either language specific or universal.

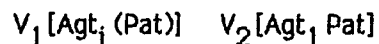
Marantz 1984, 242

This will also apply to the V<sub>Ptc</sub>'s that add arguments, such as *pe* 'Benefactive/Malefactive'. However, feature percolation, like headedness (of which it is a consequence) is not of great salience in the analysis of the V-V's. As an example, the Sequential *thwi méthm ʔe á le phá na* 'the dog saw and ate the hide book', cited above, cannot be fully accounted for by percolation of semantic roles only. If the analysis is limited to role percolation, the V-V *méthm ʔe* 'see and eat' would end up with the Experiencer<sup>14</sup> of *méthm* as well as the Agent and Patient of *ʔe*, and the example sentence would be expected to include three participants, not two. The reason this does not happen, of course, is that the Experiencer and Agent, as the highest-ranking arguments of their respective verbs, are coindexed by control and so count as a single argument for

syntactic realization. The derived argument structure would be as follows, using 'Exp' (Experiencer) for the higher-ranked argument of *métha*:



In fact this type of representation of derived argument structure implies that *all* semantic role specifications percolate, of head and nonhead alike, in that all are represented: in the preceding example if  $V_1$  were  $V_{a/t}$  the derived argument structure would include two Agents:



The failure to differentiate head from nonhead must mean either that this is not true percolation in the sense defined by Lieber and Selkirk, or that (some) V-V's are not headed (e.g. their constituents might be coordinate).

I am inclined to say that V-V's are headed, that all semantic role specifications of all constituents are simply brought into the argument structure, and that what percolates with sensitivity to the head/nonhead distinction is not semantic role specifications but syntactic valence features. Syntactic valence features do seem to percolate in the 'additive' fashion described in the quote from Marantz. Consider again evidence from the Sequentials: a Sequential's valence number is that of the highest member: if the head has the highest valence, no more need be said. If the non-head has the highest valence, it

will add to the compound's features those of its syntactic specifications that do not have equivalents in the head's features. It does not matter for these purposes which constituent is analyzed as head: both  $V_i-V_t$  and  $V_t-V_i$  come out as  $V_t$ 's. Therefore it must be that if the head specifies Subject only and the nonhead specifies Subject and Object, the nonhead's Object specification is added to the V-V's feature matrix.

The view of the lexical structure of the V-V's can now be summarized. V-V's are like simple verbs in specifying at least one and at most three grammatical relations, and in requiring that each grammatical relation realize a different semantic role. They differ in that the argument structure of V-V's may contain multiple instances of a single semantic role type (e.g. multiple Patients in Resultatives, multiple Agents in Sequentials), and may contain semantic roles that are not syntactically realized (e.g. the lower-ranking argument of  $V_i$  in Sequentials where both verbs are  $V_t$ ): this is the level at which the V-V's are complex. It is not clear whether we want to say that V-V's resemble simple verbs in assigning no more than one semantic role to each grammatical relation; e.g. perhaps the Obj-1 of a Directive V-V is simultaneously a Recipient(Causee) and an Agent<sup>15</sup>.

It is probably better to take the conservative view, that one grammatical relation can have at most one semantic role. In this view the Obj-1 of a Directive V-V is simply a Recipient/Causee, although the semantics has available the information that in the

argument structure the Recipient is coindexed with an Agent. That is, the syntax is simple, and complexity is confined to lexical structure.

(2) Argument linking. The Kayah V-V's provide the empirical evidence in favor of the extension of the theory of argument linking suggested in 5.2.3 above. The two types of argument linking described by Lieber for English compounds are external linking of the head's arguments, and internal linking of head's and nonhead's arguments. Neither can apply to the nonhead constituent of Causative and Sequential V-V's. For example, in the Resultative *thwí sū* 'wipe + be-dry → wipe stg dry' the Patient argument of *sū* does not link to *thwí*. If it links to anything it is to the NP Object, outside the compound; since *thwí sū NP* is grammatical, we must either allow nonheads to link arguments externally, or else abandon the principle that all argument specifications must be satisfied.

The clearest evidence is from arguments of nonheads that are realized syntactically and are not coindexed. In this category are the lower-ranking arguments of second verbs in Directive and Sequential V-V's; e.g.:

(26)	vē dǎ phjá PhāΛ ʔĩthoə ~	Sbj	Obj-1	Obj-2
	I s let take P. knife			
	I let Pha'a take the knife.	dǎ [Agt Rec <sub>i</sub> ]	phjá [Agt <sub>i</sub> Pat]	

(27)	vē cwá phjá ʔĩthoə	Sbj	Obj-x
	I s go take knife		
	I went and took the knife.	cwá [Agt <sub>i</sub> ]	phjá [Agt <sub>i</sub> Pat]

In both of these there is no doubt that the complex verb takes a grammatical relation, Obj-x or Obj-2, which has the semantic role Patient and realizes the Patient argument of

the nonhead constituent<sup>16</sup> of the complex verb.

The case of nonhead arguments that are coindexed, as the Agent of *phjá* in both of the above examples, is less certain. There are at least three possibilities:

1. The Agent argument of *phjá* (i.e. the contrulee, with the coindexed head argument being the controller) is not linked, at least in the sense of 'linking' as syntactic realization. This follows from the conservative view of lexical structure described in the previous section.

2. The Agent specification of *phjá* is satisfied by the coindexing with an argument of  $V_1$ . This would involve a change in the definition of 'linking', since the link is not to a syntactic constituent but to another argument specification.

3. The Agent of *phjá* is actually linked to the Object NP, with the controller argument simply mediating the linking.

### 5.3.3.3. Afterword.

(1) Descriptives as modification. The one V-V type that might be described in terms of internal linking by the nonhead is the Descriptive, if its meaning is seen as applying the meaning of the second verb to the meaning of the first verb. The second verb then would link its single argument (Theme/Patient) within the compound. E.g. *æ phrē* 'eat quickly' could be interpreted as attributing quickness to the action of eating, hence the Patient argument of *phrē* 'fast' would be linked to *æ* 'eat'. In other words, the second verb of a Descriptive V-V is modifying rather than predicating; therefore it links its



argument compound-internally, to the lexical-category verb that is its head.

Note that arguments in the usual sense are typically referring expressions; it may be that a verb can properly be called an argument only when, as in the Descriptive V-V's, the (other) argument-taking morpheme is modifying it. Arguments of predicating morphemes may still be required to be referring expressions (i.e. NP's, or major categories in general).

(2) Modals. The Modal V-V's can be analyzed as the mirror image of Descriptives:  $V_2$  is the head, and there is no control, i.e. no argument of  $V_1$  is realized.  $V_1$  is a modifier, probably to be treated by whatever semantic analysis turns out to apply to  $V_2$  of Descriptives. I have no compelling evidence for this analysis. Treating the Modal verb as a modifier is at least consistent with the semantics of modality; for instance, English sentences with modals often have equivalent (synonyms or near-synonyms) with the modal meaning expressed in an adverb:

(28a) John must go to the doctor on Tuesdays.

(b) John obligatorily goes to the doctor on Tuesdays.

(29a) The final report may be postponed.

(b) The final report is optionally postponed.

(3) Complex V-V's. Here I will suggest a way in which the present analysis might be extended to account for complex V-V's. This last section represents research in progress, and is presented as an indication of a line of analysis which shows promise, but has not been fully developed at the time of writing.

Let us take as an example a string *nō cwá ʔíchi* 'command+go+split'. In terms of the verb types referred to in Chart 2, this is  $V_t-V_a-V_t$ . There are two possible ways of parsing this string, namely [<sub>Drv</sub> nō [<sub>Seq</sub> cwá ʔíchi ] ] and [<sub>Seq</sub> [<sub>Drv</sub> nō cwá] ʔíchi ]. In fact only the former is acceptable, meaning 'tell sbdy to go split', while the latter is not (the meaning would be 'x tell y to go and x split'); this was accounted for in 4.4.1 by stating that a Directive has scope over a following Sequential, while a Sequential cannot have a Directive as  $V_1$ .

These facts also fall out of the present analysis. First, since the third verb *ʔíchi* 'split' is a  $V_t$ , it cannot function as  $V_2$  of a Descriptive, and so the structure [[nō cwá] ʔíchi] must include a control relation between the argument structures of the first constituent *nō cwá* and the second constituent *ʔíchi*. Taking *nō cwá* as a compound  $V_t$  for the moment, two argument structures with control are possible:

ScD nō-cwá [Agt<sub>i</sub> Rec] ʔíchi [Agt<sub>i</sub> Pat]

OcD nō-cwá [Agt Rec<sub>i</sub>] ʔíchi [Agt<sub>i</sub> Pat]

So far there is nothing that would rule out either of these. Let us then consider the argument structure of *nō cwá*: being  $V_t-V_a$ , it also cannot be [-cn], and so must also include a control relation between its component argument structures. Since *nō* 'command' belongs to a closed class of verbs with a specialized meaning, it is plausible to attribute to it a feature requiring Ocn. Therefore there is only one possible argument structure for *nō cwá*:

$n\bar{o}$  [Agt Rec<sub>i</sub>] cwá [Agt<sub>i</sub>]

Consider now what would result if the relation between *nō cwá* and *ʔíchi* were Scr:

$n\bar{o}$  [Agt<sub>i</sub> Rec<sub>j</sub>] cwá [Agt<sub>j</sub>] ʔíchi [Agt<sub>i</sub> Pat]

This would give the unacceptable 'x tell y to go and x chops', and is probably to be ruled out because of the nested control relations: the arguments indexed *j* are 'inside' the arguments indexed *i*. In contrast, note the acceptable *cwá nō ʔíchi* 'go and tell sbdy to split', whose argument structure would be

cwá [Agt<sub>i</sub>]  $n\bar{o}$  [Agt<sub>i</sub> Rec<sub>j</sub>] ʔíchi [Agt<sub>j</sub> Pat]

This can be allowed because there is no nesting of the two control relations. Note that the account just given seems not to need to refer to differences in constituent structure, unlike the description in terms of c-command relations that was given in 4.4.1.

Finally, it seems that we must also prevent a control relation from 'passing over' an uncoindexed argument structure. This will prevent a Descriptive from occurring inside a Resultative; e.g. *\*thú phrē sū* 'wipe dry fast' would have the argument structure:

thú [Agt Pat<sub>i</sub>] phrē [Pat] sū [Pat<sub>i</sub>]

in which the control relation between the Patients of *thú* 'wipe' and *sū* 'dry' skips over the Patient argument of *phrē* 'fast'.

**Morphosyntax, Part II**

(other constructions, clausal and interclausal syntax)

## 6. The Clause.

### 6.1 Definition and Types.

In this chapter we will examine the structure of the clause, considering first the question of what configurational structure to assign to the linearly-ordered elements NP, VC, PP and SPtc (6.2), then turning to some discussion of clause constituents in terms of grammatical relations (6.3-5).

The clause is defined as any construction including a verb and terminable by *to*, the negative. The clause may also be delimited by the occurrence of the obviative third-person pronoun, which will be described in 6.3 below.

The linear order of elements in the clause can be represented as below:

(NP<sub>1</sub>) VC (NP<sub>2</sub>) (NP<sub>3</sub>) (PP<sub>1</sub>) (PP<sub>2</sub>) (ClP) (CPtc)

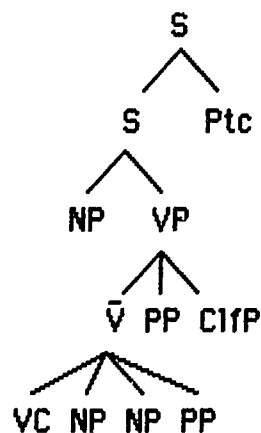
VC is Verb Complex, PP is Prepositional Phrase. ClP is Classifier Phrase, consisting of Quantifier (usually a numeral) and a Classifier. CPtc is Clause Particle, one of a small class of morphemes generally having to do with the realis-irrealis gradient; the most common member is the negative *to*.

## 6.2. Constituents

I assume that the Cptc is adjoined to the clause (see [1] below). It may follow VC, NP, PP or ClfP; it is thus also possible that Cptc is a constituent of VP or a sister of VP (and NP<sub>1</sub>). Semantically, however, it relates to the entire proposition, not to any smaller portion; e.g. *to* is a sentence-level negative.

6.2.1. The VP. The second break that can be made is between NP<sub>1</sub>, which bears the Subject relation, and what I will call the VP. This VP may have a further division into  $\bar{V}$  and  $\bar{V}$ -external elements. I suggest that the clause in Kayah is best represented as below

(1)



In order to support the recognition of VP as a constituent, I will briefly describe two constructions in Kayah that need to refer to a unit [VC + XP], i.e. some portion of a clause excluding both Subject NP and Cptc.

Source expressions.

There is a construction involving a sequence [V XP][V XP] within a single clause, where XP is NP or PP. The general meaning is 'V<sub>2</sub> from having V<sub>1</sub>-ed'. When V<sub>1</sub> is *ʔo* 'be at, have', it is the most usual way to express the notion Source:

(2) [ʔo dʒ so kū] [tā the nɔ́] to  
 be-at at tree inside fall go-out at-all NEG  
 It didn't fall out of a tree. (8.4)

The brackets enclose the two [V XP] constructions; the question of how to label these brackets will be addressed below. This example illustrates the monoclausal nature of the construction: the negative *ʔo* has scope over both verbs in the sequence. In fact the main point of this particular utterance is to negate the first verb: the speaker then goes on to say where 'it' did come out from.

This construction cannot include a ClfP in the first of the sequential units; in the formula given above, XP cannot be ClfP (the first unit [VC XP] also, of course, cannot include a clause particle, which would turn the construction into a two-clause sequence). This suggests that NP and PP, the categories that can occur in the first unit, may form a constituent with the VC. If we label that constituent  $\bar{V}$ , the 'source' construction can be described as a sequence of  $\bar{V}$ 's, and could be accounted for by some such phrase structure



rule as

$$(3) VP \rightarrow (\bar{V}) \bar{V} PP ClfP$$

or possibly

$$(4) VP \rightarrow \bar{V} PP ClfP$$

$$\bar{V} \rightarrow (\bar{V}_1) \bar{V}_2$$

It is not clear whether the 'Source'  $\bar{V}$  should be adjoined at the VP or  $\bar{V}$  level.

The label  $\bar{V}$  is consistent with common practice, in which  $\bar{V}$  is an intermediate-level category containing the verb and its subcategorized arguments: the PP that may occur in the first [V XP] unit is always the specified, 'inner' Locative, as in (2). The inadmissibility of ClfP in the same position is then explained by the fact that ClfP never embodies a specified argument.

#### The lè construction

The abstract Noun *lè* 'place for V-ing, thing for X' must be followed by a modifying expression containing a Verb; e.g. *lè mā* 'bedroom' (*mā* 'sleep'), *lè ʔi* 'toilet' (*ʔi* 'defecate'). This postposed modifier may include an NP Object, as in

*khru lè bó thé*                      machine + *lè* + weave + cloth → loom

However, it cannot include a Subject:

\**lè phre khū ʔiò thā*              *lè* + person + male + bathe + water →  
men's bathing place

(\**lè ʔiò thā* 'bathing place' is fine). *lè* may thus be subcategorized for a following VP

or  $\bar{V}$ ; unfortunately I lack data that could show whether the modifying expression could contain a ClfP (as explained above,  $\bar{V}$  should not be able to contain a ClfP). Still the postposed modifier of /e' is a second construction that must refer to some unit [VC + XP].

It should be noted that the evidence for the hierarchical structure given in (1) above is by no means as strong as it is in English and other SVO languages that have a major break between Subject NP and the remainder of the clause. The evidence of the source expressions and the /e'-construction does give support to the recognition of a constituent made up of the verb (VC) and its subcategorized non-Subject arguments, which I have labelled  $\bar{V}$ , but there remains the possibility that it is surrounded by a more 'flat' structure, as

$$[_S \text{ NP } [_{\bar{V}} \text{ VC NP NP PP}] \text{ PP ClfP } ]$$

in which case  $\bar{V}$  could equally well be called VP. The existence of a distinct VP, something that includes both the  $\bar{V}$  and the non-subcategorized PP and ClfP, must be considered a provisional assumption at this stage.

The matter of the further constituent analysis of the VP hinges on the grammatical relations holding between the Verb (or, perhaps, the VC) and the constituents following it. Accordingly we now turn to an examination of the grammatical relations in the Kayah clause.

As an anchor for discussion I repeat the summary of grammatical relations given above (3.5):

	order	configuration	category
Subject	1st before V	sister of VP	NP
Object-1	1st after V	sister of VC	NP
Object-2	2nd after V	"	NP
Oblique-1	3rd after V	sister of VC <sub>or</sub> of VP	PP
Oblique-2	4th after V	"	PP
Extent	5th after V	sister of VP	ClP

6.2.2 Selected and non-selected grammatical relations. Most theories distinguish selected from non-selected grammatical relations. This distinction closely parallels that between selected and non-selected semantic roles (3.2.2 above); in fact it may be taken as simply its reflection at the syntactic level, depending on one's theory.

Chomsky's 'Projection Principle' is one version of such a theoretical stance.

Selected grammatical relations subcategorize verbs; each verb's lexical entry marks which of these grammatical relations it occurs with. Non-selected grammatical relations occur freely with all or most verbs. Subject and Object are always selected; time, place, and other 'adverbial' expressions are not selected. Some grammatical relations are selected in some instances and non-selected in others, again paralleling the semantic roles I have referred to as 'variable' (3.2), such as Locative.

Selected grammatical relations are often said to appear obligatorily. The following would then be ungrammatical because of non-appearance of selected grammatical relations:

(5) \*Sam put the book.

(6) \*They discussed for an hour.

(5) lacks a selected locational expression (e.g. *on the table, away*), and (6) lacks an Object (*the problem, Kim's proposal, etc.*).

However, I prefer not to emphasize obligatory appearance of selected grammatical relations, for the reasons given in 3.2.2 above in connection with obligatoriness of semantic roles (i.e. that in Kayah a role may be both semantically obligatory and linguistically represented by zero).

A feature that is useful in examining the selected/nonselected status of Kayah grammatical relations is the fact that selected relations can realize a wide range of semantic roles, the role realized in a given instance having a direct dependence on the selecting verb. Non-selected grammatical relations, in contrast, have no such dependence or only a very indirect sort. The influence exerted by the verb on the semantic role realized by selected grammatical relations is most clearly seen in Subject and Object, as the following familiar types of example show:

- (7) Al broke the glass.      Sbj = Agt, Obj = Pat
- (8) The glass broke.      Sbj = Pat
- (9) Al likes cookies.      Sbj = (?)Experiencer, Obj = (?) Cause
- (10) The car seats five.      Sbj = Loc, Obj = ? (if *five* is an Obj)

The varying semantic roles realized by Subject and Object in such sentences are most easily stated as lexical specifications of the verbs. Non-selected grammatical relations, in contrast, are unvarying in their semantic roles; e.g. *when*-clauses are always time expressions. When they do vary, the different semantic values do not depend on the main verb of the sentence. For instance, English *for*-PP's may realize Benefactive (*for John*), temporal extent (*for three days*), and other notions (*for example*), but any of the three may occur indifferently with the same verb; e.g. *write a note for John*; *write letters for three days*; *write a play, for example*.

A range of examples similar to the above can be given for Kayah:

- (11) Phā ʌ    jó    kè      ʔimū  
 (name) bend broken-off stick

Pha'a broke the stick/bent the stick so it broke.

- (12) ʔimū    lé kè  
 stick    broken-off

The stick broke/is broken.

- (13) Phā ʌ    sí plɔ tō    khó mū  
 (name) like    cake, bread

P. likes cake.

- (14) phú cè tā hi  
 child fall house  
 The child fell off the house.

The first three sentences are exactly similar to the English sentences (7-10) in their grammatical-relation/semantic role pairing. The last sentence above has Subject as Patient and Object as Locative (Source). The variability of the semantic roles realized thus indicates that, as in English, Subject and Object are selected grammatical relations in Kayah.

In fact the 'Object' relation exemplified above is only one of two post-verbal NP positions in Kayah, as indicated by the terms Obj-1 and Obj-2 given above; the single postverbal NP in (11-14) is more like Obj-2 (see 6.4 for more discussion). Obj-1 realizes roles like causee, recipient, and beneficiary (which may be all grouped under Recipient; see 3.3); also standard of comparison and comitative. Again, the choice between these is controlled by the verb: causee with Directive V-V's, recipient with 'root' ditransitive verbs (see 6.4) like *dá* 'give' and *śíche* 'sell', and beneficiary with Verb Particles *pe* 'for' and *cwə* 'help', and so on.

