General Diachronic Course of Proto-Austronesian Casemarkers

Prepositions and Topic Markers: Sources of Austronesian Casemarkers

Joseph C. Finney

Overview

This is an exploration in Diachronic Natural Language Syntax (DNLS). It moves toward reconstructing what the spoken natural language may have been like in PAN, possibly 6000 years ago. What it attempts to reconstruct is the Surface Structure, including the word order. It deals with syntax on the level used in reference grammars, such as Mosel and Hovdhaugen's (1992) Samoan Reference Grammar, and the reference grammars of Philippine and Micronesian languages produced by the University of Hawai'i.

Abbreviations include P proto, AN Austronesian, MP MalayoPolynesian, PN Polynesian. I shall sometimes use "PCM" for "prepositions and casemarkers" or even "prepositions, topic markers, and casemarkers". Why not include "topic" in the set of cases? Partly because the development of topic markers into Nominative or Absolutive case markers has some analogy to the development of space prepositions into abstract casemarkers. Note, too, that topics, like vocatives and exclamations, are not syntactically within a clause. Latin grammars include Vocative as a "case", but the ancient grammarian Panini considered Sanskrit to have only seven cases, not eight, as Vocative, not being in a clause, does not enter into syntactic relations. I like Li's concept that prepositions are one type of casemarker, though it is convenient to distinguish prepositional phrases sharply from Det phrases.

In his Ph.D. dissertation done under Starosta, Li (1973: 111) remarked, "Prepositions, determiners, personal pronouns, verbal formations, and occasionally word order...are case markers in Rukai." He added (op. cit., 116): "[P]reposition (P) is not a sister category of noun (N). Of the four main case forms treated in this study, two of them are marked on P's, while the other two are marked on N's. P's are marked for the two main case forms: [+L] and [+I], and N's are marked for these two:

[+NM] and [+AC]." [This is a distinction based on the difference between a PP, prepositional phrase, which modifies a verb or noun (and is thereby adverbial or adnominal/adjectival), and DetP, determiner phrase, which is a substantive.] This system allows us to regard, for example, Mayrinax **ki** "to, toward, into" as a preposition, without preventing it from being a "casemarker" in paradigmatic tables in which its position is analogous to that of Nominative casemarkers, which cannot be called prepositions. So, in this paper, whenever I distinguish "prepositions" from "casemarkers", read that as distinguishing casemarkers that are prepositions from casemarkers that are not prepositions.

The concept of "topic" used here is much the same as that used by Foley and Van Valin (1984), as follows. "Pr[agmatic]P[ivot]s in at least some languages developed out of earlier topics. But...[p]ivots are clause-internal, whereas topics occur external to the clause... Topics are normally set off from the rest of the sentence by a pause...Topics play no such central role in the syntax of any language...; they seem instead to be an adjunct to a sentence rather than an integral part of it..."

From the situation prevailing in conservative languages it is the consensus that PAN had verb-initial word order. It is precisely because the topic is **not** part of the clause that it was allowed to come first.

A general principle of diachronic Universal Grammar is that particles of Locative usage (at, in) acquire Ablative usage (from); and particles of Ablative usage acquire Genitive usage (of); and perhaps Locatives can acquire Genitive usage directly without first acquiring Ablative usage.

As I shall use capital letters **C** and **V** for generalized consonant and vowel, I'll use a lower-case **c** for the reconstructed PAN consonant usually written *C. Dyen had used the lower-case **c** for another phoneme, but that shouldn't cause a problem, as neither Ross's (1992) nor Wolff's (1995) sound chart uses the symbol *c any more.

The general scenario proposed is as follows. The ancient markers were light CV syllables. The vowels were **a** for nonspecific common nouns, and **i** for personal pronouns and proper names of persons, and the default value was **u**.

Consonants were zero, **n, k, d, s**, and both **c** and **t** (which fall together as **t** in MP).

Of the reconstructed prehistories of casemarking, the most clear cut are those of the space prepositions, which were: ablative **nV**, locative **V**, allative **kV**, locative **dV**, allative **tV**, locative **cV**, and locative and general oblique **sV**. Every one of those has living languages today that preserve the original space usage that they had in PAN. Every one of them, of course, has developed new usages in some languages. The most conservative of all is the **nV** particle which consistently glosses 'of', 'from', and (agentive) 'by'. (When major syntactic reanalysis changes Ergative case to Nominative, as it did in Proto-Central Eastern MP, the casemarked **na** pronoun from **nV** + demonstrative **a** becomes Nominative in some contexts.)

