CONSTRUCTION MARKERS AND SUBGROUPING OF FORMOSAN LANGUAGES[1]

RALEIGH FERRELL

1. TOPICALISATION IN FORMOSAN LANGUAGES

Obligatory sentence focus (topicalisation) is basic to all Formosan aboriginal languages. The verb in all instances is inflected to show the role of the topicalised Noun Phrase (such as agent, object/goal, locus, instrument, cause, beneficiary). The topicalised NP itself is then marked by either (1) an over phrase-introducing particle (Construction Marker), or (2) a fixed-order function slot for the NP, or (3) both of these syntactic devices. These devices, whether overt CMs or constituent NP order, are equational: the same device which identifies a NP as being in primary relationship with the Verb Phrase also equates NPs in non-verbal (equational) sentences.[2]

Non-topicalised NPs are also marked by either CMs or constituent NP order. Typically, all non-topicalised NPs may be marked by identical,

non-equational CMs, except that in languages having a discreet genitive/partitive marker the latter is used to mark the agent/actor in non-Agent Focus sentences.[3]

In certain Formosan languages, the topicalisational CMs perform a dual role by being differentiated to indicate proximity or specificity as well as the focus/non-focus roles outlined above. In at least one language, focus/non-focus roles are marked strictly by fixed NP order; and CMs, which obligatorily introduce each constituent NP, function solely as articles indicating relative proximity or specificity. These unusual developments distinguish these languages from other western Austronesian ones, and heighten the interest of Formosan languages to comparative Austronesian syntactic studies.

2. TOPICALISATIONAL VERB INFLECTION

The primary importance of the verb inflections in Formosan focus (topicalisation) system is attested to by the widespread preservation of cognate forms for these inflections. This is in contrast to the variability and relative instability of overt CMs and other NP-marking devices connected with topicalisation.

In Formosan languages generally the various roles which may be filled by topicalised NPs are subsumed structurally under four verbal inflections, herein referred to as AF (Agent Focus), OF (Object Focus), RF (Referent Focus), and IF (Instrument Focus).[4] Figure 1 shows topicalisational verb inflections in various Formosan languages.

	Amis	*Atayal*	*Bunun*	*Favorlang*	*Paiwan*	*Pazeh*	*Saisiat*	*Sediq*	*Siraya*	*Thao*	*Yami*	*Tsou*	*Saaroa*	*Rukai*
AF	/m/	/m/	m-	/m/	/m/	mV-	/um/	/m/	/m/	/m/	/m/	/m/	/um/	w-
OF	-ən	-un	-un	-en	-ən	-un	-Vn	-un	-en	-in	-ən	-i(ʔ)	(ʔ)	ki-(ʔ)
RF	/m/+-an	-an	-an	-an	-an	-an	-an	-an	-an	-an	-an	-a(ʔ)	-a(na)	ta-+-a(n)
IF	sa-	sə-	-is	i-	si-	si-+-i	si-	sə-	(ʔ)	sa-	i-	-(n)eni	-(a)ɬa	sa-+-a(n)

Figure 1. TOPICALISATIONAL VERB FOCUS INFLECTIONS IN FORMOSAN LANGUAGES

	3-CM Systems						2-CM Systems			
	Amis	*Atayal*	*Kuvalan*	*Paiwan*	*Siraya*	*Yami*	*Puyuma*	*Sediq*	*Rukai*	*Saaroa*
CM=	o	Ø	a	a	ta	u	a	ka	ka	ka
CM≠	to	sa	ta	tua	tou	su	ḍa	Ø	sa	na
CMgen	no	na	na	nua	ki	nu				

Figure 2. GENERAL FOCUS CMs IN SOME FORMOSAN LANGUAGES

3. TOPICALISATIONAL MARKING OF NOUN PHRASES

3.1. *3-CM SYSTEMS*

Amis, Kuvalan, Paiwan, Siraya, Yami, Pazeh, Thao, Favorlang and Ciuli-Atayal each have three topicalisational NP Construction Markers. In these 3-CM systems, the topicalised NP is introduced by an equational CM (CM=), the agent of non-Agent Focus sentences is introduced by the genitive/partitive CM (CMgen), and all other non-topicalised NPs are marked by a non-equational CM (CM≠).