Subject, Obj-1 and Obj-2 are thus selected relations; the remaining grammatical relations of Kayah are the ClFP-Extent expression and the two types of PP.

ClFP. The Kayah Classifier Phrase, or Extent expression, is clearly non-selected, always occurring freely. With respect to semantic range, although it is possible to distinguish semantic values like temporal extent, frequency, and number of

participants, these are determined by the type of classifier heading the ClFP: the co-occurring verb has no bearing on the choice. All three can occur with the same verb:

(15a) ʔa cwá dǝ hóhé sō nā      They went to school for three days.

(15b) ʔa cwá dǝ hóhé sō phó      They went to school three times.

(15c) ʔa cwá dǝ hóhé sí sō      Three of them went to school

PP. PP's introduced by the preposition *nɛ́* must be distinguished from all others: this construction poses a number of problems of analysis, which deserve a section of their own (6.5 below). PP's with *nɛ́* always precede other PP's; in what follows, what I say about 'PP's' does not apply to those with *nɛ́* unless the latter are specifically mentioned. The constituent PP, which can be given the grammatical-relation label *Obl*, is the normal realization of the Locative role. As in many other languages, this can be either selected ('inner' Locative) or non-selected ('outer' Locative); see discussion in 3.3.

*Obl* can also express other 'adjunct' notions than locative, depending on the preposition employed; in such cases it is non-selected. E.g. *hú* 'like, as, as if':

(16) ʔa | dǎ    lá te á | li    phá | hú [ phē    ʔirō ] | hé nǎ

3 give instead book skin like father sing say NA

He gave them a hide book instead, as Father sang, it's said (100.4)

Other non-locative prepositions include *chǎ́* 'when (future)' and *tí* 'as much as, as big as' (see 7.2 for a complete list of prepositions).

The selectional status of the Kayah grammatical relations as discussed so far can be summarized:

selected: Subject, Object(s), Obl

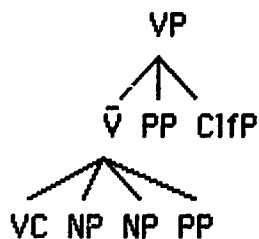
non-selected: Obl, Extent

uncertain: *né*-PP

Obl appears under two headings since it may be either selected or non-selected, as described above.

Assuming selection to be equivalent to subcategorization, it can be seen that the facts outlined above support the constituent analysis we have proposed for the Kayah clause. Subcategorization is usually assumed to be able to refer only to the sister nodes of the category in question; thus the constituents bearing the grammatical relations on the 'selected' list above must be sisters to the VC. Note that this requires VC to be capable of bearing subcategorization features; as we have seen, one way of allowing this is to analyze VC as  $\bar{V}$ , possibly a (quite complex) compound. The proposed analysis of the VP is repeated below:

(17)



As has already been discussed, the first  $\bar{V}$  in the 'source', or serial  $\bar{V}$  construction may contain a PP, but that PP is probably restricted (more evidence is needed) to a selected PP, like the inner-locative PP appearing with *ɔ* 'be at' in (6.2-2). How the *né*-



PP fits into this structure will be discussed below.

### 6.3 Subject (and Topic).

Non-embedded clauses in Kayah may include two NP's preceding the verb:

- 1            2  
 (18) bó        nλ | sɔ t̄ɔ̄ | khé k̄λ | hú pwā na | to mē ɔ  
          rice-plant Nλ weed overgrow COM like every year NEG CNC  
          It is true that the weeds didn't overgrow the rice the way they do  
          every year. (176.5)

- 1            2  
 (19) he khrē | ʔa | ʔe | lū | nλ  
          earth bug 3 eat 3OBV Nλ  
          Earth bugs ate it. (84.4)

- 1            2  
 (20) ʔā    tana | vē lù    vē bwí    vē lù tɔr̄λ | tā p̄hē ŋē  
          this one-year 1s merit 1s fortune 1s luck fall ahead front  
          This year my luck, my fortune keeps on getting worse. (178-9)

These three examples are typical. (18) has a 'fronted Object' followed by a Subject, (19) has a 'fronted Subject' followed by a 'resumptive pronoun', and (20) has a time expression followed by a Subject. At first glance these are all reminiscent of phenomena that have been discussed under the rubric of 'Topic'; I will now proceed to show that this impression is correct, and to clarify how I am applying the often vaguely-defined term 'Topic' to Kayah.

6.3.1. Subject vs. Topic: General. Li and Thompson's well-known paper (1976) offers this summation:

In conclusion, the topic is a discourse notion, whereas the subject is to a greater extent a sentence-internal notion.

This should be expanded on slightly: while topic is basically a notion relevant to the level of discourse, it has important consequences at the level of syntax; although subject is basically a syntactic-level notion, there are discourse-level conditions that bear on it. It is this criss-crossing of characteristics that makes both subject and topic such troublesome things to define and distinguish.

A distinction must be made between topic as a constituent and topic as the orientation-point of a process, namely the process of discourse-building. To illustrate, consider the following:

(21) That squirrel, I think maybe we should stop feeding him. Yesterday I saw him trying to open the kitchen window.

The first sentence contains a Topic constituent, *that squirrel*. The following discourse then takes that NP (or perhaps more properly, its referent) as the orientation-point of the process of information-giving; or in more familiar terms, the following sentence is 'about' the squirrel. The second sentence contains no Topic constituent; Topic is not an obligatory sentence constituent in English. It does, however, contain a Subject, which *is* obligatory in English. It also, in a sense, 'has' a topic, in that

there is something that it is about, namely the squirrel. It furthermore contains an item that shows the syntactic/semantic role played in it by that topic, namely the pronoun *him*.

But that pronoun is not a Topic constituent; rather it refers back to the Topic constituent *that squirrel*. The same function is performed by *him* in the first sentence.

In the present discussion of Topic and Subject, we will confine ourselves to distinguishing the two as constituents. Let us begin by surveying the sort of properties typically associated with Subject. These are selected from the many described by Keenan (1976), but are organized somewhat differently.

Morphosyntactic marking. 1) associated with a particular case (nominative). 2) tends to be leftmost NP.

Syntactic control. 1) tends to be controller of verb agreement. 2) controller (binder) of certain reflexive morphemes.

Semantic role. 'normally express[es] agent of the action, if there is one', unless the sentence includes some non-basic construction (e.g. passive).

Discourse status. tends to be 'old information', 'definite', identifiable.

As this list shows, Subject is a multi-level phenomenon, and so is 'primarily syntactic', as in the formulation above, only by contrast to Topic.

Topic contrasts with Subject in lacking all morphological, syntactic and semantic properties listed. It does not control verb agreement or the Subject-controlled type of reflexive morpheme, and has no necessary semantic role in the sentence. At the

discourse level, Subject and Topic overlap almost entirely, the main difference being that the requirement of definiteness is more stringent for Topic than for Subject. The discrepancy varies cross-linguistically, English being more tolerant than many languages of 'new', 'indefinite' Subjects. It is no accident that the examples chosen by Li and Thompson to demonstrate the optionality of definiteness/givenness for Subject are both from English: *A couple of people have arrived* and *A piece of pie is on the table* (L.&T. examples 3 and 4). Note that even in English these would be more likely to be *There's a piece of pie on the table* and *Here come some people*, with non-referential Subjects.

The fact that the bulk of the contrast between Subject and Topic lies at the levels of morphosyntax and semantics can be seen as following from one simple difference:

Subject is a constituent of the clause, while Topic (as a *constituent*, as discussed) is extra-clausal. Since agreement, reflexivization, semantic role patterns, and the like are all intra-clausal phenomena, it follows by definition that Topic should have no direct connection to them.

In generative theory Topic is often said to be a constituent of some category of a level above the clause (S) such as S' or S". This is of course one way of reflecting the extra-clausal status of the Topic; there may be others. I do not adopt that particular formalism, partly because it often goes along with a concept of Topic constructions as derived by a movement transformation. I prefer to leave open the possibility that Topic-Comment structures are basic.

6.3.2. Subject in Kayah. In seeking to define Subject as a grammatical relation in Kayah, we immediately find that most of the non-discourse-level properties of Subject are inapplicable. There is no case marking or verb agreement. The Kayah morphemes with reflexive-like meanings are not pronouns, but include one full noun (*ne* literally 'body') and a *ki*-class particle (*du* 'on one's own, of one's own accord'). The syntax and semantics of both of these morphemes need further investigation; at present no firm data are available on their interaction with the Subject. The semantic-role dependency of Subject on the verb is also of limited use as a diagnostic, since while Topics need not have a definable semantic role in the following sentence, they very often do. Examples are (18) above, in which the Topic relates to the Patient of the following clause, and (19), in which it relates to the Agent.

It might thus seem that there is no useful distinction to be drawn in Kayah between Subject and Topic. Possibly one could speak of first Subject and second Subject, along the lines of Chao's approach to Chinese. For instance, in (18) above the first Subject *bo* 'rice plants' has predicated of it *sət̚ khé ka hú pwā na to* 'weeds didn't overgrow [it] the way they do every year', that predicate itself consisting of *sət̚* 'weeds' as Subject with the remainder as Predicate.

There is, however, one minor but noteworthy syntactic phenomenon that allows a distinction in Kayah between Subject and Topic. This is an alternation between two of

the third person pronouns. Briefly, the third person pronouns are

ʔa	he/she/it/they; more specific, foregrounded
lū	id.; also foregrounded; the other one
ʔū	they, people, other people, someone; less specific, backgrounded; (also used in humilific self-reference)

The pronouns *ʔa* and *lū* can alternate within a clause as a way of keeping track of two third-person protagonists. *lū*, which I will refer to as the obviative form (abbreviated OBV) can only be used if *ʔa* or some other non-coreferential third-person NP appears before it in the clause. Compare the following two examples:

(22) ʔa chū ʔa thwi  
 3 stab 3 dog  
 He stabbed his (own) dog. (4/19)

(23) ʔa thū lū thwi  
 3 stab 3OBV dog  
 He<sub>i</sub> stabbed his<sub>j</sub> dog. (4/19)

More generally, successive (foregrounded) non-coreferential third-person NP's require an alternation between *ʔa* and *lū*. There are examples of more than one turn of alternation:

(24) ʔa ʔé ka lū dʒ ʔa kè  
 3 call return 3OBV at 3 country  
 They<sub>i</sub> called to them<sub>j</sub> to return [with them<sub>i</sub>] to their<sub>j</sub> country. (233.4)

The successive third-person NP's need not both be pronouns:

- (25) tamò chá ʔé lū  
 sun shine much 3OBV  
 The sun shone strongly on it (161.1)

The point to stress here is that this alternation operates only within the clause, thus providing a means of determining the boundaries of the clause. With the beginning of a new clause, the alternation must begin again. In the following example, note also that the third-person NP triggering the alternation may be realized as zero (here symbolized  $\emptyset$ ).

- (26) ʔa | ʔé ʔo jwā | lū | bɛ́ tē ||  $\emptyset$  | ʔo jwā kɛ |  
 3<sub>i</sub> call wait 3OBV<sub>j</sub> at what (3<sub>j</sub>) wait COM

lū | to | bō ʌ || ʔa | khrwā ʔíchō kɛ | lū ||  
 3OBV<sub>i</sub> NEG then 3<sub>i</sub> follow curse Kɛ 3OBV<sub>j</sub>

Whenever he called to them to wait they didn't wait for him;  
 then he followed cursing them (157.1)

Here the three occurrences of *lū* show that there are three clauses, each of them containing a third person Subject NP that triggers the appearance of *lū* as the Object. Note that the *ʔa-lū* alternation is not a 'switch-reference' marker in the usual sense of the term. For instance, in the third clause in (26), the Subject *ʔa* is coreferential with the nearest preceding NP *lū*, but it would have also been *ʔa* if the preceding NP had not been coreferential. This is because it is the first NP in the clause, and so must be *ʔa*; coreference or lack of it with NP's in preceding clauses is irrelevant. Switch-reference

usually marks identity of referents across clauses; the *ʔa-lū* alternation marks identity of referents within clauses. Another point is that the third-person NP triggering the alternation need not be a Subject, although it often is, as in all three clauses in (26). It is also possible for a non-Subject to be the trigger:

- (27) *vē dʰ pè kó k̄ ʔa lū te*  
 Is give TRN temporarily COM 3<sub>i</sub> 3OBV<sub>j</sub> thing  
 I lent him<sub>i</sub> his<sub>j</sub>. (10/8)

Thus, triggering the alternation is not a property of the Subject alone; it is a property of all NP's within the clause. It does allow us to distinguish Subject from Topic.

The alternation is not triggered by Topics, as is shown by the following examples in which a Topic is followed by a noncoreferential Subject:

- |-----Topic-----| Sbj  
 (28) *ʔa dō l̄ t̄p̄w ja ʔa phjá ʔo m̄l thó phe*  
 3 thick CMP one-clf (?) 3 take sleep cover supplanting  
 The thicker one [blanket], he took to sleep under (273.2)

- |----Topic--| Sbj  
 (29) *h̄l ca p̄k̄w ʔa sí jw bó mé*  
 clothes Sgaw 3 want weave look  
 Sgaw Karen clothes, [if] he wants some [I'll] weave some and see (319.5)

In each example the Topic is not coreferential with the following third-person Subject NP, but it does not trigger the alternant *lū* for the latter. We interpret this as



evidence that the Topic is not a constituent of the clause, while the Subject is.

We thus have a syntactic property capable of distinguishing Subject from Topic.

Naturally, this property is not available as a diagnostic for testing every preverbal NP, since it depends on certain features not found in every clause; namely the occurrence of several non-coreferential third-person NP's of which at least one is an overt pronoun and is not the first in the clause.

Several other characteristics of Topic in Kayah should be mentioned. One is that, although examples like (28-29) in which Subject directly follows Topic, are not uncommon, very frequently there is something setting off the Topic from the following clause, such as the particle *m* (see 8.3) or the very common morpheme *ma* (9.4). Secondly, sentences beginning with only one NP could conceivably represent a) Subject and no Topic; b) Topic and no Subject; c) a conflation of Subject and Topic. One type of sentence with this pattern deserves some comment, e.g.

(30) khrw      ʔi:chi    lɔ̃      ʔ  
           firewood chop use-up NS

The firewood is all chopped. (3/3)

(31) phú cè dʔ cwá    lɔ̃      ʔ  
           child let go use-up NS

The children have all been allowed to go. (2/20)

(32) ʔā    tʔmē    bɔ̃      sō    nā    jʔ    ʔ  
           this one-clf weave three day PTC PTC

This loom-set [I've] been weaving three days (306.2)

In these a two-argument verb is preceded by an NP that could in other circumstances appear as its Object. Similar examples are common in other languages of the area.

(33) Chinese zhèjian shì zǎo fābiāole

this-clf affair early publish-PFV

This matter has long been published (Chao, p.70)

(34) Thai kàj tua nán kin léew

chicken CLF that eat PFV

That chicken has eaten/has been eaten.

In Chao's view these preverbal NP's in Chinese are Subjects, since 'the direction of action in an action verb in the predicate need not go outward from subject to object' (p. 72). The verbs could then be treated like the English verbs *open, break, drop* and so on (*The book dropped / John dropped the book, etc.*). There is one bit of evidence that the Kayah sentences are to be analyzed as in b), as having Topic but no Subject, or perhaps zero Subject. Consider

(35) pe phú pe pò kəŋəə X síjə íx rá ʔa

1p child 1p YS youngest care-for more Rλ 3

Our child, our youngest brother, [we] care for him the most. (97.2)

Here if the NP *pe phú pe pò kəŋəə* were Subject, the Object pronoun would have to be the obviative /i/, since the two are noncoreferential third person NP's. *pe phú pe pò kəŋəə* must then be outside the clause, a Topic. One might also claim that there is a Subject, in the form of a phonologically empty pronoun in the position marked X. This X would have to be marked as first person: if X were third-person, the Object pronoun would have to be /i/.

Theories that utilize empty NP's do, in fact, allow them to be marked for person; their feature matrix includes features like person, number, and gender, but no phonological features (Chomsky 1981). However in the analysis I have presented of the Kayah sentences, the only necessary assumption is that there is a clause boundary in the position X; I have not defined the clause as requiring a Subject:

[pe phú pe pò kəneə]<sub>NP</sub> [<sub>S</sub> [<sub>VP</sub> síjɔ́ lɛ́ rá ʔa ]]

The occurrence of ʔa proves only that there is no noncoreferential third-person NP preceding it in the clause. This may mean that there is a preceding empty *non*-third-person NP, but it may also mean that there is simply no preceding NP at all.

6.3.3. 'Inverted Subjects'. Mention should be made of a class of verbs in which the participant of which the state is predicated typically appears post-verbally, while the Subject position, if present, is occupied by the pronoun ʔa. Many of these denote bodily sensations or emotions, e.g.

(36) kēsé 15 | vē nè

itch use-up 1s body

I itch all over. (127.2)

(37) ʔa mo pa | lū síplɔ́ | to

3 happy DUR 3OBJV heart NEG

Their hearts weren't happy. (89.5)

Presumably these verbs specify in their lexical entries both the single argument and its realization as Obj-x.

#### 6.4 Objects

A VC may be followed by zero, one or two unmarked NP's (i.e. without preposition). In the discussion that follows, I refer to the first of two unmarked postverbal NP's as 'Obj-1', and to the second as 'Obj-2'; in the case of a single unmarked postverbal NP, I use 'Obj-x'. This reflects the question that we will be examining: which of the two Objects should be identified with the single Object? We will first look at the semantics of Obj-x

Let us use 'root  $V_t$ ' to refer to lexical verbs that subcategorize for Subject and Object, or to VC's containing such verbs but no argument-adding morphemes (whether Particle or second verb in a V-V). 'Derived  $V_t$ ' will then be a VC including a lexical  $V_t$  and some argument-adding morpheme. With root  $V_t$ 's, Obj-x is a fairly typical Object relation in the semantic roles that it realizes: Patient with verbs specifying Agent and Patient; 'content' of perception with verbs like *metha* 'see', *nidā* 'hear'; 'goal' (subsumed under Locative?) of verbs like *sijw* 'want', *āphri* 'buy', and *mē* 'look, look for'.

With derived  $V_t$ 's, Obj-x may have additional semantic roles, which can be grouped in the following way:

1. 'content' or 'goal' of emotional state, with Particle *rā*

(38) ?a síp) du rā ?a phē

3 angry Rλ 3 father

He's mad at his father. (10/29)

(39) ʔa sí nì so rál ʔa mē

3 miss Rλ 3 wife

He misses his wife. (10/29)

(40) taphreʔukhré béswá rál panè

(name) be-friends Rλ buffalo

T. was friends with a buffalo. (19.3)

2. Comitative, with *ɾí*-class Particle *kā*

(41) ʔa cwá kā vē

3 go COM 1s

He goes with me.

3. Standard, with *ɾí*-class Particles *lǎ* and *khɾw*

(42) ʔa ʔibe cè lǎ vē

3 speak able CMP 1s

He can speak better than me. (9/29)

(43) vē ʔire phrē khɾw kā ʔa to

1s work fast equal Kλ he NEG

I can't work as fast as him. (10/31)

4. Benefactive/Malefactive, with Descriptive Particles *cwá* 'help' and *bébw* 'show the way to', and *ɾí*-class Particle *pè* 'to sbdy's benefit/detriment'

(44) ʔa s̄ ɩ̄ pè ɩ̄

3 rot use-up BEN 3OBJ

It all rotted 'on' them. (155.7)

(45) ʔa cwá bébw kulā d̄ cho khā

3 go show-way European at mountain upper-surface

He takes Europeans up the mountains. (6/25)

## 5. Causee, with Directive V-V's

(46) *vē dǎ cwá ne to*

1s let go you NEG

I won't let you go.

(47) *ʔa . nā ʔo nē phú cè*

3 command sit child

He told the children to sit.

6. point of orientation (Locative), with *talwá*-class Bound Directionals (4.3.7.2)(48) *ʔa cwá talwá vē hi*

3 go pass 1s house

He went past my house.

I have no evidence of any difference in syntactic behavior between Obj-x with root  $V_t$ 's and with derived  $V_t$ 's. Of course, this is not a very strong indication of anything, since Kayah lacks any syntactic phenomenon that requires reference to Object, such as passive, raising constructions, case-marking, or verbal cross-indexing.