The following examples from Paiwan show how 3-CM systems typically work. Note that in non-Referent Focus sentences, locative NPs may be introduced by secondary, CM-like markers[5] such as the Paiwan i *'at, in'*, which may either co-occur with the CM≠ or may result in deletion of the latter.[6]

'(The) man hunts wild-pigs in the mountains with a spear':

AF	q/m/alup	a	caucau	tua	vavuy	{tua / i tua / i}	gadu	tua	vuḷuq
	hunts	CM=	*man*	CM≠	*pig*	CM≠	*mtn*	CM≠	*spear*
OF	qalup-ən	nua	caucau	a	vavuy	{tua / i tua / i}	gadu	tua	vuḷuq
		CMgen		CM=		CM≠		CM≠	
RF	qalup-an	nua	caucau	tua	vavuy	a	gadu	tua	vuḷuq
		CMgen		CM≠		CM=		CM≠	
IF	si-qalup	nua	caucau	tua	vavuy	{tua / i tua / i}	gadu	a	vuḷuq
		CMgen		CM≠		CM≠		CM=	

Saaroa must also be included among the languages having 3-CM systems. Although Saaroa has only two overt topicalisational CMs (CM= and CM≠), structurally the equivalent of a 3-CM marking system is maintained by assigning the function of a CMgen to a Ø-marker. The following examples are illustrative:

m-aci?i	ka	caciḷi
die	CM=	*one person*

'one person died'

k/um/ita	na	aḷaḷiamu
see[AF]	CM≠	*fly*

'(they) saw a fly'

saḷa?a takoḷu
trail wildcat
'the wildcat's trail'

Squliq-Atayal, on the other hand, has only **CM≠** and **CMgen.** In this language, the role of **CM=** is filled by lack of an overt **CM** before the topicalised **NP**:[7]

?malup squliq sa bizuak
hunt man **CM≠** *pig*
'the man hunts the pig'

ŋasal na squliq
house **CMgen** *man*
'the man's house'

3.2. *2-CM SYSTEMS*

Bunun, Puyuma, Saisiat, Rukai and Sediq each distinguish by overt marking only two categories of **NP**. The 2-**CM** systems lack a separate genitive/partitive **CM** but differ from Squliq-Atayal in that all non-topicalised **NP**s, including the agentive **NP** in non-AF sentences, are marked by identical non-equational devices (see Figure 2). The following examples are from Puyuma:

t/əm/bəl a marəwadi ḍa ṭau
bury[AF] **CM=** *brothers* **CM≠** *people*
'the brothers buried the people'

romaʕ ḍa ṭau
house **CM≠** *man*
'the man's house'

Thus whereas Saaroa and Squliq-Atayal each have only two overt **CM**s but mark by **Ø** the categories of **NP** introduced in Paiwan respectively by **CM=** and **CMgen**, conversely in 2-**CM** systems two separate categories of **NP** are overtly marked by identical **CM**s. Constituent **NP** order then must play a more important role in 2-**CM** systems than in 3-**CM** ones to indicate the case-like roles of the various non-topicalised **NP**s.[8]

Tsou also appears to have a general 2-**CM** system. But as will be discussed in 3.4, Tsou has separate **CM**s in each of the two categories to show several degrees of proximity or specificity and warrants separate treatment.

3.3. *Ø-CM SYSTEMS*

In Kanakanabu, there are no overt topicalisational **CM**s; all topicalisational **NP** marking is accomplished by constituent **NP** order:

ni-macay cau
died man
'the people died'

ivatu vavuru k/um/aənə vuruŋa
come pig eat[AF] *eel*
'the pig came and ate the eel'

Aside from fixed NP order, one way in which topicalisational ambiguity is avoided in Kanakanabu is by a preference similar to that of Squliq-Atayal (see 3.1) for non-AF sentence construcions where non-agent NPs are present in the sentence.

Maga and the other two so-called "Lower Three Villages Rukai" languages[9] also achieve topicalisational NP marking strictly by NP order. However, as will be seen in 3.4, these languages do have obligatory NP-introducing particles which fulfil quite different functions from topicalisational CMs.

3.4. *GENERAL AND SPECIFICATIONAL CMs*

Besides the general topicalisational CMs discussed thus far, two additional types of specificational CMs are found in Formosan languages. These two types of specificational CMs indicate respectively (1) proximity or definiteness, and (2) named individuals.

Atayal, Puyuma and Rukai have parallel sets of general and definite (or proximal) CMs. The latter indicate either that the NP is nearby or gives it a specificity similar to that expressed by the definite article in European languages:

	Squliq-Atayal		*Puyuma*		*Rukai*	
	General	Specific	General	Specific	General	Specific
CM=	ø	qo·	a	na	ka	ku
CM≠	sa	s-qo·	ḍa	ka-na	sa	ki
CMgen	na	na-qo·				

Tsou, on the other hand, is said to have a complex system of topicalisational CMs which simultaneously indicate focus relationships and relative proximity for each NP:[10]

	In Sight			*Out-of-sight*			
	Near	Away	Remote	Known	Hearsay	Speaker sees, Hearer not	Hearer sees, Speaker not
CM=	e	si	ta	o	na	yao	eyo
CM≠		ta		to	no		

Maga and the other two "Lower Three Villages Rukai" languages have CM-like specificational particles which, although they have nothing to do with topicalisation, are obviously related to the specificational topicalisational CMs of other Formosan languages. These obligatory NP-introducing particles in Maga are:

na (visible)
ku (out-of-sight)

Thus, in the following Maga examples topicalisational NP marking is effected by NP order, and the obligatory CM-like particles serve roughly as definite/indefinite articles:

latbi na vḷakə
weep *child*
'the child weeps' (and the child is visible)

latbi ku vḷakə
weep *child*
'the child weeps' (and the child is out-of-sight)

The second type of specificational CMs, found in Amis, Atayal, Bunun, Kuvalan, Paiwan, Puyuma, Siraya, Yami, Saisiat and Favorlang, indicate that the NP represents a named individual. Cognate CMs fulfilling the same functions are common among Philippine languages. Examples from Paiwan and Squliq-Atayal are:

	Paiwan	*Squliq-Atayal*
CM=(pers)	ti	i
CM≠(pers)	tʸai	∅, -an
CMgen(pers)	ni	ni

4. SYNTACTIC, LEXICAL AND PHONOLOGICAL SUBGROUPING OF FORMOSAN LANGUAGES

A tripartite subgrouping of Formosan languages (Atayalic, Paiwanic, Tsouic) has been suggested by lexicostatistical studies (see Dyen 1971). On both lexical and phonological grounds there is little doubt concerning the close genetic relationship between Squliq-Atayal, Ciuli-Atayal and

Sediq (the Atayalic subgroup). Tsou, Kanakanabu and Saaroa also cluster well to form the Tsouic subgroup; it is likely that Rukai and the "Lower Three Villages Rukai" are also Tsouic languages.[11] All the other Formosan languages are thus far considered to fall into the catch-all Paiwanic grouping; comparative phonological work has not yet progressed to the point that further subgrouping can be done with any degree of confidence. Nor have the relationships between the three major subgroups been determined.

I have noted (Ferrell 1972) apparent confirmation of the three-way subgrouping of Formosan languages from the standpoint of relatively superficial differences in transitive verb syntax. It appears, however, that some syntactic processes such as conjunct verb constructions may link Tsouic and Paiwanic somewhat more closely than either of these relates to Atayalic.

The present study has indicated the relative instability of topicalisational NP-marking devices. It will be seen from Figure 3 that considerable flexibility is possible in the actual marking mechanism for NPs in focus constructions (3-CM, 2-CM and Ø-CM systems) without serious consequences to the basic focus system itself, and the choice of one or the other of these systems cuts across all subgroupings established by lexical and phonological comparison. Nor does geographical distribution indicate any strong area influences.