We now go on to consider Obj-1 and Obj-2. VC's that subcategorize for Subject, Obj-1 and Obj-2 can be referred to as  $V_d$ , with the same distinction between root and derived as used above with  $V_t$ . Root  $V_d$ 's, which are not numerous, share the semantic element 'transfer of possession'. They frequently occur with the Bound Directional *pe*, which can be given the TRN (for 'transfer'). In most cases Obj-1 has the Recipient role and Obj-2 is the Patient (specifically, the thing transferred). Examples:

- (49) vĕ dǎ́ (pè) Khōmè rŭ  
 1s give TRN (name) money  
 I give K. money.
- (50) ʔa ʔíchĕ (pè) vĕ thé nā dō  
 3 sell TRN 1s pig two clf  
 He sold me two pigs.
- (51) ʔa bo ʔe (pè kā) thé che  
 3 feed for-use TRN COM pig feed(n.)  
 He feeds the pigs [their] feed. (10/29)
- (52) sárá ʔīswá phú cè li  
 teacher teach child letter/book  
 The teacher teaches the children (their letters). (2/1)
- (53) ʔa bulĕ vĕ haca  
 he exchange 1s clothes  
 He exchanges clothes with me. (10/31)<sup>1</sup>

With *kwī* 'ask for, request', Obj-1 seems to be Source rather than Recipient:

- (54) já kwī tíí ʔū rŭ ma  
 go-and ask-for constantly 1s money PTC  
 [you're] always asking me for money! (136.6)  
 (humiliific use of *ʔū*)

- (55) vĕ | kwī khjā sé | ʔa | lō tóthé  
 1s ask-for back again 3 bicycle  
 I ask him for the bicycle back. (10/29)

This structure, [VC Obj-1 Obj-2], is the only way of casting these events in a single clause. There is no form having the Recipient in a prepositional phrase, and thus no question of 'Dative Movement' or the equivalent.

With derived  $V_d$ 's, Obj-1 usually corresponds to the added Obj-x of derived  $V_t$ 's. The following set of examples is grouped into the same six categories as those used in the preceding discussion of Obj-x with derived  $V_t$ 's:

(1. there is no good equivalent to content of emotional state)

2. Comitative

(56) ʔō    tu    kāl | vē | thāl ʔiphɾè  
 drink PTC COM 1s whiskey  
 Drink (whiskey) with me. (10/20)

3. Standard

(57) ʔa | ʔe ʔé    lɛ́ | vē | dī  
 3 eat much than 1s cooked-rice  
 He eats more than me. (9/29)

4. Benefactive

(58) ʔa | me cwà    | Mīl    i    Thīm    hi  
 3 do help (name) (name) house  
 He's helping M. build T's house. (10/20)

(59) ʔa vē pè    lū    he so  
 3 dig BEN 3OBV earth  
 He dug out earth for her [i.e. doing her job]. (36.6)

5. Causee

(60) | nō        já    dō |    lū    i    thé pānā  
 command go-and forge 3OBV chisel  
 He told them to go forge a chisel. (94.1)



(61) ʔaphē | ʔiswá khē | ʔaphú | sɔkɪ̄l̄

father teach paddle 3-child boat

The father teaches his child to paddle a boat. (2/1)

The single exception to this equivalence between Obj-x of derived  $V_t$  and Obj-1 of derived  $V_d$  is the sixth category, point of orientation (Locative). When a root  $V_t$  is followed by a Bound Directional, the result is not a  $V_d$  but a new  $V_t$ , with the 'extra' argument expressed, if at all, as object of the Preposition *né*. However, either the Patient of the main verb or the Locative of the Bound Directional may be 'extra' :

#### 6. Locative with /ə/-prefixed Bound Directionals

(62a) ʔa | khē talwá | sɔkɪ̄l̄ | né sɔkhō

3 paddle pass boat Né snag

He paddled the boat past the snag (fallen log).

(62b) ʔa | khē talwá | sɔkhō | né sɔkɪ̄l̄

3 paddle pass snag Né boat

(same meaning)

For discussion of the function of *né*, see 6.5 below.

The most straightforward way to account for this difference between categories 2-5 and category 6 is not by reference to a contrast between Locative (Cat. 6) and Causee, Standard, Benefactive, etc. (Cat. 2-5), but in terms of syntactic subcategorization. The argument-adding morphemes *kā̄*, *lɔ̄*, *khɪw*, *cwà* and others found in categories 2-5, besides specifying a semantic role, must also add to the verb a syntactic feature

allowing Obj-1; the Bound Directionals like *ta/wá* do not. Therefore the VC's in categories 2-5 have a syntactic valence of [Sbj Obj-1 Obj-2], while those in category 6 have the valence [Sbj Obj-x]. The behavior of all categories can then be described in terms of the hierarchical realization principles described in 5.3.3, as follows. The highest-ranking semantic role (Agent, in the examples) is first 'linked' to Subject; if there is only one other argument specified, it is second-highest by default and is realized as Obj-x. If there are two non-Subject arguments, there are two possibilities. In categories 2-5 the Patient (or other more saliently affected) becomes Obj-2 as usual, and Obj-1 is available for the remaining argument. In category 6, only Object-x is available to the non-Subject arguments.

It might seem that example (62b), in which the Locative has 'pre-empted' the Patient for the Obj-x realization, violates the realization hierarchy. However there is already evidence that Bound Directionals like *ta/wá* must specify Obj-x realization for their Locative argument, as demonstrated by example (48) (see also 4.3.7.2). Evidently such a specification is enough to allow the apparent pre-emption of the Patient, which can nevertheless appear in an Oblique phrase introduced by the preposition *né*. Although the proper analysis of PP's with *né* is not entirely certain, I will suggest in the next section of this chapter that this constituent is not subcategorized for, but is always available for the expression of a 'backgrounded' participant.

To sum up, there are the following semantic affinities:

Obj-x with root  $V_t$  .....Obj-2 with root  $V_d$

Obj-x (not Locative) with derived  $V_t$  ..... Obj-1 with root  $V_d$

Obj-x (Locative) with derived  $V_t$  ..... Obj-x with root  $V_t$

if it were not for the behavior of Obj-x with derived [+Locative] verbs, we would be inclined to identify Obj-x with root  $V_t$  as Obj-2 and Obj-x with derived  $V_t$  as Obj-1. But in light of the facts in (25a-b) it is best to interpret Obj-x as a neutralization of the distinction between the two relations Obj-1 and Obj-2.

#### 6.5. $n\acute{e}$ -PP's.

The form  $n\acute{e}$  has a variety of functions. Some of these are distinctive enough that it makes sense to speak in terms of separate morphemes, which nevertheless have a semantic family resemblance. In other cases it is unclear whether there is a division to be drawn between syntactic functions, and hence between separate morphemes.

Subcategorized. With some verbs (only two known thus far)  $n\acute{e}$  is required with the following argument, which may be a NP or a clause; in the latter case  $n\acute{e}$  can be called a complementizer:

(63)  $\text{?a} \mid \text{l}\bar{a}$       takhrē kāl h\acute{e} [  $\text{?a} \text{ ?ó} \text{ p}\bar{i}l\acute{o}$  t\acute{i}kwa d\text{ } m\acute{e}k\text{ } l\acute{u} N\text{ } ke ]

3 intrusively similar COM    3 blow flute flute beat drum N\text{ } PRH

It's as if he's blowing flutes and pipes and beating drums. (59.2)

(64) me sɛə | nɛ́ [ ʔū pwe ] to  
do same 3 celebrate NEG

They don't do it like a festival.

Cf. also, with following NP:

(65) ʔa sɛə nɛ́ vɛ́ te  
3 same 1s-thing

It's the same as mine.

It also appears in the pattern forming ordinal numerals, NUM-CLF-nɛ́-ʔā; e.g. sō ʔú nɛ́

ʔā 'the third volume', nɛ́ be nɛ́ ʔā 'the fifth one [bird]'.  
..

But there are many occurrences of nɛ́ in which it is not subcategorized in the way that it clearly is with sɛə and takhrɛ́ above; at the same time, the semantics of the nɛ́-PP in these occurrences are not unrelated to the argument structure of the verb or VC.

One such use can be labelled 'Quasi-coordination'. In it nɛ́ has the flavor of a coordinating conjunction, at times simply joining NP's:

(66) təcù | vɛ́ sí ʔe kāl lɔ nɛ́ dāl  
bland 1s want eat COM gourd melon

For blandness, I want to eat gourds and melon (28.1)

(67) ʔa khé ka nì thuú sō be nɛ́ takhɔ tədō  
3 shoot come get bird three CLF muntjac one-clf

He shot three birds and a muntjac. (2/22)

(68) mo du de dwá dʒ vɛ́ hi nɛ́ Phāɔ hi  
gong big put put-away at 1s house (name) house

The big gongs are kept at my house and Pha'a's house. (2/22)

- (69) ʔa nɛ ʔa phā́ jǎ ʔo tʷə lé hē rē lē  
 3 3 GMo go-and be-at PTC pumpkin trellis base

He and his grandmother went and lived at the base of a pumpkin trellis. (19.5)

In these the meaning is something like 'and, moreover', with the conjoined NP's acting like a single NP in being able to appear as Obj-x (66-7), object of preposition (68), or Subject (69). In the following examples, however, *nɛ* does not appear between two NP's, but precedes a single NP (70-71) or has no associated NP's at all (72). Its meaning seems to be more like a sentence adverbial 'moreover, additionally'.

- (70) ʔū | tɛpló rǎ | nɛ ʔa pɔ  
 3 put-on RA 3 finial

They also put on its<sup>2</sup> finial. (47.4)

- (71) ʔe chā dɬ, ʔe pò nɛ khómū  
 eat chicken egg eat additionally bread

[I] ate eggs, and I also ate bread.

- (72) tɛ | ʔo kǎ | nɛ pā ke  
 thing exist COM IRR PRH

[we] may soon be rich as well.

In some such cases it would be possible to maintain the view of *nɛ* as a conjunction by claiming that there is a phonologically null pronoun preceding it. The meaning of [VC*nɛ*NP] would then be 'to V X and NP' where X = something definite, mentioned, recoverable, etc.; thus a more literal rendering of (71) would be 'I ate eggs, I ate them and bread'. But this interpretation is not possible with (72), in which there are no

postverbal NP arguments at all<sup>3</sup>. It is also not appropriate for (70), as is clear from the context (glosses simplified):

(70b) ?ū rō ?e thō rλ || ?ū sú plī phō cha se || sú plī thō rλ ||  
 3 hew finish 3 wash soap-pod wash finish  
 plo bū rλ thwí || plo bū thō á pā pō rλ || ?ū tapló rλ  
 smear white lime smear white finish finish 3 put-on  
 né ?a po || tapló thō ?a po ?ā nλ ma ||...[continues]  
 finial put-on finish finial this be-so

When they've finished hewing it, they wash it with soap-pod; when they're done washing it, they paint it white with lime; when they're done painting it white, they also put on the finial; when they've put on the finial, well then . . .

The point is that *?ū tapló rλ né ?a po* does not describe an additional instance of putting-on, since the finial is the first and last thing to be put on the *?ilū*: there is no X such that *?ū tapló rλ X né ?a po* 'they put on X and the finial'. Rather a meaning like 'additionally' is applied to the entire clause, putting it as an additional step in the procedure of preparing the *?ilū*.

There is a further use, or set of uses, of *né* with little or no coordinative flavor, in which it seems to mark an argument that is 'extra', in a sense that will be described below. These extra arguments can be subdivided into the following three types.

Goal

(73) ʔa | s̄ | phre m̄ hɿ jē | ɲé síne

3 ram-in woman shirt old gun

He loaded an old blouse into his gun. (228.6)

(74a) phú cè vī the l̄ ɲé hi

child throw go-up stone house

The child threw a stone up at the house. (5/11)

(74b) phú cè vī the hi ɲé l̄

(same meaning as 12a)

(75a) ʔa | khē təlwá | sɔkɿ̄ | ɲé sɔkhō

3 paddle pass boat snag

He paddled the boat past the snag (fallen log).

(75b) ʔa | khē təlwá | sɔkhō | ɲé sɔkɿ̄

(same meaning as 13a)

(76) sī | pā be | ʔídw̄ | ɲé l̄

2p chop strike machete stone

You chop striking your machete on a stone. (157.4)

Instrument

(77) ʔa chū s̄ lū ɲé ʔithoə

3 stab die 3OBV knife

They [tried to] stab him to death with a knife. (354.4)

(78) ʔa khé lū ɲé síne

3 shoot 3OBV gun

They shot him with a gun. (354.)

- (79) ʔa chū li mi ɲé tɔthé  
 3 stab red fire iron  
 They poked the fire red with an iron. (355.5)
- (80) ʔa|chú ʔú lɔ pī chæ | v̄e te ɲé mí | n̄l p̄w  
 3 burn smolder use-up complete mine fire two CLF  
 She burned up two of mine [blankets]. (272.3)
- (81) ʔa phō ɲé dīpò du n̄  
 3 cook pot big NA  
 They cooked [it] in a big pot. (356.4)  
 (could also be listed under Goal)

Other

- (82) [ʔū | kwā s̄l | ʔū nè | ɲé sɔ ] s̄  
 3 chop die 3 body tree those-who  
 the ones who get killed chopping trees (255.4)
- (83) v̄e | n̄ d̄l p̄e k̄l | Ph̄a | r̄w̄ | ɲé ʔa p̄ò  
 1s command give TRN COM(name) money 3 YS  
 I told Pha'a to give his younger sibling money.  
 (could also be listed under Goal)
- (84) taphō thē | khe le | ɲé īp̄ana  
 stub wound foot nail  
 [he] stubbed his toe and wounded it on a nail.
- (85) v̄e pí cwí p̄ù ɲé sɔ  
 1s whip pull ox tree  
 I whipped the ox to make it pull the log.  
 - I got the ox to pull the log by whipping it.



(86) ʔa me ʔo nē bja vē nē liú

3 do sit damage 1s book

He made me sit on and damage the book.

(87) ʔa bule vē nē haca

3 exchange 1s clothes.

He gave me clothes in exchange. (10/31)

Regarding realization of semantic roles, it does seem that in some of these examples the *nē*-PP can be interpreted as having a semantic role specified by the verb. This applies best in the 'Goal' examples, most of which do contain verbs specifying a Locative<sub>Goal</sub> argument; *ʔā* 'chop' would seem to be an exception, however. In addition, it is not true that the specified Goal is required to appear as a *nē*-PP; as might be expected, it easily occurs as PP with *dě*. For *vī* 'throw' cf. *vī the dě hi dō kū* 'throw [something] up into the house' (8/3); for *sō* 'push in with a tool' cf. *sō tāite dě síne kū* 'ram something into the gun'; and so on.

As for the 'Instrument' set, this semantic role is one that can be seen as 'derived' rather than basic, as discussed in 3.2.3 above. The frequent appearance of Instrument(-like) arguments in the *nē*-PP is in fact one of the reasons for my decision to exclude Instrumental from the repertoire of (linguistically-encoded) semantic roles, as I hope to make clear shortly.

The final category, 'other', provides the key. To the extent that the semantic role of

the *né*-PP in these sentences can be identified, it seems to duplicate the semantic role of some other NP in the sentence; in particular, Patient in (82, 85, 86). The *né*-PP in (83) can also be said to duplicate a semantic role under the assumption that causee of Directives and recipient with transfer verbs are both Recipients. Given this duplication of roles, we cannot say that the *né*-PP 'has' those semantic roles without violating the constraint against multiple instances of one role type in the same clause. In terms of the general semantics of the situation described, the *né*-PP participant might have been eligible for these roles, had it not been pre-empted or demoted by a second participant with greater eligibility. This eligibility apparently has much to do with saliency of effect: in all the examples (82-87) a case can be made that the Object-x participant is more saliently affected than the *né*-PP participant. In all but (83) the Obj-x is higher on an animacy scale, either human as opposed to nonhuman (82, 84, 86, 87)), or animal as opposed to plant (85). In (83), the competition is for Obj-1 position: both participants are human, but it is plausible to class the effect on the causee (being caused to perform an action) as more salient than the effect on the recipient of goods.

I believe that the *né*-PP in the 'goal' and 'instrument' types is best interpreted in the same way: in spite of the labels, these should not be considered to realize any specified semantic role. In this view, the function of *né* is to mark some participant as demoted, backgrounded, peripheral, etc. The relation of that participant to the action may be understood to be *like* one of the linguistically-encoded semantic roles, but in its

function as object of the *ne*-PP it does not count as a bearer of that role. Consider (85): the VC specifies Agent/Subject and Patient/Obj-x, these being assigned to *ve* and *pu* respectively. However the VC is made up of two verbs, both specifying Agent and Patient and the Object-x NP refers to a participant that can be inferred to be the recipient of the action of  $V_1$  but not of  $V_2$ . It is then inference that tells the hearer that *so* 'tree' is in a Patient-like relation to the VC, and particularly to the VC's component verb *cwi* 'drag'. Furthermore it can be said that *pu* appears as Object-x because it outranks *so* for the Patient role, by virtue of being more saliently effected. (Cf. 3.2.3: effects on animates are always more salient than effects on inanimates) also *pu* is a potential Agent that is cast as Patient by the (marked) type of Resultative V-V, with two  $V_i$ 's. We may then say that the grammatical meaning of the *ne*-PP is 'backgrounded participant'. Any more specific interpretation of the relation of that participant to the action is the result of inference. The inference may be based on information that includes knowledge of the argument structure of the verb, but also includes extra-linguistic knowledge. There is a distinction to be made between the purely-linguistic process that assigns semantic roles to constituents and this mixed process, involving both linguistic and extra-linguistic knowledge, that arrives at interpretations of the *ne*-PP.

In this view the 'Instrument' role can be seen as a special case of a more general phenomenon of backgrounded participants. In support of this it may be noted that the

verbs in (77-81) may also take the 'Instrument' participant as Obj-x. For *khé* 'shoot' cf. *ʔa thwā jasú, khé rá síne* 'it turns into gunpowder, [and you] shoot [it in] a gun' (170.3). For *chū* 'stab, poke', cf. the expression *chū bó* 'to transplant rice plants', where *bó* is the thing that is inserted, thus parallel with *ʔithoə* in (77) and *tóthé* in (79). These facts harmonize well with Fillmore's treatment of Instrument as a Patient that has been 'set aside' (3.2.3 above); additionally examples (82-87), in which the backgrounded Patient is marked with *né* but is not Instrument-like, are akin to Fillmore's *load the truck with hay* in demonstrating that 'backgrounded Patient' is the more general concept, and so presumably more important.

The boundary between these 'extra-argument' *né*-PP's and the quasi-coordination type is unclear. On the one hand, it may be that even the latter can be given an interpretation in which *né* precedes a backgrounded participant, the only difference being that it is not very severely backgrounded. On the other hand, it might be possible to reduce some of the extra-argument type to something like coordination. The reversability of the NP's in some instances (74a-b, 75a-b) could be evidence in either direction: with the former hypothesis, one could say that the reversible cases involve participants whose degrees of Object-eligibility are so close that it does not matter so much which one is backgrounded as long as one is (the danger here is in reducing the notion of 'backgrounding' to vacuity). With the latter hypothesis, of course, reversability could just count as a characteristic of coordination. However this would leave many

cases in which the NP's cannot be reversed, and it seems better to predict that some difference in emphasis, discourse status, etc. will be discoverable for the seemingly reversible cases.

The interpretation I am suggesting for *né* is similar to that often given for the English preposition *by*. In passive sentences, *by* indicates that its object has the semantic role associated with the Subject of the sentence's active counterpart. In Kayah *né* has a similar function, but in relation to Object rather than Subject: the semantic roles realized by *né*-PP's have the same range as those realized by Obj-x's.

More research is needed on this construction. We need more data on reversability, alternative realizations of participants marked by *né*, degree of obligatoriness of *né*-marking, and other points. Finally, it is worth pointing out that *né* is likely to be a loan from Burmese: Written Burmese *náí*, modern *né* (creaky tone) 'and, both; by, with, from, to, *manner*, etc.' (Okell, 120-1).