Thus the fact that Thao and Saaroa both have 3-CM systems and lack person name CMs, whereas each of these languages is surrounded by respectively more closely related languages which do have the latter, suggests that the similarities between Thao and Saaroa are probably due to independent simplification, rather than to genetic or area influences.

The number and form of general topicalisational CMs in Formosan languages then are of little value in establishing genetic subgroupings. The two types of specificational CMs (see 3.4), on the other hand, may have more important implications in that their occurrence in various languages can not be the result of parallel simplification, and independent invention is not particularly likely. The non-occurrence of these specificational CMs in given languages may be due to independent simplification; but their occurrence most likely indicates either genetic or area relationships.

Separate topicalisational CMs for use with personal names are widespread in western Austronesian languages. In Formosa they are found in Atayal and all the Paiwanic languages except Pazeh and Thao. They are conspicuously absent from the Tsouic languages (see Figure 3).

Discreet topicalisational CMs indicating proximity or definiteness are found in all three major Formosan subgroups (Atayalic, Paiwanic,

Tsouic), although in Paiwanic they are found in only a single language (Puyuma).

	Personal Name CMs	General Focus CMs			Specificational (Definite) CMs
		3-CM System	2-CM System	Ø-CM System	
Amis	+	+			
Bunun	+		+		
Favorlang	+	+			
Kuvalan	+	+			
Paiwan	+	+			
Pazeh	+	+			
Saisiat	+		+		
Siraya	+	+			
Thao	(?)	+			
Yami	+	+			
Puyuma	+		+		+
Atayal	+	+			+
Sediq			+		
Saaroa		+			
Kanakanabu				+	
Maga-Rukai				+	+
Rukai			+		+
Tsou			+		+

Figure 3. DISTRIBUTION OF CM-TYPES IN FORMOSAN LANGUAGES

There appears to be little likelihood of area influence explaining the occurrence of these CMs in both Atayalic and Tsouic languages. Parallel

innovation is a possibility, but the simplest assumption is that they are a common inheritance from a proto-language. Their loss in some languages from each group, viz. (Atayalic) Sediq and (Tsouic) Saaroa and Kanakanabu, could then be due to independent simplification. This could also be the case for the Paiwanic languages individually, although their near-universal absence in Paiwanic languages might rather indicate that their absence represents a shared innovation of the Paiwanic languages as a group. Their occurrence in (Paiwanic ?) Puyuma, on the other hand, either (1) strengthens the case for the existence of these CMs in Proto-Formosan, or (2) must be attributed to independent invention (if Puyuma is in fact a Paiwanic language), possibly as a result of area influence.

In summary, the occurrence and specific number of general focus CMs in Formosan languages can tell us little regarding either genetic relationships or language contacts. The occurrence of special (definite/specificational and Personal Name) CMs, on the other hand, may be considered to support the three-way subgrouping of Formosan languages in that (1) the Tsouic subgroup is marked by the absence of Personal Name CMs, and (2) the Paiwanic subgroup [with the unexplained exception of Puyuma] is marked by the absence of definite/specificational CMs. Meaningful genetic subgrouping of present-day Formosan languages on the basis of CMs alone would not be feasible, however, since some languages (Atayal and Puyuma) have all three types of CMs whereas other languages from different subgroups (e.g. Sediq, Saaroa) lack both specificational and Personal Name CMs.

5. IMPLICATIONS FOR PROTO-FORMOSAN SYNTAX

The universal distribution of obligatory sentence topicalisation in present-day Formosan languages suggests that topicalisation by focus inflection was characteristic of whatever proto-language was ancestral to all these languages.[12] In this Proto-Formosan focus system, as in the modern languages, verb focus inflections indicating the case-like roles of topicalised NPs were of primary importance.

As in the modern languages generally, topicalised NPs were probably marked by overt CMs. Present data suggest the possibility that CMs in the proto-language either served the dual function of topicalisational NP marking and indicating general/specific or near/remote distinctions, or that there were separate sets of CMs performing these functions.