Constituency of PP's revisited. Since the  $n\acute{e}$ -PP does not realize a specified semantic role, it must not be a constituent of  $\bar{V}$ . We have thus arrived at a picture of the PP's that may be summarized as follows:

$n\acute{e}$ -PP	nonselected	sister of $\bar{V}$
locative PP	$\pm$ selected	daughter or sister of $\bar{V}$
other PP's	nonselected	sister of $\bar{V}$

Only these combinations are possible:

$n\acute{e}$ -PP + other nonselected PP

selected PP + nonselected PP

nonselected PP + nonselected PP

Evidently the  $n\acute{e}$ -PP must precede any other PP, since we do not appear to get

\*[  $n\acute{e}$ -PP + selected PP]. Examples:

selected (locative) + nonselected (other)

- |      |                                                                          |                 |                 |                 |                                     |
|------|--------------------------------------------------------------------------|-----------------|-----------------|-----------------|-------------------------------------|
|      |                                                                          | PP <sub>1</sub> |                 | PP <sub>2</sub> |                                     |
| (88) | ʔa                                                                       | phā             | nλ   ʔo   dʒ hi | thw̄ i          | hú [ pe ʔo ʔā ] nλ                  |
|      | 3                                                                        | grandmother     | Nλ              | be-at           | at house edge like 1p be-at this Nλ |
|      | His grandmother lived at the edge of the village, as we do here. (204.4) |                 |                 |                 |                                     |
|      | (ʔo specifies an inner Locative)                                         |                 |                 |                 |                                     |

né-PP + nonselectedPP<sub>1</sub>    PP<sub>2</sub>

(89) ʔa|pā    be | ʔídw| né    l̩̀| d̩́ mi    klē

3    cut strike knife    Nε rock at forest among

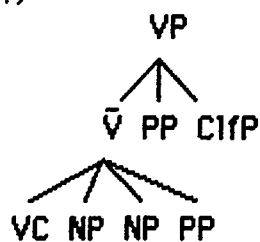
He cut striking his knife on a rock in the forest. (10/31)

(pā does not specify a Locative)

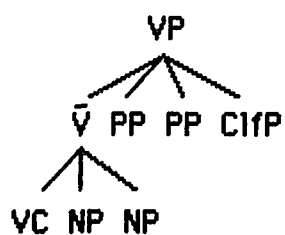
This means that the structure given in (1) and (17) above is actually one of two

alternatives for a maximal VP:

(a, = 1, 17)



(b)



## 7. NP, PP, ClfP: Syntax and morpheme classes.

The possible kinship of these three categories, which provides justification for grouping them together in one chapter, will be discussed below. I will also present some additional detail concerning the classes of Noun, Preposition and Classifier, plus aspects of the internal syntax of the NP, PP and ClfP.

### 7.1 NP.

Several types of nouns may be distinguished:

Common nouns	thwi 'dog', so 'tree', hi 'house', bwí 'fortune'															
Names	personal names: Phāλ, Khōλ, Sēthūphē, Pǝmè place names: Mē Lē, Thā Médʁ Lē Khā, Phremesō															
Pronouns	<table> <thead> <tr> <th></th> <th>Sg</th> <th>Pl</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>vē</td> <td>pe</td> </tr> <tr> <td>2</td> <td>ne</td> <td>sī</td> </tr> <tr> <td>3</td> <td>ʔa, lū</td> <td></td> </tr> <tr> <td></td> <td></td> <td>ʔū 'somebody' (ʔū ... pē 'who?')</td> </tr> </tbody> </table>		Sg	Pl	1	vē	pe	2	ne	sī	3	ʔa, lū				ʔū 'somebody' (ʔū ... pē 'who?')
	Sg	Pl														
1	vē	pe														
2	ne	sī														
3	ʔa, lū															
		ʔū 'somebody' (ʔū ... pē 'who?')														
Localizers	khā 'upper surface', kū 'inside'															
Classifiers	be 'clf for flat sheets', vē 'clf for seasons', cʁ 'clf for kinds, types'															

Pronouns and Localizers are listable classes, Common nouns make up an open class, and Classifiers probably do also (e.g. the name of any container can be a classifier). Four categories can usefully be distinguished (grouping names and pronouns together), on the basis of the type of modifier they may take.



1. Common nouns may be modified by NP's and/or postposed clauses, the clause often consisting of only a single verb.

2. Localizers may be modified only by preposed NP's.

3. Names and pronouns generally are not modified<sup>1</sup>.

4. Classifiers must (being bound forms) be modified by a Quantifier.

Some notes on the pronouns:

1. Both *ʔa* and *ʔū* are unmarked for number. The rule determining their occurrence has already been described in 6.3.2 above.

2. *ʔū* is indefinite, backgrounded, often to be translated as 'other people' or 'they' as in 'they say' (*ʔū hē n̄*, standard expression in legendary narratives; see 9.3.3). It means 'who?' when the sentence ends in the particle *pē* (cf. 8.2, 8.3). It may also be used as a humilific first-person pronoun: this is the only instance I know of in which Kayah has linguistic marking of status, something that is usually not found in the languages of the hill cultures in Southeast Asia (compare the elaborate status-marking apparatus of Thai).

3. *sī*, besides being the second-person plural pronoun, is also a bound noun meaning roughly 'and the rest, and things like that', e.g. *ʔamɔ sī* 'Amaw and that group' (similar to Mandarin *Zhāngsān tāmen*), *thwí k̄ sī* 'the lime-box and all that sort of thing', and

(1) *sā ré to ma hú ʔū tā sā thā sī*

die good NEG be-so like 3 fall die water

Dying badly is like those who are drowned and so on. (255.3)

Here *sī* is modified by the preceding clause *ʔū tā sā thā* (which must be considered

to be nominalized; see 9.2), forming an NP that then acts as object of the preposition *hú*.

At this stage of research, it seems best to describe three different types of NP, with the hope that all may eventually be assimilated under a single structure, expressible by a single set of phrase structure rules expanding NP.

1. Lexically-headed NP. As the name states, the head of this type is a lexical noun: a common noun or a localizer, possibly with modifier; or a name or pronoun, usually not modified, and thus making up its own NP.
2. Clf-headed NP. These may consist of a ClfP alone (Quantifier and Classifier), or a ClfP preceded by either a demonstrative or a nominalized modifying clause.
3. Expanded NP. A partial combination of the two: a lexically-headed NP modified by (or modifying?) a following ClfP.

7.1.1. Lexically-headed NP's. The order of constituents in the lexically-headed NP can be described in terms of a default rule and specific exceptions to it. The default rule is that nominal modifiers precede the head, and verbal modifiers follow.

<u>nominal+head</u>				<u>head+verbal</u>			
ne mò		your	mother	rù du		big	snake
1	2	1	2	1	2	2	1
chā dā		chicken	egg	phre bómō		the person [who]	opened [it]
1	2	1	2	1	2	1	2
				kājē pā tha		people [to]	cut sesame
				1	2	3	1

A combination of the two is seen in:

phremò hɔ̃ je	tattered woman's skirt
1    2   3	3    1    2

The postposed verbal modifier is in fact a clause, since in addition to the verb it may contain constituents which relate to the verb in exactly the same way as do the constituents of a clause; e.g., *kajē pā tha* includes, besides the head *kajē* and the verb *pā*, the NP *tha* functioning as Obj-x in relation to *pā*. The other type of modifying clause, which modifies a following ClfP, can be interpreted as nominalized, thus falling in with the general rule for modifiers, prehead for nominals and posthead for verbals. Both of these types of what will be called attributive clauses will be described further in Chapter 9.

#### Nominal modifiers.

The statement that nominal modifiers precede the head is actually a generalization to which there are semantically-definable classes of exceptions. The general modifier+head order signifies possession, as in:

ne mò	your mother
Phāλ hi	Pha'a's house
ʔū ro kè	other people's country

or some more general meanings that may be seen as extensions of the possessive meaning, as:

chā dA chicken egg

thé ja pig meat

cho khā mountain's apex → on top of the mountain

bése thā face-water → tears (bése may be short for bése plo 'eye')

khε bó leg-rings (item of costume)

Locative expressions including localizers are built on this type; for instance *dó khrā* *kū* 'inside the bottle-gourd', more literally 'at the bottle-gourd's inside', with *khrā* modifying *kū*. In other words the Localizers are not equivalent to English prepositions, except semantically: *kū* here is the head of an NP which in turn functions as the object of the Preposition *dó*. For a complete listing of Localizers see 7.1.4.

Exceptions to the general rule are of three types:

1. artifact-material. Examples:

dipo. tathé iron pot

dibē sɔ wooden bowl

bē?ū taphé cotton cloth

?iswí thé ja pork curry

dó sɔba wall of boards

dó twa split-bamboo wall (examples all 5/2)

These are best analyzed as head+modifier, since the classifier used with the whole expression is the same as that used with the first constituent. For example, *dibē sɔ*

takes the Classifier *ba*, as in *dībē sɔ nā be* 'two wooden bowls', which as the same Classifier as is taken by *dībē* alone (*dībē nā be* 'two bowls'); *sɔ*, however, takes a different Classifier: *sɔ nā bō* 'two trees'. This can be taken to indicate that *dībē* is the head, and one of the features of the whole expression that it determines is the feature of association with *ba*.

There are examples with the opposite order, material+artifact, in which the head is a bound noun: in *sɔ ba* 'wooden flat-thing → board' *ba* may also be a classifier for thin flat objects (mats, pages, hats, paper money); it also occurs in *ce ba* 'paper', *tēú ba* 'fish scale', and *kū ba* 'dandruff' (for *kū* see 2.4). In *sɔba khu* 'board floor' and *twa khu* 'split-bamboo floor' *khu* is also a localizer meaning 'on (the upper surface of)'. It also occurs in many compound nouns, such as *plā khu* 'shoulder' (*plā* 'arm'), *ŋé khu* 'lower back', and *ʔí khu* 'land, world'.

2. generic-specific. Names of plants and animals often begin with the general term for the kind, such as *sɔ* for plants, *thu* for birds, *tē* for fish, and so on. It is not entirely clear how to analyze this type. On one hand, it is often said that the expression and its head exhibit an 'is-a' relation (a blackberry 'is-a' berry, and so on). The 'is-a' relation indicates that the general term is the head: a *tē lɔ* 'rock-fish' 'is-a' fish (*tē*), not a rock (*lɔ*). On the other hand the specific (second) member of the compound often has no identifiable use outside the compound: the second element of *tē phjǎ* 'a kind of large-headed fish' is surely not to be identified with the verb *phjǎ* 'take'. *tē phjǎ* must

be simply '*phjá*-fish'. For a different case, consider *tē thó* 'a kind of eel-like fish'. This *thó* might be identified with *so thó* 'oar', as if *tē thó* were 'oar-fish'; another way to look at it is to say that both *so* and *thó* function as disambiguating prefixes ('the *thó* that is wood', 'the *thó* that is a fish'). The same point can be made about *thu khwi* 'parrot' and *so khwi* 'vine'.

3. ethnic designations. These may either precede or follow the head, with the latter having the connotation 'characteristic of, X-style'. Thus *phré hmca* 'a Shan's clothes', meaning any clothes that a Shan might happen to have, versus *hmca phré* 'Shan-style clothes', which need not belong to a Shan. This may mean that ethnic designations have dual class membership, in both nouns and verbs: as nouns they precede the head, with possessive meaning, while as verbs they follow the head. This is certainly true of English ethnic/national designations, except that the two classes they belong to are nouns and adjectives: *a Russian, three Chinese* (nouns); *a German car, they are Russian* (adjectives). The same semantic distinction as that seen in Kayah can be demonstrated: *a German's car* could be a Pontiac, Toyota or a Renault as long as it belonged to a German, while a *German car* would have to be one manufactured in Germany.

7.1.2. Clf-headed NP's. Classifiers may be considered to be a special type of noun. This might be shown by saying that Clf's have the same feature matrix as other nouns, with an additional feature like

+  $l_Q - l_N$

which is meant to show that the Clf is a bound morpheme, being required to combine with a preceding Quantifier to form a lexical-level noun (this is a simplification: for expressions with the order Clf-Q; see 7.3.2).

Syntactically, ClfP's may function as Topic:

(2) nā mjō ?a ré ?é to

2 kind 3 good so NEG

Neither kind is very good (both kinds are not so good). 9/22

(3) təhe ?a khé rā kɪa || təhe ?a khé rā síne

one-CLF 3 shoot Rλ bow one-CLF 3 shoot Rλ bow

One group shot (with) guns, one group shot (with) bows. (226.7)

And possibly as an Object:

(4) pe ?íìò kā təcɜ rλ

1p plant COM one-CLF PTC

We planted one kind... (93.6)

although this might better be analyzed as having a zero Obj-x pronoun, the ClfP then being in the familiar Extent position ('we planted one kind *of it*').

ClfP's may be modified by either a preceding Demonstrative or a preceding nominalized clause. The Demonstratives are

?ā                      this

nλ                      that

Example of a clause (nominalized, but not overtly marked as such; see 9.1) modifying a following ClfP:

(5) *ʔa ʔe tēú tahe*

3 eat fish one-CLF(groups)

the ones who were eating fish (198.6)

A ClfP may not, however, be modified by a postposed clause<sup>2</sup>. Thus we get *ʔa ré to tahe* 'the group that's not good' (271.2), but *tahe (ʔa) ré to* would have to be a clause, 'one group is not good'.

ClfP's with modifier may be Obj-x:

(6) *bó pjà taloa tí ʔā tamē*

weave BEN medium-sized X-much this one-CLF

Weave a medium-sized one like this. (305.1)

[more strictly, 'medium-sized to this extent']

(6) is to be analyzed with *taloa tí ʔā tamē* as an NP, functioning as Obj-x, and consisting of a clause modifying a ClfP, the clause consisting of the verb *taloa* and the PP *tí ʔā*.

ClfP's with modifiers can also be the object of a Preposition, e.g. :

(7) *dś the takjā*

at go-up one-CLF(sides)

above, up there (see 9.1.2)

We have now seen Clf-headed NP's in several typical NP functions. To this we may add that a ClfP with modifier may be a referring expression, designating a person, place or thing. Unfortunately there is a problem in completely identifying Clf's with nouns and ClfP with NP, which may be summarized as follows:



<u>constituent</u>	<u>function</u>
lexical-headed NP	Topic, Sbj, Obj(x,1,2), obj of P
CIfP	Topic, Extent
CIfP with modifier	Topic, Obj-x, obj of P

The sticking point is the close association between CIfP and Extent position: only CIfP appears as Extent, and CIfP does not appear in the NP positions Sbj, Obj-x or object of preposition. This suggests that CIfP is a restricted type of NP, which becomes a more 'ordinary' NP when modified by a Demonstrative or clause. Further elucidation of these matters, perhaps including a statement in terms of  $\bar{X}$  theory (e.g., that CIfP is  $\bar{N}$  rather than NP), will have to await additional research.

7.1.3. Expanded NP's: NP+CIfP. The name is chosen arbitrarily, and there is some doubt as to the status of these expressions as unitary NP's. The uncertainty arises because in many cases the CIfP associated with a noun (or NP) is syntactically independent of it, as the Extent expression (3.3, 6.2.2). Of course it is possible to have Sbj-VC-Obj-CIfP, but even in that case I prefer to analyze the Object NP and the CIfP as separate constituents of the Predicate Phrase, since a Locative PP can always be inserted between the two. When the Obj is followed by a CIfP including a demonstrative, there is more of a flavor of direct modification, e.g.:

- (8) ...hé rá ʔa pò vè du lǎ nɛ tǎhe  
 say Rɛ 3 sibling big CMP that one-CLF(groups)  
 ... [he] said to those older siblings of his. (55.1)

My data does include examples of what seem to be Subject NP's consisting of lexical-headed NP plus ClfP, the ClfP seemingly modifying the NP directly:

| --- NP ----- | ClfP |

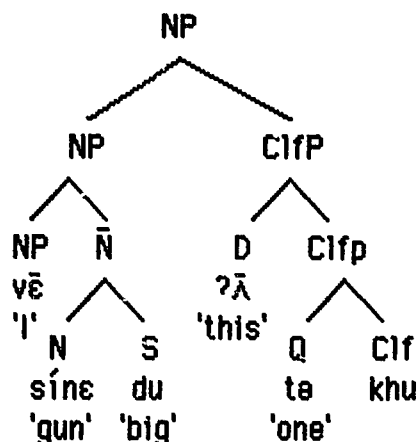
- (9) ʔa m̀̀phrè phrem̀̀ taphre ka dǎ khjā  
 3 old-person woman one-CLF come at back  
 One old woman came behind. (167.5)

Also the informant would accept expressions like *thuú ʔā tǎx ʔa vi* 'this kind of bird is good-tasting', but it is not clear to what extent these are artifacts of translation to or from Thai, which does allow noun and classifier phrase to be directly in construction with each other; compare the Kayah sentence, just given, with its Thai translation:

<u>Kayah</u>	thuú ʔā tǎx ʔa vi
	bird this one-clf 3 delicious
<u>Thai</u>	nók jàaŋ ní ʔarɔɔj
	bird clf this delicious

The one situation in which NP+ClfP is found with regularity is in response to elicitation concerning the proper Clf to be associated with a given noun; thus *dībē sɔ nā* *be* 'two wooden bowls', cited above. While this is at best a highly marked form of discourse, such expressions are probably legitimate nominal sentences (8.1). With these qualifications, I offer a tree structure for the (hypothetical) Expanded NP *vē sīnɛ du ʔā*

*takhu* 'this big gun of mine':



The basic structure [<sub>NP</sub> NP ClfP] might also subsume the Clf-headed NP: if we assume that the nominalized modifying clause is expressed by a rule NP →  $\bar{S}$ , the Clf-headed NP would then be [<sub>NP</sub> [<sub>NP</sub>  $\bar{S}$ ] ClfP].

#### 7.1.4. Localizers

Localizers form a closed class of nouns that cover much of the semantic territory of English prepositions. The Kayah construction [noun+localizer] is usually to be translated as an English [preposition+noun], but the localizers are not structurally equivalent to prepositions; they are not postpositions). In the Kayah construction, the localizer is the head and the noun is its modifier, thus *dě́ khrā́ kū́*, cited above, is 'at the bottle-gourd's inside → inside the bottle-gourd', *dě́ tǝ́ khu* is 'at the table's upper surface → on the table', and so on.

This follows Chao's analysis of the corresponding Chinese category, and borrows his term for it; cf. also Thompson's similar analysis of Vietnamese 'Relator Nouns'. Chao,

however, does not insist on the nounhood of his localizers: 'Though substantive in form, they are translatable into prepositions. For this reason, they are also called postpositions.' (p.621-2). The Kayah localizers are unambiguously nouns, more so perhaps than the Chinese class, since most of them may also form ordinary compound nouns, as can be seen in the following list. For each Localizer I give a general gloss, examples of locative use (A), and of occurrence in other compounds (B). Note that the general gloss applies only to use as Localizer (e.g. the first item *kū* could be given a general gloss 'hole' in its non-Localizer functions).

<i>kū</i>	inside	
A. <i>dó lē kū</i>		in the ravine
<i>dó pjā kū</i>		in the bag
B. <i>ʔa kū</i>		hole
<i>thā kū</i>		spring, well
<i>kadā kū</i>		window
<i>sō kū</i>		three (holes, springs, windows, etc.)
<i>khu</i>	on (the upper surface of), above	
A. <i>he khu</i>		on the ground
<i>hi khu (1)</i>		on the house
B. <i>plā khu</i>		shoulder
<i>twa khu</i>		floor
<i>hi khu (2)</i>		roof
<i>ʔikhu</i>		land, earth, world

khā	on top of, at the apex of
A. <i>dǎ cho khā</i>	on the (peak of) mountain
B. <i>lē khā</i>	headwaters
<i>pīthoə khā</i>	knife edge
kīō	outside
A. <i>dǎ hi kīō</i>	outside the house
B. <i>li kīō</i>	book cover
lo	on the non-horizontal surface of
A. <i>dǎ là lo</i>	on the cliffside
kīē	in among, in stg not construed as an aperture
A. <i>dǎ dō kīē</i>	in the village
<i>dǎ mi kīē</i>	in the forest
rōkīē	beside
A. <i>dǎ hi rōkīē</i>	beside the house
<i>dǎ rōkīē cáci</i>	on the left (side)
B. <i>rō</i>	side
<i>rō khwí</i>	rib
lē	bottom, base of, underneath
A. <i>dǎ hi lē</i>	under the house
<i>dǎ dō lē</i>	below (downhill from) the village
B. <i>plā lē</i>	armpit
<i>tēú kē lē</i>	gills

- chá** next to, near the base of
- A. *dé dō chá* near and downhill from the village  
*dé Phān chá* next to Pha'a
- B. *ke chá* mouth (of stream); foot (of tree)  
 (?) *ne nā chá* cheek
- ŋē~béseŋē** in front of
- A. *dé ŋē* in front  
*dé hi béseŋē* in front of the house (*bése* is 'face')
- khjā~békhjā** in back of, behind
- A. *dé hi békhjā* behind the house  
*ka dé khjā* come later
- B. not noun-related; cf. VPtc *khjā sé* 'back again, in response'; Bound Directional  
*kakhjā* 'backwards'
- ple~ple kū** in (the narrow space) between
- A. *dé dō ple* inside the wall (e.g. a lizard)  
*dé X né Y ple kū* between X and Y (e.g. between people standing in a row)
- B. *ʔa ple* a crack
- klē mē kū** in the middle of, between
- A. *dé dō klē mē kū* between the villages
- B. *taklē mē* one half
- thū** on the edge of
- A. *dé dō thū* on the edge of the village
- B. *thō thū* river bank
- ŋū thé** uphill from
- ŋū lē** downhill from

<b>takjā</b>	in the direction of
A. <i>dš phrè dš takjā</i>	towards the Shan village
<i>dš tē takjā ē</i>	in which direction?
B. <i>plā nā kjā</i>	two arms

## 7.2 PP.

Prepositions are bound morphemes that form a PP with a following NP (which may be a nominalized clause). There is one exception to this: *chā* takes only clauses or zero, plus the SPTc *pā* 'irrealis' (required with zero, optional otherwise)

The Prepositions fall into three groups:

### a) Locative

<i>dš</i>	at, when (past); distal, unmarked for evidentiality
<i>mú</i>	" known by inference or hearsay, not in sight
<i>bš</i>	" proximal, in sight

### b) Extensive

<i>bá</i>	as much as, _ much
<i>tí</i>	as big as, _ big
<i>tš~thš</i>	as long as, _ long

### c) Miscellaneous

<i>chá</i>	when (future)
<i>phú~hú</i>	like, as
<i>cé~cəə</i>	the __ part, the ones who __

a) Locative. The difference among these three is that of evidentiality, i.e. the speaker's basis for knowing the truth of the proposition he utters. *mú* and *bš* have

positive evidential connotations, as indicated; *dě* seems to have less, and may be considered unmarked. It is by far the most common preposition. All three may be used to indicate time, but it must be past time; future time requires *chá*. *dě* at times is better translated 'as for':

- (10) *dě* ?a mē nλ rλ ... *dě* ?a phē nλ ...  
 3 wife Nλ PTC 3 father Nλ

The wife, for her part, ... the father, for his part, ... (53.2)

b) Extensive. These usually precede lexical NP's (not nominalized clauses); e.g.

- (11) ?a béseplɔ ?o t̃i chā dλ  
 3 eye have chicken egg

He had eyes as big as chicken eggs. (95.5)

They may form interrogative sentences in conjunction with the SPtc *tē*, as:

- (12) ?a thū t̃x tē ɔ  
 3 long PTC HUH

How long is it, huh? (10/18)

And *bá* CLF *tē* is the standard expression for 'how many?'.  
 .

c) Miscellaneous: a note on *chá*

*chá* indicates future time. Its commonest uses are 1) *chá pā* meaning 'soon'; 2) with short time expressions, most of which can be analyzed as clauses, e.g.



chá mō hé pā	this evening ( <i>mō</i> 'sun', <i>hé</i> 'late, be evening')
chá no pā	later on (for <i>no</i> as a verb cf. 4.2.4)
chá lē the tan̄ pā	on the coming first day of the waxing moon ( <i>lē</i> 'moon', <i>the</i> 'ascend', <i>tan̄</i> 'one day')
chá pā ro pā	tomorrow ( <i>pā ro</i> 'tomorrow', analysis uncertain)

Finally, note the possible relation of *chá* to the homophonous Localizer meaning 'place nearby, at the base of'.

Besides filling the post-verbal slot that I have referred to as Obl (Ch. 6), PP's are also quite common as Topics. This is especially true of time-when expressions referring to the past, which usually appear in that position--for instance, the standard story-telling opener *dá njā nā* 'long ago' (*njā* 'be a long time'). Cf. also:

- (13) dá | vē | jì | bó | nā | vē | bá | hē | chā | ʔa | ri | mē  
 at 1s thresh rice N<sub>λ</sub> 1s divine go chicken 3 good PTC  
 When I was threshing, I divined<sup>3</sup> about going [to work]; [the indication] *was*  
 auspicious! (183.6)

Future-time expressions are introduced by *chá* and are normally in the Obl position.

There are also examples that can be interpreted as PP's functioning as Object(-x):

- (14) ʔa | lā | sínē | rá | phé | céə | rā | má | rá | nā | phé  
 3 PTC understand R<sub>λ</sub> only part-that write down R<sub>λ</sub> N<sub>λ</sub> only  
 They know only what they write down. (302.3)

- (15) ʔa|nō hē dā rā |dē ʔa rō tēphre i nā  
 3 order go give Rλ at other one-CLF Nλ  
 He got another person to go give it. (100.5)

Kayah PP's are thus not very different from NP's. Most of the meanings associated with the prepositions of English and other languages are not borne by prepositions in Kayah, namely:

directional/configurational

into	nō (verb)
under	iē (Localizer)
from	ʔo X V (V̄ series)

temporal

after	clause sequence, 'having done X, then did Y'
-------	----------------------------------------------

case- or semantic-role marking

for	pē (Verb Particle)
of	(concatenation of NP's)
to (Recipient)	Obj-1 (structural position)

The 'locative' Prepositions especially can be thought of as evidential markers of NP's, which only incidentally tend to be locative in meaning. In this view, it may be the Localizer constituent of a locative expression, rather than the Preposition, that assigns or marks the PP/Oblique grammatical relation.

### 7.3 ClfP

Since we have already dealt with the syntactic characteristics of the ClfP as a clause constituent in 3.3, 6.2.2 and 7.1, this section will be confined to discussing the lexical categories Classifier and Quantifier and how they combine in the ClfP.

7.3.1. Classifiers. The choice of classifier depends to an certain extent on the noun that is being counted. Nouns are marked lexically for a certain classifier or classifiers:

Noun	Classifier
təple 'cabbage'	to
ʔa ple 'crack', kədā 'door'	kū
thuú 'bird', bē lǝ 'cup'	be
sɔ 'plant'	mò (smaller plants)
"	bō (trees)

Although there is a lexical association between the noun and classifier, in the sentence there is (usually) no direct syntactic relation between the NP and the ClfP, as we have noted (3.3, 6.2.2, 7.1.3).

Several types of Classifiers can be distinguished, such as

Unit Clf. counts common nouns, associated with shape

Measure. often also the name of a container

Time. a) units of time such as nā 'day', na 'year'; b) Clf's counting instances of an action, including the general time Clf phó 'a time'. There are also verbs that double as

Clif's, some of which denote instances of an action, as:

	<u>as verb</u>	<u>as Clif</u>
mū	hit	strokes, blows
khē	to step	steps

But others denote instances of an action only in the sense of entities resulting from the action:

	<u>as verb</u>	<u>as Clif</u>
tuū	sever	stumps
khri	be in bits	bits, shards, hills
co	wrap	packages
ʔikhu	to wind	-
khu	-	spools of thread, guns

Regarding the last-mentioned, a spool of thread can be understood to be the result of an act of winding; the relation to guns must be metaphoric, perhaps relating specifically to revolvers.

This leaves a number of Classifiers that do not fit the into the types easily, such as *mjō*, *cɔ* both 'kind, sort, type'.

There is an extensive area of overlap between classifiers and common Nouns; or in other words, many nouns may act as their own classifiers. Thus *dō* 'village', *sō dō* 'three villages', *mi* 'name', *ʔimi* 'one name'. The common-Noun version of some of these is Bound, and so must be accompanied by some other morpheme; if there is nothing more specific in meaning the other morpheme may be *ʔə*, which may then be considered a

## derivational prefix :

pɔɔ	classifier for small round things
síne pɔɔ	bullet (síne 'gun')
?apɔɔ	a seed
kū	classifier for holes (also Localizer 'the inside of')
là kū	cave (là 'cliff')
?akū	a hole
bō	classifier for lengths
taré bō	candle (taré 'wax')
dí klɛ bō	sugarcane
té ta bō	pencil
plē	classifier for ears of grain
bó plē	ear of rice
tachē plē	elephant tusk
?a plē	ear of (any) grain

A different sort of overlap is seen in *hɪ* 'house', which is also a classifier for humans meaning 'household'. As a common noun, however, its associated classifier is *mē*.

List of some common Classifiersunit Clf's

mò	smaller plants
dō	animals
phre~sí	humans (see 7.3.2.2)
mē	larger semi regular shapes: houses, drums, heads, stomachs, livers, larger fruits, hammers, hills, wheels, etc.
bō	lengths: ropes, snakes, intestines, worms, tongues, trees, vines, lizards, cigarettes, sprouts etc.
ko	the general classifier; also: beds, sticks, ridges, stoppers, splinters, pincers, noses, certain bones
plɔ	small round things: smaller fruits, stars, buttons, eggs, cakes of soap, scabs, grains of sand, etc. (cf. Thai lûuk)
be	flat-faced and winged things: birds, leaves, fish, dishes, cups, spiders, sickles, drinking glasses, axes, saws, moon, sun, doors, hoes, boats, teeth, paddles, etc. etc. (cf. Thai baj)
ba	sheet-like things: mats, paper objects (money, cards, pages, pictures), cloth, mosquito nets, umbrellas, hats, the earth, etc.
pw	clothing: shirts, towels, shawls

Measure Clf's

thē	span (the distance between outspread thumb and middle finger)
plè	cubit (from elbow to fingertip)
kfi	fathom (distance of spread arms)
cw	handful

khwē the volume of a packbasket

de eight *khwē*

and the name of any basket (*khri*, *dī phó*, *phó mó*, *su du*, etc.)

**7.3.2. Quantifiers.** Quantifiers include the numerals and a few additional morphemes.

pwā every

chī whole, the entire

bá how many? (but *bá* shares features with certain 'extensive'  
Prepositions; 7.2)

E.g. *pwā phre* 'every person', *pwā cx* 'every sort', *chī hi* 'the whole household'.

The remainder of this section will be devoted to a description of the numeral system  
and how numerals combine with classifiers.

Counting in Kayah.

The basic numerals are

1	tə- (prefix)
2	nā
3	sō
4	lwī
5	ŋē
6	sō swá
7	sō swá tə-
8	lwī swá
9	lwī swá tə-
10	chá (basic form) ~ chā (in 20-90)
100	(tə)je
1000	(tə)rí
10,000	(tə)só

The last three forms always are accompanied by a 'multiplying' digit: *təje* 'one hundred', *nāje* 'two hundred', and so on (the proper translation for *je* or *rí* on its own would be 'hundred' and 'thousand').

The Kayah numerals are emphatically bound morphemes, to the extent that if a speaker is asked to recite them, he or she will usually recite numeral-classifier constructions, using the classifier for small round objects *ɲɔ*, thus: *təɲɔ*, *nāɲɔ*, *sōɲɔ*, *lwīɲɔ* and so on (does this indicate that numbers in the abstract are thought of as small round lumps, like counters?)



The morpheme *swá* also appears in *bé swá rá* 'be companions with' and *khō bé swá* 'friend'. It is evidently verb-related, as shown by its occurrence in the verb 'be companions with' ('friend' presumably has the verb modifying a noun *khō*)<sup>4</sup>, and can be glossed 'to double, make a pair'.

Thus 'six' is literally 'three doubled', 'seven' is 'three doubled plus one', and so on. These 'analytic' numerals appear to be a recent innovation: other Kayah dialects preserve the monomorphemic forms of 6-9. Below are listed these numerals in one West Kayah dialect, along with the forms they would probably have if preserved in East Kayah

	West	East (hypothetical)
6	ꠘꠦꠄ	*có
7	núo~dá	*nwá~dá
8	ꠘꠦꠄ	*swá
9	núá	*nó

The numbers 11-19 (the teens) are formed by *chá* followed by the units numeral: *chá lwī* '14', *chá sō swá* '16', and so on. The numbers above 19 are formed by *chá* plus a 'multiplier' unit. The relative ordering of numeral and classifier varies, and can be described by a series of rules.

**Rule 1:** Numerals that end with *swá*, *chá~chá* or *je* follow their classifier; all others precede the classifier<sup>5</sup>. Thus the recitation above would continue *ꠘꠦ ꠘꠦ swá*, *ꠘꠦ ꠘꠦ swá ꠘꠦ*, *ꠘꠦ ꠘꠦ swá ꠘꠦ*, *ꠘꠦ ꠘꠦ swá ꠘꠦ*, *ꠘꠦ ꠘꠦ swá ꠘꠦ*.

**Rule 2:** *chā* in the numerals 20–90 acts like a classifier, preceding multipliers that end in *swá*, thus:

20 *nā chā*

30 *sō chā*

60 *chā sō swá*

70 *sō swá tēchā*

The function of the tonal allomorphy of 'ten' can now be seen: those tens whose multiplier follows 'ten' would otherwise be indistinguishable from the teens, which all consist of 'ten'+ 'unit'. Minimal pairs are

*chá sō swá*      16                      *chá lwī swá*      18

*chā sō swá*      60                      *chā lwī swá*      80

*chá sō swá tē-*      17

*chā sō swá tē-*      61

The significance of the mid tone of 'ten' in 20–90 can thus be seen as that of signaling a multiplicative (rather than additive) relationship with the adjoining numeral.

For example, the difference between '17' and '61' can be symbolized algebraically as:

$$10 \cdot (3 \times 2) + 1 = 17$$

$$(10 \times (3 \times 2)) + 1 = 61$$

Notice that while expressions like *sō swá tē-* are probably best considered compounds<sup>6</sup>, *tē-* 'one' is phonologically dependent on an element outside the expression, the following morpheme *chā* 'ten', deriving its vowel color from it. Although I have

treated *tə-* as a prefix, no other prefix has this sort of dual dependence; recall that *tə-* also stands out as the only prefix having full productivity.

Rule 3. When the classifier refers to humans, it has the following three allomorphs:

phre	following <i>pwā</i> 'every', and numerals ending in <i>tə-</i>
zero	when the numeral ends in <i>swá</i> or <i>chā</i>
sí	otherwise, <i>preceding</i> the numeral

Thus:

təphre	1 person
sí nā	2 people
sí sō	3 people
sí lwī	4 people
sí ŋē	5 people
sō swá	6 people
sō swá təphre	7 people
lwī swá	8 people
lwī swá təphre	9 people

This may be a factor in the required use of the classifier *plɔ* in 'abstract' counting as mentioned above: if no classifier were supplied, *sō swá* and *lwī swá* would not mean 'six' and 'eight' in the abstract, but 'six people' and 'eight people'.

Apart from the cases covered by Rule 3, the placement of the classifier follows Rule 1 regardless of the internal structure of the numeral. Thus even though the numeral 'ten' precedes its multiplier in 'sixty' *chā sō swá* and follows it in 'seventy' *sō swá tachiā*, the classifier precedes both, since both end in a numeral that requires the classifier to

precede. To illustrate (classifier underlined):

20	<u>pɿɿ</u> nā chā
21	nā chā tə <u>pɿɿ</u>
22	nā chā nā <u>pɿɿ</u>
26	<u>pɿɿ</u> nā chā sō swá
27	nā chā sō swá tə <u>pɿɿ</u>
30	<u>pɿɿ</u> sō chā
31	sō chā tə <u>pɿɿ</u>
60	<u>pɿɿ</u> chā sō swá
70	<u>pɿɿ</u> sō swá təchā

Further complications arise above one hundred. The classifier is often repeated, once with the hundreds and a second time with the tens, as in the following examples with *na* 'year':

<u>na</u> taje nā <u>na</u>	102 years
<u>na</u> taje chā tə <u>na</u>	111 years

The first occurrence of the classifier may be omitted; thus '111 years' may also be *taje chā tana*. The classifier follows Rule 1 even to the extent of appearing 'inside' the construction:

( <u>na</u> ) taje <u>na</u> lwī swá	108 years
( <u>na</u> ) taje <u>na</u> chā	110 years

This suggests that the construction is not unitary, but perhaps coordinate, as if it were to be glossed e.g. 'one hundred years and ten years'. A similar impression is given

by the possible occurrence of *í*, presumably the *í*-class Particle 'new situation', after

the hundreds expression:

taje *í* na sō swá                      106 years

na nā je *í* ηē chā sō swá              256 years

But *í* appears to be optional:

(16) na sō je sō swá tachā lwī swá tana

year 3 100 3 double 1 10 4 double 1 year → 379 years (3/21)

Here *je* is not followed by *í*.

## 8. Sentence Types and Sentence Particles

### 8.1 Sentence Types.

This study has been mainly devoted to what may be called intra-clausal grammar, particularly the grammar of the simple clause, i.e. without embedding of clause within clause. In this chapter I will look briefly beyond the simple clause, by situating the simple autonomous clause within a more inclusive inventory of sentence and clause types. I will then go on to describe the class of Sentence Particles, the only clause constituent not yet considered. This also seems the best place for a discussion of questions and question words, although only a few of the question words are Sentence Particles

Let us begin with a set of definitions, some repeated:

A *verb* is any morpheme that can stand on its own in construction with *lāī*, *ī*, *pā*, or some other member of the *ra*-class Verb Particles.

A *clause* is any construction that both (a) contains a verb and (b) can be terminated by the Sentence Particles *to* 'negative' or *pā* 'irrealis'.

A *sentence* is any construction that can stand on its own as an utterance bearing an illocutionary force; it may consist either of one or more clauses, or of an NP. Let us call the former a *verbal sentence* and the latter a *nominal sentence*; both types may end in a Sentence Particle (although only a subset of the SPTc's may terminate a nominal sentence). A verbal sentence may also be classified as an *autonomous clause*; clauses that cannot stand on their own (do not qualify as sentences) are termed *non-autonomous*

clauses. Examples:

verbal sentence = autonomous clause

(1) nō nól hóhé ! to ē ɔ

enter at-all school NEG QUES HUH

Aren't you going to school, hey? (130.4)

nominal sentence

(2) thwá ke

cat PRH

(on hearing a noise:) Maybe it was the cat. (in conversation)

non-autonomous clauses

(a) modifying a preceding noun

(3) kajē [ ré to ]

person good NEG

a bad person

(b) modifying a following ClfP

(4) ʔa ʔe tēú tahe

3 eat fish one-CLF(groups)

the ones who were eating fish (198.6)

(c) as object of a preposition

(5) ʔa phā nʌ|ʔo| dʒ hi thw̄| hú [pe ʔo ʔā nʌ]

3 grandmother Nʌ exist at house edge as 1p exist this Nʌ

His grandmother lived at the edge of the village, like we live here. (204.4)

The difference between preposed and postposed *attributive clauses* (AC) parallels the contrast between nouns (which typically precede what they modify) and verbs (which typically follow what they modify); for more on this point see 9.2. The preposed AC's

end in *m*; since clauses ending with *m* also resemble nouns in being capable of acting as object of a preposition, I interpret *m* as a nominalizer (see also 8.3).

We therefore have, in addition to [ $\pm$  autonomous], another possible feature for classifying clauses: [ $\pm$  nominalized]. We have seen examples of [-autonomous +nominalized] (4 above), [-autonomous -nominalized] (3), and [+autonomous -nominalized] (1). It turns out that the fourth possibility also exists: there are what appear to be autonomous clauses ending with *m*, which might be analyzed as autonomous nominalized clauses. Autonomous nominalized clauses might further be taken to be instances of nominal sentences. Since they are extremely common (they seem indeed to be a widespread Tibeto-Burman trait), I have instead generally treated these autonomous nominalized clauses as verbal sentences with *m* functioning, like the (other) Sentence Particles, to mark illocutionary force. See, however, the discussion of *m* (8.3 below).

It will be seen in the following section that only a small number of Sentence Particles can occur in non-autonomous clauses; the large number that cannot include most of the illocutionary-force markers. This fact reinforces the definition of autonomous clause (verbal sentence) in terms of ability to bear illocutionary force: since an embedded clause generally cannot have any illocutionary force distinct from that of the main clause, we would not expect illocutionary-force markers to occur in it.



## 8.2. Sentence Particles

The Sentence Particles (SPtc), as the name indicates, terminate the clause; as suggested above (6.2), they are adjoined to the clause in a configuration like ...]<sub>S</sub> Ptc]<sub>S</sub> Ptc]<sub>S</sub>. There are some 18 SPtc's known; they can be sub-classified on the basis of two formal properties. The first is ability to reduplicate (cf. 2.4 above); the second is occurrence in non-autonomous clauses.

Formal properties. Certain SPtc's can reduplicate, as in the following:

- (6) dǎ cwá vĕ to || vĕ cwá to to  
 let go 1s NEG 1s go NEG NEG

[if you] won't let me go, I won't go, then. (0.5)

- (7) m sí?ichĕ ké rò he he  
 afraid AMB cold LEST LEST

Mm, I'm afraid it'll be cold, too. [e.g. in addition to raining] (2/24)

- (8) thĕ phra kĀ ke ke  
 pig to-sound COM PRH PRH

It might also be a pig making noise. (2/24)

Others cannot:

- (9a) vĕ cwá kĀ ní | \*vĕ cwá kĀ ní ní  
 1s go COM EMP

A: I'm going along! B: I'm going along too!

compare:

- (9b) vĕ cwá kĀ ní || vĕ cwá kĀ kĀ ní  
 (same meaning as 9a)

For a similar example:

(10a) ma Pímò te mē | \*ma Kāmè te mē mē

be-so (name) 's PTC

It's Pimo's, and it's also Kameh's!