NOTES

1. This paper was presented at the First International Conference on Comparative Austronesian Linguistics, Honolulu, 1974. Grateful acknowledgement is given to the Centre National de la Recherche Scientifique, Paris, under whose auspices I carried out ethnographic and linguistic fieldwork in Taiwan in 1968-70, and to the National Science Foundation for grant #GS-28818 for linguistic fieldwork in summer 1971.

2. Compare examples from Paiwan, in which this equational relationship is indicated by a CM, and Atayal in which it is shown by phrase order:

Paiwan

S VP + NP k/m/an a alak *(eats* CM *child) '(the) child eats'*

S NP + NP alak a vavaian *(child* CM *woman) 'daughter; (the) child is female'*

Atayal

S VP + NP m-aniq laqi *(eats child) '(the) child eats'*

S NP + NP laqi knairil *(child woman) 'daughter; (the) child is female'*

3. Other types of CMs, such as those introducing relative, coordinate or subordinate clauses, those marking verb/topic transposition (and thus additional emphasis on the topic), those showing genitive/partitive relationships, etc., are mentioned here only where they bear upon topicalisation.

4. In Paiwan, for example, inflection of the verb for AF shows that the topicalised NP is agent or actor; OF may indicate direct object, patient or goal; RF may be locus, beneficiary or indirect object; and IF may be instrument or cause. For obvious semantic reasons individual verbs may be "defective", i.e. not susceptible to inflection for topicalisation for one or more of these potential roles.

5. These prepositional CM-like markers should probably be considered to be conjunct verbs; cf. my communication to the First International Conference on Comparative Austronesian Linguistics, Honolulu, 1974, "Conjunct Verbs and Verb-Object Incorporation in Formosan Languages".

6. In such cases, inclusion of the CM≠ gives relative specificity to the locational NP, whereas deletion of the CM≠ makes the locative NP more general; in other words, when the CM≠ co-occurs with the "prepositional" particle, it is semantically comparable to the definite article in English.

7. Note that in Squliq-Atayal, although the general CM≠ (sa) exists its use is in fact rare due to an apparent preference for non-AF constructions when object, referent or instrument NPs are included in the sentence.

8. For instance, to differentiate agent from beneficiary in non-AF sentences.

9. The relationship between Rukai "proper" and the so-called "Lower Three Villages Rukai" is problematic. Whether the latter, i.e. "Maga" (Tərdkanuu), "Tona" (Koŋadavanu) and "Mantauran" (Opunoho) are actually three separate languages or dialects of a single language is also undecided.

10. This follows Ogawa and Asai's (1935) analysis.

11. Lexicostatistically Rukai appears to be closer to Paiwan. However, the propinquity and close cultural ties between the Paiwan and Rukai suggest heavy borrowings; and structural considerations appear at this point to indicate closer genetic ties between Rukai and Tsouic.

12. "Proto-Formosan" in this paper is not meant to infer that all Formosan languages form a single subgroup of Austronesian. Comparative research has not yet provided a decisive answer to this question. Proto-Formosan here means simply whatever ancestral language all Formosan languages did share in common. Thus Proto-Formosan might conceivably turn out to mean, for example, Proto-Formosan-Philippine, Proto-Western Austronesian, or even Proto-Austronesian.

REFERENCES

DYEN, Isidore

1971 'The Austronesian Languages of Formosa'. In: Thomas A. Sebeok, ed. *Current Trends in Linguistics*, vol. 8, *Linguistics in Oceania*. The Hague.

FERRELL, Raleigh

1972 'Verb Systems in Formosan Languages'. In: Jacqueline M.D. Thomas and Lucien Bernot, eds *Langues et Techniques, Nature et Société*. Paris.

OGAWA Naoyoshi and Erin ASAI

1935 *Taiwan Takasagozoku Densetsu-shū* (Myths and Traditions of the Formosan Native Tribes). Taipei: Taihoku Imperial University.