(10b) ma Pímò te mē ||ma Kāmè te te mē

(same meaning)

The second formal property is that of occurrence in non-autonomous clauses, found only in *to* 'negative' and *pã* 'irrealis, future'. Examples with *to*:

(11) dī [ vā lāi to ] ʔo pa dɛ dī ʔo kũ

rice cooked yet NEG have DUR at pot inside

The rice that is not yet cooked is in the pot. (11/24)

(12) ma | kajē [ mò ʔo to phē ʔo to ]

be-so person mother have NEG father have NEG

[they] are people without a mother or father. (267.2)

and with *pã*:

(13) ma | dɛ [vē | dá pè kã | vē pò | pã nã] tãbe

be-so at 1s give TRN COM 1s YS IRR Nã one-CLF

It's the one [a tool] that I'm going to give to my brother. (11/24)

(14) pãnè [ cwi ʔĩthá pã ] ʔo tōútē

buffalo pull plough IRR have where

Where's the buffalo that will pull the plough? (11/24)<sup>1</sup>

Since these two formal properties do not coincide completely, they delineate three subclasses of SPtc's:

	<u>subclasses</u>		
	A	B	C
reduplicates	+	+	-
occurs in non-autonomous clause	+	-	-

Some at least of the SPtc's can terminate nominal sentences --which is indeed the evidence compelling us to recognize those constructions as sentences. Thus the class B SPtc *ke* in (2) above; also the negative in *ʔū rə mǝ to* '[it was] not other people's country'

There is a rough correlation of semantic values with the three subclasses. A relates to what can be called polarity: positive/negative, realis/irrealis; C consists entirely of markers of illocutionary force, while B is a mixture; e.g. the B class SPtc *he* 'possible undesirable event' combines polarity ('possible') with expression of the speaker's attitude ('undesirable').

By 'illocutionary force' I mean the 'interpersonal' level of meaning (the term is from the works of M.A.K. Halliday) in general, which can be loosely defined as concerned with the speaker's attitude towards his/her utterance, especially what the speaker expects the hearer to do with the propositional content. For example, the hearer may be expected to

believe the proposition to be true (assertion), carry out the action it describes (directive imperative), supply information marked in it as unknown (question), and so on. Note that the meanings of Kayah illocutionary force particles do not correspond directly to a theoretical system of speech-act types like that developed by Searle. There are many markers for a single speech act type (e.g. commands), with various shades of meaning; conversely there is no SPTc marking the Commissive type, which is rather marked by the Quasi-modal Verb *kha* (see 4.2.5). Also, there are SPTc's marking exclamation, which can only be said to be a speech act type if one says that what the hearer is expected to do is understand that the speaker is surprised.

The illocutionary force-marking function is typical of morphemes of this type, i.e. sentence-final bound morphemes, often unstressed. This is true at least for Southeast Asian languages, in which they are often known as Sentence Particles or Final Particles (cf. also Okell's Verb-Clause Markers). It is often said that the lexical function of tone in these languages limits the possibilities of exploitation of phrase and sentence intonation for expressing illocutionary force, hence such meanings must be encoded in this class of lexical items. What is certain is that these illocutionary force markers are invariably among those aspects of the language that are the most difficult to capture and explain when the investigator is a speaker of English or some other language that does not encode these meanings in individual morphemes. Research on the Kayah SPTc's is still in a preliminary stage, and the descriptions below are in many cases simply tags for items

whose meanings have yet to be thoroughly studied. This is the reason for the many particles with the same tag (e.g. 'urging'); they are probably not synonymous, but the distinctions between them have not been uncovered.

Subclass A.

*tə*

The negative. This particle may have some relationship to the phonologically aberrant *təə* 'only' (cf. 1.4). *təə* occurs only after Extent expressions (i.e. post-verbal ClfP's), while *tə* does so marginally at best. Consider the following contrastive sets:

(15a) *rū* ?ə ηē cwè təə

money have 500

[I] have only 500 [Baht]. (2/27)

(15b) *rū* ηē cwè ?ə tə

money 500 have

[I] don't have 500 [Baht]. (2/27)

(16a) ?a khé be thuú sō dō təə

3 shoot strike bird three CLF

He shot only three birds. (2/27)

(16b) nʌ sō dō ?a khé be tə

that three CLF 3 shoot strike

He didn't shoot [shot at and missed] those three. (2/27)

In the (a) sentences, substitution of *tə* 'not' is either unacceptable or not preferred.

Given these facts, a case could be made that 'not' and 'only' are two translations of a single morpheme (assuming that the phonological difference could also be accounted for).

The latter translation would hold when the clause contains an Extent expression; the former otherwise. Note that the difference in (15–16) above may be the presence of a ClfP inside the clause: in the (b) sentences the ClfP is in the extra-clausal Topic position (clearly so in (16b), which also has a Subject; arguably also in (15b), but with zero Subject). In support of the semantic kinship of 'not' and 'only', one could cite French *ne...que* 'only', in which the first element is (or is homophonous with) the negative; and perhaps also nonstandard English *ain't got but [quantity]*. Note also that *nothing but* is a near equivalent to *only*.

Examples of reduplication of the negative have been given above; the morpheme in the sense 'only' may also reduplicate:

- (17) dī pɔ ʔo tamē too || pē ʔo phé tamē too too  
 pot have one-CLF      bottle have only one-CLF  
 There's only one pot; and only one bottle as well. (2/27)

*pā*

Irrealis, future hypothetical, upcoming. Very frequent in clauses introduced by the Preposition *chá*, which indicates near-future time-when expressions. E.g.

- |                    |                                                                                        |
|--------------------|----------------------------------------------------------------------------------------|
| chá m̀ hé pā       | this [coming] evening (m̀ 'sun'+hé 'evening')                                          |
| chá lē the tan̄ pā | the [coming] first day of the waxing moon<br>(lē 'moon', the 'ascend', tan̄ 'one day') |
| cwá chá pā         | will go pretty soon                                                                    |

*pā* contrasts with *ke* (subclass B), which expresses uncertainty about events that \

are (or may have been) completed, usually in the past. The Kayah equivalent of 'maybe, possibly, might' is not a modal auxiliary but a use of *pā*, as in *hē pā ē, to ē* literally 'may come, may not', corresponding to *may come, maybe will come*, and very often followed by *sīŋe to* 'don't know'.

Subclass A may also be known as the Clause Particles: since these SPtc's are the only ones that can occur in all sorts of clause, it has been useful to single them out in discussing intraclausal syntax. But from the perspective of the sentence *to* and *pā* are simply a special subtype of a larger class.

Subclass B.

*ke*

past or perfective irrealis. Examples:

ʔiphri ka ke [they] may have bought [it].

(18) pe lā síplɔ nō lū to nʌ ke

1p PTC understand 3OBV NEG Nʌ

'We just don't understand them. (45.5) (note the sequence of three SPtc's)

*he*

lest, possible undesirable event. Examples:

sí de khrā, síʔichē ʔa ʔ hē            [I] want to put it out to dry; I'm afraid it may  
mildew (cf. I want to put it out do dry *lest* it  
mildew)

sī súplī pè ʔa, ʔa kāsé và lū hē ní    You wash him, or he'll get itchy again, now! (lest  
he get itchy)    (249.4)

*me*

don't, negative imperative. Examples:

mé pathe me ní                            Don't look up, now!

thé ma kóno me me                      Don't play with the needle either.

(note reduplication)

*lē*

and what about ...? Seems to occur only in nominal sentences. Examples:

Phrèλ cwá nó to |.. ʔa mò lē ɔ        A: Phre'a didn't go. B: What about his mother?  
(146.2)

vē λ lē                                        Where's my grandchild?

Context: in a narrative, the grandmother speaking to people who had gone fishing with her grandchild and were returning without him/her.



Subclass C. The remaining SPtc's will be grouped according to general types of illocutionary force; some of the SPtc's already discussed that count as illocutionary force markers will be repeated, in parentheses, under the appropriate heading.

### Interrogative

ē yes-no question

?ō tanā ?o kō ē Are you free today?

also occurs in the idiom

dš tē taCLF ē 'which one?'

pē discontinuous component of ?ū ... pē 'who?'

(lē)

ɔ prompt-question; can follow other interrogative indicators, or may turn a statement into a question, somewhat like English *huh*

ne cwá ɔ you're going, huh?

There is also a form *ʔē* that appears at the end of questions, but it is not a SPtc; see 8.4 below.

### Imperative

kɔ urging, offering for consideration, suggest that you do this/agree with this

tε~tεē urging, let's

dūú urging, you go ahead and

pó urging, let's, want to ...?

ke urging

(me)

ní (see 'Assertive' below)

(cf. also Verb Particle mā, imperative)

Assertive

nλ neutral assertion

tako λ concessive assertion, nevertheless, still

(19) ʔū bé b́é tē'ʔū síjw k̄ nλ tako Δ

3 endowed at what 3 want COM Nλ

However rich they are, they still want some. (269.6)

mó concessive: sure, but ...

lé counter-assertion

(20) pē já pé ĺé sé ʔū lé

1p go-and dumb more back-again 3

[we should have been smarter,] but we're *dumber* than them! (109.3)

mē counter-assertion, contrary to some other statement or fact; possibly differs from the preceding by connoting irrealis; cf.

(21) ʔa me s̄ā á lū mē

3 do die NS mutually

They would have killed each other [if I hadn't stopped them]. (241.2)

ní strong assertion or imperative: be sure and pay attention to what I say

lā~lá exclamation

(cf. also Verb Particle wā: medium-strong assertion, sure it's true that...)

Multiple Sentence Particles. Below is a rough indication of the co-occurrence

possibilities of the SPTc's:

to	na	me	ē	teē
pā		ke	mē	o
		he		ní
			*	

\*among the SPtc's that can appear in this position are *lē, pō, dūū, kɔ, mō, lē* and possibly others.

This chart treats *na* as just another SPtc; for its special characteristics see below.

### 8.3. Unique and problematic particles

#### 1. *ke* 'if'

In addition to the SPtc described above, there is also a morpheme *ke* that appears between Subject and VC, with the meaning 'if':

mò ke ʔo to ma ...                      if they have no mother ... (11.3)

pe ke bwí rÉ rA ...                      if our luck is good ... (30.6)

Clearly, conditionality and irrealis are closely related. This pre-VC *ke* might be classified as a *khwe*-class Verb Particle (4.2), but it freely occurs before the *khwe*-class VPtc's; the (other)members of that class co-occur only in circumscribed conditions. It is probably better to set up a one-member class for this *ke*, with notice taken of its possible relation to the SPtc. If that relation is accepted, *ke* becomes analogous to form classes found in several other languages of the area. For Burmese Oke11 describes a type of 'subordinate marker' that is 'suffixed to either [verbs or nouns]'; other classes of this

sort are the 'movable particles' of Vietnamese (Thompson) and the 'unrestricted particles' of Lahu (Matisoff).

2. *m*, assertion marker, nominalizer/complementizer, demonstrative.

The syllable *m* is variegated and elusive in its function, both syntactically and semantically. A first division can be made between the distal demonstrative *m* 'that' and what may be called the particle(s) *m*. Two particles can be distinguished, both of them extremely common: a topic-marker and what I have been calling the nominalizing Sentence Particle.

The topic marker follows NP's:

Topic marker

(22) Míʌ ɲʌ ma vĕ phĕ mĕ

(name) Nʌ be-so 1s father PTC

Mi'a is my father! (266.1)

(23) phrè jwi ɲʌ kwi ʔe kè

Thai request for-use country

The Thais asked for some land. (202.6)

(24) ʔa phú ʌ ɲʌ nō ka ló lū đǎ sí nō ʌ

3 child grandchild command go inter 3OBJV at West

His descendants were told to go bury him in the West. (201.5)

The nominalizing SPTc occurs with both autonomous and non-autonomous clauses:

Nominalizer

a. with autonomous clauses

(25) ʔa ka ʔe lɔ̃ lū ɲΔ

3 come eat use-up 3OBV

They came and ate them all up. (206.2)

(26) the me ʔichē tò | ʔa mīú ʔa mwī | dʒ hi dʒ kū | ɲΔ

go-up do fear arrive 3 (name) 3 namesake at house inside

[He] went up and frightened Mi'u's namesake in the house. (242.6)

b. with non-autonomous clauses

(27) dʒ pè kʌ ʔa təkʰʌ ja ɲΔ taphre

give TRN COM 3 muntjac meat one-clf

the person that you gave muntjac meat to

(28) dʒ vē jì bó ɲΔ

at 1s thresh rice Nʌ

when I was threshing (183.6)

(29) bʒ ʔa phʌ ʔo rʌ ɲΔ

at 3 grandmother exist Rʌ Nʌ

where the grandmother lived (209.2)

At the present state of knowledge I do not want to insist too much on the function of *ɲ* being nominalization. Concerning non-autonomous clauses with *ɲ*, their nominal nature resides largely in the two characteristics of preceding modified ClfP's and of acting as objects of Prepositions, and *ɲ* seems to be optional in both cases. There are some examples that show what looks like nominalization in the classical sense of 'that which S':

(30) [ chá lū nɿ ] ma ʔū pē

fight each-other Nɿ be-so who

Who was it who was fighting? (227.4)

(31) ʔa chá nɿ thwā rū ʔa kōdó nɿ thwā thē

3 clear Nɿ become silver 3 muddy Nɿ become gold

The clear one [container of water] turned to silver, the muddy one turned to gold. (40.5)

(32) [ ʔəə thē ʔiswá tíí nɿ ] ma ʔā təphre kə

dght.-in-law go-up teach constantly Nɿ be-so this one-CLF PRMPT

This is the one you go teach all the time, huh? (220.1)

But we would expect to find nominalized clauses in other functions as well, such as object (of verbs like *know*, *believe*, *see*) or subject (e.g. *That he arrived late annoys me*, *For John to arrive late would be surprising*). Not enough is known about clauses as arguments of verbs (see 9.3), but they do not seem to require *na*:

(33) nìhō ʔa ʔé mò ʔé phē to

hear 3 call mother call father NEG

[we] don't hear him call his mother and father. (249.6)

Concerning autonomous clauses with *na*, nounhood is even harder to demonstrate. As indicated previously, we could place them in the category of nominal sentence, needed independently for lexically-headed NP's that stand as sentences. The semantic connection between nominalization and assertion is well described by Matisoff in relation to the Lahu particle *ve*, which resembles *na* in many respects:

From this point of view, every verb occurring in the environment  $\text{---} + \text{ve} + \{, / * / P \}$  would be considered 'objectified' or 'reified'. Its verbality is set up as a neutral fact, endowed with a reality like that inhering in physical objects.

Matisoff 1973, 362

Finally, it should be pointed out that  $\mathcal{M}$  has at least some of the properties of a complementizer. If  $\mathcal{M}$  is a complementizer, clauses with  $\mathcal{M}$  are to be considered to be  $\bar{S}$ , and what I have been calling 'nominalization' may be better seen as the properties that  $\bar{S}$  has in common with NP. As Rothstein (1985) points out, NP and  $\bar{S}$  group together in being the typical argument constituents, while the other maximal projections, VP, AP and PP, are the typical predicates. To put it another way, VP, AP and PP require external arguments (in Williams' sense), but NP and  $\bar{S}$  do not (although they are not barred from having them either). As before, however, we are left with the task of explaining the function of  $\mathcal{M}$  in autonomous clauses; or in these new terms, of explaining the existence and prevalence of independent  $\bar{S}$ 's.

#### 8.4. Excursus: interrogative (and indefinite) morphemes.

The rubric 'interrogative morphemes' includes a Sentence Particle, a bound morpheme best analyzed as a noun, and one discontinuous constituent of a compound.

Yes-no questions are signaled by the SPtc:  $\bar{e}$  ʔō tənā ʔo kō  $\bar{e}$

Are you free today?

Question-word questions (WH-questions) are formed with the following question

words:

ʔitē	what?
me tē	why?
tō ú tē	where?
bó kē tē	when?
bá CLF tē	how many?
bś tē	where? (nearby)
hú tē	how?
dś tē taCLF $\bar{e}$	which one?
ʔū ... pē	who?

$\bar{a}$  is the pronoun 'they, other people, someone', while  $\bar{p}$ , whose only function is to trigger the interrogative meaning of  $\bar{a}$ , must be considered a SPtc: e.g. it may follow the negative, as in  $\bar{a}$   $\bar{m}$   $\bar{t}$   $\bar{p}$  'who doesn't look?'.  $\bar{t}$  may be considered a bound Noun,  $\bar{a}$  'what, anything' being its free equivalent. Most of the other question words are



analyzable:

me tē	verb+Obj-x	do what → why?
bá CLF tē	preposition+object	as-much-as-what clf's → how many?
bá tē	preposition+object	at what → where?
hú tē	preposition+object	like what → how?
dá tē taCLF ē	preposition+object+Sptc	

in the last, the object of *dá* is an NP consisting of N *tē* and ClfP: 'what one-clf → which one?' (*dá* here is not locative)

*bá kē te* 'when?' is not analyzable. *tō ú tē* 'where?' is probably grammaticalized from an expression including a Resultative V-V with the second verb being *tō* 'strike, correct, exactly', as if 'to V exactly at what?'. Cf. *métha ʔa tō bá tē* ~ *métha tō ʔa bá tē*, both 'where did [you] see him?' *ú* must then be a Preposition, perhaps derived from *mú* 'at, not in sight'.

Note that the sentence-final position of these words is partly due to syntactic factors; for instance, those that constitute a PP could only be followed by a ClfP at most. But sentence-final position is also universally a position of 'focus', i.e. not the background or reference-point established by the Topic, but its complement, the portion of information to which the hearer's attention is directed. Obviously the question-word in a question is quintessentially a focus, since it is, in effect, a blank that the hearer is being asked to fill in.

All question words may have indefinite meaning; in fact it is better to think of interrogative and indefinite as conditioned variants of a single 'meaning'. The indefinite meaning is triggered by the conjunction/verb *ma*, e.g.

*cwá bó kē tē ma ré* to go whenever is good: it's good any time

(34) *bá tē'' ma sí?ichē to ?a be phri*

how-much be-so fear NEG 3 must buy

However much it is, don't be afraid, he must buy it. (174.2)

and in other contexts also:

(35) *?a ?é ?ojwā lū bá tē ?ojwā kāl lū to*

Wherever he called to them to wait, they didn't wait for him

(cf. 6.3.2, example 9)

## 9. Interclausal Syntax

This chapter will first sketch some of the ways in which clauses may be embedded in each other, including clauses modifying a nominal (9.1.2, 9.2) and clauses embedded as constituents of other clauses (9.1.1, 9.3); finally we will consider sequences of linked clauses, with no embedding (9.4).

### 9.1 Non-autonomous nominalized clauses.

A nominalized clause is any clause followed by *m*, or by a ClfP before which *m* can be inserted. If the nominalized clause is autonomous (in which case there is no following ClfP), the *m* may be taken as equivalent to an illocutionary force marker, as discussed above (8.3). If the clause is non-autonomous and has the following ClfP, it is a *preposed attributive clause*; if there is no following ClfP I will simply term it an *embedded clause*.

#### 9.1.1. Embedded clauses. Embedded clauses are typically the objects of prepositions:

(1) b́ [Jepu hē nλ] ma tēú ?o pa mē

at Japan come Nλ be-so fish have DUR PTC

When the Japanese came, there were still [many] fish. (205.3)

(2) tē tādω b́ [ʔa sō né ʔa li nλ]

measure go-on at 3 green Ne 3 red Nλ

Measure up to where it's green and red. (287.4)

(3) ʔa da láteá li phá hú [phē ʔíro] hé nʌ

3 give instead book skin like father sing say Nʌ

He gave a hide book instead, as Father sang, it's said. (100.4)

PP's containing embedded clauses may occur in all the typical PP functions: Topic as in (1), Oblique/Locative(Goal) as in (2), Oblique/Adjunct as in (3). Note that the embedded clause in (3) lacks *nʌ*; my analysis of it as nominalized rests largely on the definition of Prepositions as taking NP objects.

### 9.1.2. preposed attributive clauses.

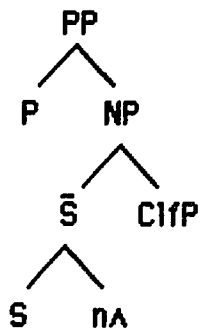
The idealized maximal form containing a preposed attributive clause (AC) is:

P S nʌ ClF

If *nʌ* is a complementizer, as suggested above (8.3), this should be revised to

P  $\bar{S}$  ClF

I assume that S-nʌ-ClF forms an NP, which functions as object of the preposition, so that the structure would be



Both the Preposition and *nʌ* may be omitted.

Examples without *nʌ*:

S-ClfP

(4) ?a ?e tēú tahe

3 eat fish one-CLF(groups) the ones who were eating fish (198.6)

P-S-ClfP

(5) dǎ thō taphre

at go-over one-CLF(people) the person [who lives] over there (265.5)

I have nothing to say at this point on what the conditions might be that bear on the presence or absence of *na*. If *na* is a complementizer, there is of course a likely parallel in English, as in *the book that I read ~ the book I read*. The impression that omission of *na* is not of especially great moment is reinforced by examples like the following two, from a single narrative:

(6) ?a bÉ rǎ tǎba

3 mold beforehand one-CLF(worlds) the first one he made (337.5)

(7) ?a bÉ rǎ nǎ nǎ ìa

Nǎ two CLF the first two he made (340.3)

The semantic relation between the AC and the head ClfP can usually be thought of in terms of a semantic role or syntactic position that the (referent of the) ClfP plays in the AC. In (4) above, the head has the Subject/Agent role; in (6) it has the Obj-x role (Goal? Patient?). Obj-1 and Obj-2 are also possible:

Obj-1

(8) kajē dǎ ?iphē dǎ iū rū nǎ cwá thō á

person at Father give 3OBV money Nǎ go finish NS

The person who Father gave money to has gone.

Obj-2

(9) mā|dǎ|vĕ|dǎ pè kǎ|vĕ pò|pā nǎ| tǎbe

be-so P 1s give TRN COM 1s VS IRR Nǎ one-CLF(tools)

It's the one that I'm going to give to my brother. (11/24)

And relations occur that are even 'more oblique':

(10) kwā nǎ mā dǎ ?a mū lípǎna kuklǎ nǎ tǎbe

axe Nǎ be-so at 3 hit nail head Nǎ one-CLF

The axe is the one in whose head he pounded a nail. (252.1)

('the one that he pounded a nail into its head')

Here the axe's role in the AC might be characterized as 'possessive' or 'genitive'.

As the preceding examples show, there may or may not be an element (usually a pronoun) present in the AC, marking the position that relates to the head. If the head relates to Subject, there is generally a pronoun in that position in the AC, as in (4); cf. also *pū rĕ to tǎbe* 'the group that's not good' (271.2). If the head relates to an Object, there usually is no pronoun in that position, as in (6-7, 9).

Embedded clauses as locatives. A distinct subtype of nominalized clause that should be mentioned occurs in Locative expressions. Most simply, and quite commonly, these are of the form [pp dǎ V (nǎ)], where V is any of the Type B Directionals (4.2.1.2) (with the probable exception of *tó* and *tǎka*). Usually '\_\_\_ there' is a good translation:

dǎ the nʌ	up there
dǎ nō nʌ	in there
dǎ thō nʌ	over there

These can be viewed as an abbreviation of *dǎ V NM takjā* where the nominalized clause *V NM* modifies a ClfP headed by *kjā* 'side, direction'. A more explicit syntactic analysis might have empty categories representing the semantic connections, e.g.:

(11) [<sub>pp</sub> dǎ [<sub>NP</sub> [<sub>S</sub> e<sub>i</sub> | the [e<sub>j</sub> nʌ] takjā<sub>j</sub>]]

Here e<sub>i</sub> is a Subject pronoun with indefinite or arbitrary reference, and e<sub>j</sub> is the Locative expression, indicating by its indexing the role that the modified head *takjā* plays in the embedded sentence. An extremely literal translation reflecting this analysis might be 'at the direction<sub>j</sub> in which<sub>j</sub> one<sub>i</sub> ascends'.

That these expressions are true clauses is demonstrated by their ability to contain constituents<sup>1</sup> in addition to the verb:

dǎ lɛ Thóká hi nʌ	down there at Thoka's house
dǎ the dō du	up there at the big village

And nominalized clauses can have Locative meaning without containing a directional verb:

(11) tē      tǎdu bǎ ?a sō   nē ?a li nʌ  
 measure onward at 3 green Nɛ 3 red Nʌ  
 Measure up to where it's green and red. (287.4)

This simply points up the fact that the minimal 'Locative clauses' described above are a subtype of nominalized clause. It is convenient to give them a separate description, but there seems no need to distinguish them formally.

### 9.2. postposed attributive clause.

Attributive clauses following a noun often consist only of a single verb, 'adjectival' or action:

pōcī du	big <i>cedi</i>
bē ʔū bū	thin cloth
dī chwí	cold cooked rice
kəjē vī	person who drives, driver
lè de	place to put [stg]

The postposed clause may contain other constituents:

VPTc:	ʔiswí hé chílū	curry+hot+excessive → curry that's too hot
	súpɿɿ cō du	rope+tie+own+accord → self-tying rope [a magical object]
Obj-x:	pənè cwī ʔíthá	buffalo+pull+plow → a buffalo to pull the plow
Subj:	thwá kúmì thū	cat+tail+long → cat whose tail is long, long-tailed cat
	chwí phōchí ʔà	chill+ mosquito+bite → mosquito-bite fever: malaria
V-V:	mò kò ré	doctor+inspect+good → doctor who examines well



SPTc:            kajē pā tha pā            person+cut+sesame+IRR → people who will  
harvest sesame

kajē mò ʔo tɔ phē ʔo tɔ    people without mother or father

PAC's tend to be distinctly shorter than their preposed nominalized counterparts, the vast majority containing at most five morphemes. It is not yet clear whether this is best stated as a formal restriction, and if so what exact form the restriction should take. It does seem that NP arguments have an upper limit of one.

The modifier of the abstract noun /e`'place/thing for V-ing' (6.2.1) is a type of postposed AC. The construction headed by /e` may also modify a preceding noun, giving /e` some of the flavor of an attributive marker:

khruw lè bó thé    machine+lè+weave+cloth → loom

swá mo lè njā    friend+happy+lè+laugh → a happy friend to laugh with

Any postposed AC has an equivalent preposed AC, although the modified head must differ: noun for the former, ClfP for the latter, e.g. *kajē vī m̀r̀r̀aká* (postposed), *ʔa vī m̀r̀r̀aká (na) taphre* (preposed). The question arises of the semantic or rhetorical difference between the two. Definiteness is a factor: in this case, the preposed construction would most likely be translated with the definite article, 'the person who drives (the) car(s)'. The postposed version could have the same translation but could also be indefinite: 'a person who drives, a driver'. The postposed construction is more susceptible to the connotation of purpose, as in 'a person to drive', cf. also 'buffalo to

pull the plow' above. A third point is that only postposed AC's are found in lexicalized compounds:

tétabō thλ ʔo pencil+water+have → pen

chwí béseplo bē chill+face+yellow → jaundice

chwí phōchí ʔà mosquito-bite fever: malaria (cf. above)

Lahu has a similar distinction between preposed and postposed modifying clauses; Matisoff suggests that the postposed variety ('RRC', right relative clause) 'ascribes some more or less *permanent quality*' (490; italics in original). However permanence is not essential in the Kayah equivalent; e.g. *phre nō* 'the person ordering' can simply pick out the speaker of a command, whether or not that person is characteristically or habitually a commander. What is more important is the different types of nominals that the two types of AC modify; in particular the difference in discourse-pragmatic status between common nouns and classifiers. A unique feature of classifiers (apart from time classifiers and others that do not count any common noun) is that they cannot be used unless some antecedent (in a non-technical sense) is available, whether in the linguistic context or the real-world setting. E.g. I can use the expression *ʔā tamē* 'this one' only if some house<sup>2</sup> is in view or has been mentioned. Classifiers, then, are 'given' or 'evoked', but in a special sense: while the common noun may have definite reference, it may also merely name a category, or which the ClfP picks out a specific instance(s) or member(s). The consequence is that preposed AC's are part of an NP that is by definition 'given', in

the sense described; while the NP including a postposed AC has no particular inherent pragmatic status: it may be 'given' or not.

### 9.3. Clause as argument of verb

9.3.1. Clausal Objects. The perception verbs *méthla* and *nihō* can take a clause as

Obj-x:

(12) méthla [ Doa phē ka ] to na  
see (name) father come NEG NA

[I] haven't seen Do'a's father come back. (316.4)

(13) ne méthla mo [ Phēlúdu me hú ?ā ] to  
2 see happy (name) do like this NEG

You are unhappy seeing P. act like that; seeing P. act like that make you unhappy.

(11/21)

(14) nihō [ ?a ?é mò ?é phē ] to  
hear 3 call mother call father NEG

[we] don't hear him call his mother and father. (249.6)

The bracketing given, in particular the exclusion of *tə* from the embedded clause, is arrived at in (13) and (14) by common sense, which dictates that the negation must apply to the matrix clause. If it applied to the embedded clause we would get *We heard him not call his mother and father*, which does not make sense, and *You are happy not seeing P. act like that*, where the context makes it clear that it is a case of P. acting in some way and the addressee being unhappy about it. Both possibilities would be all right in

(12) (the other being / see [that] *D's father hasn't come back*), but I follow the bracketing as in (13-14) out of consistency.

The most remarkable aspect of these clausal arguments is that they usually, perhaps always, are not terminated by *m*. This poses a serious problem for the analysis of *m* as a complementizer, since clausal arguments are a prototypic function of  $\bar{S}$ 's. Certainly more research is needed on clausal arguments; for the time being it can be said that they are nowhere near as widely used in Kayah as in more familiar languages: this section is in effect concerned with the subcategorization features of three verbs only.

The third verb with this feature is *ŋjǎ* 'encounter some bad event'; like the perception verbs, it takes either NP or clausal Obj-x:

- (15) ne ?íre to chá pā ma ne ŋjǎ ?íkwa ní  
 you work NEG soon be-so you encounter stick PTC  
 [If] you don't work you're going to encounter a stick, now! (11/9)  
 [i.e. you're going to get a beating]
- (16) ne sí ŋjǎ [lo khrū pò] mē  
 you want encounter car bump PTC  
 You're going to have a collision! (11/9)

The Obj-x in (16) consists of Subject NP *lo khrū* and verb *pò*.

9.3.2. Clausal Subjects? There are occasional examples that appear to have a clause as Subject, or perhaps Topic.

- (17) [ʔa sí tərē] se ʔo kāl ʔū təcʰ to  
 3 ashamed useful COM 3 one-CLF(sorts) NEG  
 Being ashamed isn't *any* use to him! (265.1)
- (18) [ bós kʌ bʰʰ ʔa khu nʌ né ʔa kē nʌ ] twà kāl né ke  
 weave striped at 3 top Nʌ Ne 3 base Nʌ pretty COM Ne PRH  
 Weaving it striped at both ends might be pretty. (277.3)
- (19) [ʔa ka nós to] ʔo á nāl iē á  
 3 come at-all NEG have NS two month NS  
 He hasn't been back for two months. (2/20)

Here again *m* seems not to be required. Indeed, we cannot rule out the possibility that these may be *sequences* of autonomous clauses; e.g. 'he hasn't come back; it's been two months'.

9.3.3. Quotatives. The pattern for reported speech (or, less commonly, reported thought) is

S NP<sub>1</sub> hé (rál) NP<sub>2</sub> nʌ

where S is a clause representing the quote, the reported speech; NP<sub>1</sub> denotes the speaker of the quote, the person who uttered S; *hé* is the verb 'say'; and NP<sub>2</sub> denotes the hearer of the quote, the person to whom S was uttered. Examples:

- (20) [ne chá māl phrè tē Sethūphē] ʔa hé nʌ  
 2 fight PTC Shan PTC (name) 3 say Nʌ  
 'Attack the Shans, Sethuphel' she said. (353.1)

- (21) [dɔ phrē phrē tarú ʔó chwa chwa] hé rá ʔa pò vè du lɛ nɔ təhe  
 beat fast go-ahead blow strong say Rλ ʔ sibling big CMP that one-CLF  
 'Beat [the drums] fast, go ahead and blow [the flutes] loudly,' he said to his  
 older siblings. (54.6)
- (22) [ma sí ʔɛ ma ʔo kɔ dw pa tɔmjō  
 be-so want lucky be-so have COM self DUR one-CLF(kinds)  
 sí thwā ʔo kɔ dw pa tɔxʃ ʔa hé rá ʔa lɔ nɔ  
 want lucky have COM self DUR one-CLF(kinds) ʔ say Rλ ʔ grandchild Nλ  
 '[you] will be lucky of your own accord; [you] will be fortunate on  
 your own account,' he said to his grandchildren. (88.6)
- (23) [chá pā ʔū jù ʔe vē to he vē ʔo kē dwə λ ]  
 soon ʔ believe 1s NEG LEST 1s have different-place self PTC  
 ʔa ré ne kɔ nɔ  
 ʔ should think COM Nλ  
 He *should* think, 'Soon nobody will trust me and I'll be an outcast,' (but he  
 probably won't). (313.2)

Several analyses of this pattern are possible. One might wish to see the quoted S as a constituent of the clause whose main verb is *hé*. Since *hé* usually has a Subject, representing the speaker of the quote, the preceding S might be a Topic. However this would be pragmatically incongruous: the usual function of a Topic-Comment pattern is to take the Topic as the starting point or background to the information contained in the Comment, which is why the Topic must be in some manner given, activated, recoverable, etc. But the point of the quotative pattern is not so much to identify the speaker of the

quote as to convey the contents of the quote; indeed the quote is always 'new', and this alone would be enough to disqualify it from being treated as a Topic. Furthermore the following clause with *hé* nearly always has a pronominal Subject or none at all, and is usually uttered weakly. It has some of the feel of an 'afterthought', particularly in legendary narratives, in which virtually every sentence ends with a muttered *hé na* or *ʔi hé na*. It may well be on its way to grammaticalization as a quotative marker.

Looking back to its historical source, we may speculate that the quotative pattern is a vestige of verb-final syntax: the quoted S would then have originally been a clausal *Object* of the verb *hé*. This may not be as contradictory as it appears: the evolution would be from verb preceded by clausal Object to independent clause followed by reduced-clause afterthought, the main reanalysis being the shift of main-verb status from *hé* to the verb of the once-embedded clause.

#### 9.4. Clause sequences and *ma*

*ma* is a verb meaning 'be so, be true'. It occurs in the criterial verb environments: *ma í* 'it's so, yes' (cf. Thai *cháj léew*), *ma to* 'not so, no' (Thai *máj cháj*). There is also a more expanded pattern which can be symbolized [X*ma* Y], where X and Y may be NP, PP, or S (or possibly  $\bar{S}$ , since *na* is frequent in the slot S \_ *ma*).

When both X and Y are NP's *ma* has the flavor of a copula:

(24) ʔā hō ma ʔiū

This is an Ilu. (47.1)

(25) *bō lō mā dīklwī mō nλ ?a lō ?ə dǝ ?a kū nλ*  
 banana-heart banana plant Nλ 3 have at 3 inside Nλ

'Bo-lo' is a banana plant; the 'lo' is inside it. (298.1)

(26) *pe mā kǝjē li phú cé*

1s Kayah red child real

We are genuine Kayah. (1p.3)

(27) *ne tǝphre mā tǝklí ŋara phé nλ*

2 one-CLF turtle sin only Nλ

You[rs] is a sin involving turtles. (173.2)

As (26) shows, the relation between X and Y is not one of simple equation, but has a value that can only be called Topic-Comment: 'as for you, it's a turtle sin'. Chao cites a very similar example in Chinese, *tā shì ge Ribēn nǚrén*, which is not 's/he is a Japanese woman' but 'as for him, it's [his secretary is] a Japanese woman'. It is important to note that the order of X and Y cannot be reversed: X is the thing in evidence, the word to be defined, and Y is the information supplied that is relevant to X.

It might seem that in this expanded pattern *ma* remains a verb, with X its Subject and Y its Object. There is a serious difficulty in this view, though, since [*Xma Y*] fails to meet the definition of the clause: it cannot end with the negative *tə*. The way to negate statements like (24-26) is *Xma tə*, where it is very difficult to have anything between *ma* and *tə*. I write *Xma tə* since the pre-*ma* constituent must still be given, mentioned, in evidence, etc. But most commonly this X has attained that status by having appeared as the Y constituent in a preceding [*Xma Y*] pattern. For example



(28) A: ?ʌ ma ?ilū      B: ?ilū ma to      (\*?ʌ ma ?ilū to)

A: This is an Ilu      B: It's not an Ilu.

Thus if we were to represent the statement-denial exchange as a single pattern, it would be [X ma Y, Y ma to].

When either or both of X and Y are S or PP, several subtypes can be recognized, all related in meaning:

#### Topic-Comment

(29) thā ?o nʌ mʌ kəjē    ?íjē tā    phɾɔ    bōʌ    thwā    thā ?o nʌ  
 pond Nʌ    person jump fall cave-in then become pond Nʌ  
 The pond, now, a [person] jumped so that [the earth] caved in,  
 and it became a pond. (164.5)

(30) Sō    ?a    phē    te    dɛ    Lē    mò    kʰí    mʌ    ?a    piē    ?à    tò  
 (name) father 's at creek dark      ʒ ear bite arrive  
 So's father's, at Dark Creek, [they] ate [everything,] even the ears  
 [of grain]. (182.4)

#### Setting

(31) dɛ    ŋjá    nʌ    mʌ    ?ū    chá    lū    nʌ  
 at long-time Nʌ      ʒ fight mutually Nʌ  
 Long ago, they had a war. (230.2)

(32) dɛ    ?a    lē    nʌ    mʌ    ?akū    ?o    rʌ  
 at ʒ underneath Nʌ      hole have Rʌ  
 Under it there was a hole. (212.2)

Antecedent-Consequent

(33) vĕ ke pù nì jòkhró tādō mavĕ cǒ lúlǎ khe takjā

1s if catch get rat one-CLF 1s tie dangle leg one-CLF

If I can catch one rat, I'll tie it up hanging by one leg. (182.10)

(34) bó ré to maʔū do

rice good NEG 1s-humilific abstain

Because the crops aren't good, I'm fasting. (175.1)

Interrogative→Indefinite

(35) nō ʔe rǎ ʔitē ma ré

use RA what good

Whatever you use is good. (91.4)

(36) bá tē" ma síʔichē to ʔa be phri

how-much fear NEG 3 must buy

However much it is, don't be afraid, he must buy it. (174.2)

The basic meaning of [X $\text{ma}$ Y] can be put as 'given X, Y is pertinent'. In all the above types the X is taken for granted, recoverable, presupposed, etc., while Y is the relevant information, the consequence. Even in English, Topic-Comment, if-then, and time-event can overlap:

(37) *ma*ke hé kǎjē ŋē *ma* tā síjɔ chá lè phé  
 if say person part should care-for honor only

*ma*\_bó ŋē *ma*hé cè pa to ʌ  
 rice part say able DUR NEG PTC

If you're talking of people, they must care about honor; as for rice-plants, you  
 can't really say [anything]. (175.5)

or As for people . . . (Topic-Comment)

or When it comes to people . . . (time)

It should also be pointed out that (37) is quite typical in the multiple use of *ma*.  
 Kayah discourse of all genres is generally peppered with this morpheme, which is one of  
 the most common in the language, and often seems to function as little more than a  
 pause-filler.

In (35-36) question words are given an indefinite reading by their appearance in the X  
 constituent of the *ma*-construction. This is also consistent with the general meaning of  
 the *ma*-construction. The question word in an interrogative sentence is a non-Topic, a  
 'focus' virtually by definition, as I mentioned in 8.3 above. Appearance in the X  
 constituent, which is by definition 'topical', can be said to cause the question-word to  
 have its non-interrogative meaning.

*ma* in (29-37) has much of the quality of a conjunction, specifically one joining  
 clause with clause or else Topic with clause. Even in (24-27) it is a rather peculiar verb  
 since it seems to take Subject and Object in positive sentences but Subject only in  
 negative sentences.

Classing *ma* in its 'conjunction use' as something other than a verb would allow it to join several other connective morphemes, some of which indeed seem to include *ma* as a constituent; among them are:

bōλ ~ bō λλ	and then
λλ	(?) unmarked pause
mané	but (perhaps < 'be so' + 'and, with')
toíma	otherwise; if not for X, then Y
loíma	consequently, if...then, whenever...then
nékū	(similar to preceding)
nλhō	

These can be opposed to the Class B and C SPTc's, with which they are mutually exclusive, for the most part. The morphemes listed above may be called 'nonfinal particles', since part of their function is to signal that the clause they terminate is a nonfinal member of a sequence.

## Footnotes

Chapter 1

1. The first syllable of both words is probably the same obscure element found in other body-part terms; see 2.2.
2. 'Karenbyu'; evidently a speech of western Karenni and distinct from Sgaw, the latter being known in Thailand as *kariaj khǎaw*, also meaning 'white Karen'.
3. However Smalley does not use the term 'minor syllable'.

Chapter 2

1. A parallel can be seen in Chinese *zi* and *er*, whose root meanings are both 'child', but also occur in reduced form as essentially simple markers of noun-hood *er* even loses its syllabicity, being realized as retroflexion, plus other modifications, of the vowel of the main syllable to which it is attached.
2. Part of the Kayah female costume, consisting of large numbers of rings made of lacquered twine, gathered around the knees.
3. These two roots could result from a pre-Karen alternation of final nasal and stop, as \*ʔdoŋ and \*ʔdok; other examples of this type of alternation can be found elsewhere in Tibeto-Burman.

Chapter 3

1. It is further possible to distinguish the price as value (e.g. *five dollars*) from the price as currency (e.g. *a twenty-dollar gold piece*), as pointed out in Fillmore 1972 (72).
2. It is true that the latter response might not be so odd given the proper circumstances: if the speaker and hearer are aware that crocodiles are part of an exhibit or something else to which admission might be charged. But it is fairly clear that such special circumstances are not encoded in the lexical meaning of *see*.
3. Fillmore (1971) treats the sentence *The room is warm* as having a Locative Subject; which would imply that the verb (*be*) *warm* specifies Locative and no other role. But it seems equally possible to analyze *warm* as taking a Patient role like other stative adjectives. The fact that *room* may have a Locative role in

other sentences need not be relevant. An analogous case is *John* in *John died*, which has the Patient role regardless of the possibility of its occurring with other roles, such as Agent in *John worked*

4. These remarks oversimplify, in that grammatical morphemes like prepositions can have characteristics that fall between the clear cases represented on the one hand by pure case-markers that always mark case and are the only or the predominant means of case-marking, and on the other hand by items like *from* that have a constant, role-related value. Examples are the Particle members of English verb-particle idioms like *rely on*, *look up*, *look for*, which have little or no semantic content of their own but seem to serve principally as markers of the Object argument of the complex verb. There are also prepositions that have a 'core' value along with varying specialized values in conjunction with certain predicates; e.g. *of* usually indicates genitive, but it has different, specialized value in expressions like *rude of him*, *nice of him*, and so on.
5. However, O'Connor (1986) argues that, for Northern Pomo at least, it is an oversimplification to view the morphemes in question as marking semantic roles only.

#### Chapter 4

1. Chao in fact allows himself to have it both ways: although he speaks of 'Verbs that are used as Coverbs', and refers to constructions with Coverbs as an example of 'Verbal expressions in series', he also discusses Coverbs as a distinct form class ('K', pp. 749-67), and refers to Coverbs that have full-V analogs as cases of 'class overlap'.
2. *pe'* however is undoubtedly cognate with Burmese *pei* and Lahu *pi'*, which are synchronically verbs; cf. also Mikir *pi*, Miri *bi*, and Dulong (He dialect) *bi*<sup>55</sup>, all 'give'. The full verb 'give' in Kayah is represented by a distinct morpheme *di'* (etymology uncertain).
3. *thj khri* 'drum-fragments' is a N-N compound functioning as (Direct) Object. It might seem that '(be) fragments' is also a kind of result expression, but it is not, at least not grammatically. A closer translation might be 'he broke off a few drum-fragments'.
4. Another sort of deixis is denoted by the locative Prepositions; see Chapter 7.

5. See 4.3.4. below for a set of exceptions to this; they are, however, Locatives specified by non-verbs.
6. Here *me* might be better translated as 'do (stg) to (sbdy)'; cf.  
 ?a me nó lū təcú to  
 3 do at-all 3OBV one-CLF(sort) NEG  
 He didn't do a thing to her. (78.2)
7. The fact that the negative *to* is a Clause Particle, and so not a constituent of the VC, adds a slight complication: the latter structure should really be [[[de síplɔ [cwá cè]] to], which would be expected to mean 'not decide to be able to go'--but that reading is not possible either.
8. More accurately, *hē* temporally precedes *ʔo mā kɪɛ*, while *ʔo mā* and *kɪɛ* are alternating actions. However both temporal precedence and alternation count as values of a Sequential V-V.

### Chapter 5

1. This is a weaker type of evidence: while monoclausal structure would be likely to have this sort of restriction on expression of the first verb's Object as an effect, it is not impossible that a biclausal form might produce the same effect.
2. What is needed, but so far unavailable, is an example of the form  
 ?a<sub>i</sub> [Drv V V] vē (...) X<sub>j</sub> (...). If it can be established that the intervening non-third-person NP does not block the obviation relation in simple sentences, and if X here = *lū*, it would establish that X is accessible to obviation, hence in the same governing category (≈ clause) as *ʔa*.  
 It should also be pointed out, concerning example (6), that the free reference of *ʔa nē* 'his body, himself' does not prove that *ʔa nē* and the (matrix) Subject *ʔa* are in different governing categories. Consider the putative underlying structure:  
 ?a nō \_\_\_ e [e chū ʔa nē ]
- In the reading with *ʔa nē* coreferential with the matrix Subject, there would presumably be an empty Subject of the 'lower' clause, and *ʔa nē*, being disjoint in reference with that lower Subject, ought to be the obviative *lū*. All that is demonstrated by this is the need for further research.

3. The phenomena covered by predication theory include not only the infinitival clauses being discussed, but also constructions involving the other three phrasal categories *He ate the meat raw, I dub you a Tenderfoot Scout*, etc. For these likewise there are analyses involving PRO (e.g. He ate the meat [<sub>S</sub> PRO raw ]) and analyses without it. Predication theory does not necessarily rule out use of PRO: thus Williams has the predication theory index the S containing PRO. But neither does it require PRO: Culicover and Wilkins replace PRO-sentences with VP's indexed by predication; LFG does essentially the same but the VP's are further said to bear the grammatical 'function' (=relation) XCOMP and to have their subject determined by 'functional' (as opposed to anaphoric) control.
4. It would be possible, as a theoretical innovation, to allow phrase-level categories (in this case, the VC) to have subcategorization features. For example, Generalized Phrase Structure Grammar allows phrasal categories--i.e. nodes of trees--to be notated with various features, which could presumably include subcategorization features. Cf. also Kayne 1984, in which the intermediate-level category  $\bar{V}$  is said to be capable of taking a PP complement when dominating the proper verb; e.g. [[[keep]<sub>V</sub> money] <sub>$\bar{V}$</sub>  [in the box]<sub>PP</sub>]<sub>VP</sub>. In that case either the verb *keep* will have to pass its subcategorization features up to the  $\bar{V}$  *keep money*, or else the verb's subcategorization domain will have more than strictly local. An equivalent concept is provided by the notion of 'complex logical structure' of Nuclear/Core junctures in Foley & Van Valin 1984.
5. More specifically, in Vietnamese the structure both before and after the application of the transformation are acceptable surface structures, where in Kayah the before-application structure is blocked.
6. A Resultative V-V with *me* 'do, do to' as first verb would be the closest semantic equivalent in Kayah to the morphological causative.
7. Baker's framework would actually involve an extra wrinkle. For various reasons, stative verbs are said to occur in underlying structures with an Object but no Subject, the Object NP then obligatorily moving to Subject position. This makes no difference; in fact it reduces the argument to the question of subcategorized sentential objects only.
8. For the former point, cf. Hoekstra et al. 1980 on the notion that 'The subcategorization domain of a V can be invaded by adverbial phrases...' (p. 37, note 86); for the latter, see Culicover & Wilkins.



9. 'External argument' is more or less the equivalent of deep-structure subject; turning an adjective into a verb adds a new external argument and makes the old external argument internal--i.e. the Object. The fact that the name of the rule includes an instance of the rule's application (the derived verb *internalize*) is due to Williams; the pun is presumably intentional.
10. This does not negate the observations made at the end of 4.2.4, concerning the tendency for  $V_2$ 's in Resultatives to denote 'intrinsic processes' while  $V_2$ 's in Descriptives denote 'intrinsic states'. If it turns out to be possible to capture this distinction as a lexical feature of the verbs in question, then we could have the presence of that feature of  $V_2$  trigger the [-mp] specification of the V-V. This would have the desirable result of further reducing the possible ambiguities.
11. Carrier-Duncan's lexical entries also include distinct specifications of both argument structure and syntactic valence.
12. The argument type of the second argument may still be available to the semantics; this would help account for the understanding of the Obj-x of *vē pīl' cwi pu* 'I whip the ox to make it pull stg' as being in a sense both the whipped and the puller, although morphosyntactically it can be only a Patient.
13. More precisely, one of the two Obj relations should be specified as optional, in order to allow for Directives where the second verb is a  $V_1$  (*vē dī cwā ʔa* 'I let him go', etc.). The salient point here is that it makes sense to attribute the possibility of the Obj-1 relation to features of the Directive verb.
14. This semantic-role label is used for purposes of illustration only.
15. The theory described in Wilkins and Culicover 1986 does allow a single constituent to have more than one semantic role, but the single constituent 'may be assigned at most one role ... BY A GIVEN VERBAL ELEMENT.' (123, emphasis in original) It is not clear to me whether a compound verb counts as one or two 'verbal elements' in this theory.
16. It does not change things if Sequentials are analyzed as head-final, as suggested previously. The Sequential example could then be *phjā katha* 'take stg and go up [with it]', which is  $V_t$ - $V_i$  and takes an Obj-x that realizes the Patient argument of  $V_1$ .

## Chapter 6

### 1. Notes on these verbs:

*bo* 'feed, raise'; usually applied to feeding animals only, its typical Obj-2 argument being *che* 'animal feed'. Exception: the elaborate expression *bo phú mū mē* 'feed children and nourish wife'.

cf. *kē rō panē*/\**kajē thū* 'give buffalo/people water to drink'; uncertain whether to treat as Res or Drv V-V.

*ʔiswá* 'teach, study'; polysemous, with the following characteristics:

- |                                            |                                                         |
|--------------------------------------------|---------------------------------------------------------|
| 1a. ʔiswá NP                               | learn NP                                                |
| b. ʔiswá V (NP)                            | learn to V, learn to V NP (a Quasi-modal V-V?)          |
| 2a. ʔiswá NP <sub>1</sub> NP <sub>2</sub>  | teach sbdy(NP <sub>1</sub> ) stg(NP <sub>2</sub> )      |
| b. ʔiswá V NP <sub>1</sub> NP <sub>2</sub> | teach sbdy(NP <sub>1</sub> ) to V stg(NP <sub>2</sub> ) |
- (a Directive V-V)

*bule* 'exchange, trade'. Cf. also:

*pe bule iū haca* 'we trade clothes' (10/31); i.e. 'we exchange clothes with each other', with reciprocal Particle *iū*

*ʔa bule vē né haca* 'he gives me clothes in exchange for it' (id.)

2. Describing construction of the *ʔiū* ritual post (see Lehman 1967). *po*, here rendered 'final', is an elaborate structure of carved wood and bamboo that surmounts the *ʔiū*; the word also means 'pen, coop, corral'.
3. *te rō* 'things exist → be wealthy' is best considered a Subject-Predicate compound with no other specified arguments. Even the simple verb *rō* 'exist, be, have' is specified only for Patient (the thing existing) and Locative, realized as Subject and PP<sub>2</sub> (with *dé/bé/mú*). In the meaning 'have' the possessor is realized as a preposed modifier of the Subject, e.g. *vē rū rō* 'my money exists → I have money'.

## Chapter 7

1. There are exceptions, such as *Lwə pé* 'mute Lw'. However it is not clear whether to consider *pé* here to be a true modifier (e.g. if mute Lw is distinguished from others named Lw) or part of the name.
2. An exception to this is the morpheme *lǎ*, otherwise a Descriptive VPtc 'more, than'. Num<sub>x</sub>+Clf+*lǎ* means 'over x Clf's'; e.g.

*hō rō ŋē khwē lǎ* there's over 5 baskets of rice (2/1)

*sīnē rō khu nā je lǎ* there's over 200 guns (c)

*mū sō swá lǎ* after 6 o'clock

Note the minimal pair:

*ʔo sō swá tamū lǎ* be there over 7 hours

*ʔo lǎ sō swá tamū* there's 7 hours remaining

Thus *lǎ* can be considered a postposed modifier, although since it is not otherwise a verb this construction is exceptional. It is undoubtedly a loan from Shan, the Thai cognate being the verb *lǎ* 'be left over, remaining, in excess of what is needed' (Haas).

3. *bá chā* refers to the practice of divining by inserting splinters of wood into holes in the leg-bones of a chicken.
4. *bé* may be related to the first element in the Descriptive Particle *bé bu* 'take sbdy to V, show the way to V'. Even more speculatively, there may be a connection with *bé* 'stick to, get on', as in *hemu bé kōlò* 'dust gets in your hair'; the semantic connection would be via 'stick to → be attached to → be attached emotionally in friendship'.
5. Assuming that the numeral modifies the classifier, this ordering rule might be taken to indicate that *chā~chá* and *je* are verb-like, since verbs characteristically follow what they modify (the verbal nature of *swá* has already been seen).
6. Data on the productivity of *swá* is lacking, but my guess is that e.g. *\*jē swá* 'five doubled' or *\*nā swá* 'two doubled' are impossible.

## Chapter 8

1. This is far from being the last word on the behavior of *pā*, as the following examples indicate:

(a) *kəjē [pā tha nɿ] təhe ʔo māl dʒ vē hi pā*  
 person cut sesame Nɿ one-CLF sleep at 1s house IRR

The sesame-harvesters (people who *will cut* ses.) will sleep at my house.

(11/24)

(b) *kəjē [pā tha nɿ] təhe pā ʔo māl dʒ vē hi pā*

The sesame-harvesters (people who *are cutting* ses.) will sleep at my house.

(id.)

The latter indicates that *pā* can be in construction with NP; or possibly that *kəjē pā tha nɿ təhe* is more like a clause containing a ClfP than like a ClfP modified by a preceding attributive clause.

Chapter 9

1. These examples contain overt Locative expressions, but as Obj-x rather than PP. In an autonomous clause they would be more likely to be *le dš Thóká hi* and *the dš dš du*; I have at present no explanation of this.
2. or some other of the types of object lexically associated with *mā* (see 7.3).

## References

- Baker, Mark. 1985. *Incorporation: a Theory of Grammatical Function Changing*. Ph.D. dissertation, MIT.
- Carrier-Duncan, Jill. 1985. Linking of thematic roles in derivational word formation. *Linguistic Inquiry* 16.1, 1-34.
- Chafe, Wallace L. 1970. *Meaning and the Structure of Language*. University of Chicago Press, Chicago.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. University of California Press, Berkeley.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Foris Publications, Dordrecht.
- \_\_\_\_\_. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Linguistic Inquiry Monograph 6. MIT Press, Cambridge.
- Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. University of Chicago Press, Chicago.
- Culicover, Peter W., and Wilkins, Wendy. 1986. Control, PRO, and the projection principle. *Language* 62.1, 120-53.
- Emeneau, M. B. 1951. *Studies in Vietnamese (Annamese) Grammar*. University of California Publications in Linguistics 8. University of California Press, Berkeley and Los Angeles.
- Egerod, Søren. 1957. Essentials of Shan phonology and script. *Bulletin of the Institute of History and Philology, Academia Sinica* 29, 121-29.
- Fillmore, Charles J. 1968. The case for case. Bach, Emmon, and Harms, Robert, eds., *Universals of Linguistic Theory*. Holt, Rinehart and Winston, New York.
- \_\_\_\_\_. 1971. Some problems for case grammar.
- \_\_\_\_\_. 1972. The case for case reopened. Cole, Peter, and Sadock, Jerrold, eds. *Syntax and Semantics 8*. Academic Press, New York.
- Grierson, G. A. 1920. Vocal harmony in Karen. *Journal of the Royal Asiatic Society*, 347-8.
- Haas, Mary R. 1964. *Thai-English Student's Dictionary*. Stanford University Press, Stanford.
- Haudricourt, André-Georges. 1946. Restitution du Karen Commun. BSL 42

- \_\_\_\_\_ 1953. À propos de la restitution du Karen Commun. BSL 49
- \_\_\_\_\_ 1975. Le système des tons du Karen Commun. BSL 70
- Hoekstra, Teun, van der Hulst, Harry, and Moortgat, Michael. 1980. *Lexical Grammar*. Foris, Dordrecht.
- \_\_\_\_\_ 1980. Introduction. In Hoekstra et al. 1980, 1-48.
- \_\_\_\_\_ 1981. *The Scope of Lexical Rules. Linguistic Models 1*. Foris, Dordrecht.
- Huang, C.T. James. 1984. Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association* XIX.2, 53-78..
- Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge.
- Jones, R. B. 1961. *Karen Linguistic Studies* (University of California Publications in Linguistics v. 25). University of California Press, Berkeley.
- Kayne, R. S. 1984. *Connectedness and Binary Branching*. Foris, Dordrecht.
- Lehman, F. K. 1967. Kayah society as a Function of the Shan-Burman-Karen context. in *Contemporary Change in Traditional Societies*, Steward, Julian H., ed., 1-104. University of Illinois, Urbana.
- \_\_\_\_\_ 1979. Who are the Karen, and if so, why? Karen ethnohistory and a formal theory of ethnicity. in *Ethnic Adaptation and Identity: the Karen on the Thai frontier with Burma*, Keyes, Charles F., ed., 215-253. Insitute for the Study of Human Issues, Philadelphia.
- Lieber, R. 1980. *On the Organization of the Lexicon*. Ph.D. dissertation, MIT.
- \_\_\_\_\_ 1983. Argument linking and compounds in English. *Linguistic Inquiry* 14, 251-285.
- Marantz, Alec P. 1984. *On the Nature of Grammatical Relations*. Linguistic Inquiry Monograph 10, MIT Press.
- Matisoff, James A. 1973. *The Grammar of Lahu* (University of California Publications in Linguistics v. 75). University of California Press, Berkeley (second printing: 1982).
- \_\_\_\_\_ 1986. The languages and dialects of Tibeto-Burman: an alphabetic/genetic listing, with some prefatory remarks on ethnonymic and glossonymic complications. in McCoy, John and Timothy Light, eds., *Contributions to Sino-Tibetan studies* (Cornell Linguistic Contribution v. 5). E. J. Brill, Leiden.

- Mazaudon, Martine. 1985. Proto-Tibeto-Burman as a two-tone language? some evidence from proto-Tamang and proto-Karen. in Thurgood et al., 201-29).
- Noss, Richard B. 1964. *Thai Reference Grammar*. Foreign Service Institute, Washington, D.C.
- O'Connor, M.C. work in progress. (Ph.D. dissertation, University of California, Berkeley).
- Okeil, John. 1969. *A Reference Grammar of Colloquial Burmese* (Part I and II). Oxford University Press, London.
- \_\_\_\_\_. 1971. *A Guide to the Romanization of Burmese*. Royal Asiatic Society of Great Britain and Ireland, London.
- Perlmutter, D.M. 1978. Impersonal passives and the unaccusative hypothesis. in J. Jaeger et al., eds. *Berkeley Linguistics Society* 4, 157-89.
- Roeper, T., and Siegel, M. 1978. A lexical transformation for verbal compounds. *Linguistic Inquiry* 9, 199-260.
- Rothstein, Susan D. 1985. *The Syntactic Forms of Predication*. Ph.D. dissertation, MIT, reproduced by Indiana University Linguistics Club, Bloomington.
- Searle, John R. 1975. A taxonomy of illocutionary acts. Keith Gunderson, ed., *Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science*. University of Minnesota Press.
- Selkirk, E. 1981. English compounding and the theory of word structure. in Hoekstra et al. 1981, 229-277.
- \_\_\_\_\_. 1982. *The Syntax of Words*. MIT Press, Cambridge.
- Smalley, William A. 1962. *Outline of Khmu' Structure* (American Oriental Society essay 2). American Oriental Society, New Haven.
- Thompson, Sandra A. 1973. Resultative verb compounds in Mandarin Chinese: a case for lexical rules. *Language* 49.2, 361-79.

- Thurgood, Graham, Matisoff, James A., Bradley, David. 1985. *Linguistics of the Sino-Tibetan Area: the state of the art* (papers presented to Paul K. Benedict for his 71st birthday). Pacific Linguistics Series C, No. 87 (special number). Australian National University.
- van Riemsdijk, Henk, and Williams, Edwin. 1986. *Introduction to the Theory of Grammar*. The MIT Press, Cambridge and London.
- Williams, Edwin. 1980. Predication. *Linguistic Inquiry* 11, 203-38.
- \_\_\_\_\_. 1981. Argument structure and morphology. *The Linguistic Review*, 1.2, 81-114.
- Zubizarreta, Maria Luisa. 1985. The relation between morphophonology and morphosyntax: the case of Romance causatives. *Linguistic Inquiry* 16.2, 247-89.