PAN had phrases that occurred at the left edge of the sentence and were not in any clause. Those included topics, vocatives, and exclamations. At first they were marked by a particle only on their right edge, separating them from the clause. We trace the complex process by which (in many of the daughter language groups) topics acquired markers on their left, were moved to the right of the sentence, served briefly as clarifying afterthoughts, and then were reanalyzed as Nominatives (including Absolutives). There were two generations of markers of Topic (and specificity). An earlier wave of Topic becoming Nominative was with **kV**, and a later wave used plain **V** (zero consonant).

Comparison of sentence-constructions of Amis with those of Malagasy shows the change caught in the act. Framing of the topic between demonstratives took place before the movement took place (preserved in Amis) and persisted at least briefly after the movement (preserved in Malagasy). The sequence was (1) framing; (2) movement of the Topic phrase from the left to the right edge of the clause; (3) inclusion of the former Topic into the clause and reanalyzing it as the Nominative/Absolutive and as Spec,IP; (4) loss of the sentence-final Topic marker; and (5) relaxation of the requirement to keep the (new) Nominative/Absolutive sentence-final (as VOS has less optimality than VSO).

Successive Steps in the Development of Casemarking from PAN to PMP

Though we'll focus on specific parts of it, I'll give here, as a framework, my understanding of the comprehensive

course of theory of case-marking ancestral to the MP division, the group that includes some 90 percent of the AN languages. From study of the evidence, this is what I think happened. Perhaps this set of hypotheses will serve as a stimulus to scholars who may prove some, disprove others, and arrive at a more definitive synthesis.

1. At a very early time in PAN, the language had a set of monomoraic particles, each a single light syllable consisting of a consonant + vowel **u**. The consonants were zero, **n**, **k**, d_1, **s**, and both **t** and **c**. Thus the set was: **u, nu, ku, d_1u, su, tu** and **cu**. All of them served as prepositions of space relations such as "in", "at", "to" and "from" with perhaps additional functions. The variants with vowels **a** and **i** came later, with the possible exception of the locative particle, which may have always been **i**.

2. PAN had a particle **i**, which must have been an article or classifier used before personal pronouns and proper names of persons. The sequence **Cu i** contracted to **Cui** and then **Ci**. A **Cui** form survives in Chamorro as **nui**.

3. PAN had a particle **a**, which must have been an article or classifier for nonspecific common nouns. The sequence **Cu a** contracted to **Cua** and then **Ca**. Two **Cua** forms survive in Paiwan as **nua** and **tua** < *sua. A **Cua** form survives in Kanakanavu as **sua**.

4. Thus a paradigm developed of **Ca, Cu, Ci**, with some of the consonants. While some consonants have gone through steps 1, 2, and 3, others besides locative **i** may have begun with a lexically specific vowel, and then undergone reanalysis into the three-vowel paradigm.

5. All these particles, prepositional casemarkers (PCMs), were capable of being grammaticalized into non-prepositional casemarkers. They did so in varying time frames in various languages.

6. If there was a PCM **cu** it was homophonic with the demonstrative pronoun **cu** (whose consonant is known to be PAN **c** because it is **tsu** in Paiwan). Like all demonstratives, PAN's had distality (from "this" to "that"). PAN had six such demonstratives, **ni, d_1i, a, cu, na**, and **d_1a**, and probably also **ta** and **ti**, as we discussed in last year's SEALS paper.

7. When followed by nouns, the prepositions and casemarkers remained independent words. But when followed by demonstrative pronouns the PCMs coalesce with them to form bimoraic casemarked demonstrative pronouns.

8. The PCMs vary in the degree to which we can confidently assign semantic roles (and for some a syntactic one) to their PAN ancestors. The particle with the clearest role is **nV**. Its earliest reconstructable use was as ablative preposition "from". As the concrete space "from" spread to abstract (cause) "from", two syntactic usages branched off from it. One was the genitive, which in English is expressed by an "of" phrase or by the possessive case. The other was the non-subject agent, which in English is one use of the "by" phrase. This phrase is used in AN languages in passive constructions as the equivalent is used in English. In ergative constructions, in which the ergative agent marker becomes a Determiner and the ergative increasingly acquires characteristics of a subject, the **nV** agent marker is no longer a preposition heading a PP but is the ergative casemarker heading a DetP.

9. In some languages the phrase marked **nV** is used both as a genitive and as an agent PCM, often without a "from" space use. But genitive and agentive are related only indirectly, with both coming down from "from" phrases which denoted first space movement and then abstract source. It is no secret that space relations are the great well-spring from which other relational terms arise (and I found myself doing so in this very sentence, too).

10. There can be no doubt that MP languages have prepositions-casemarkers of **tV** form. The question is whether in AN they were **cV** or **tV** or some of each. If all the **tV** PCMs in MP languages were cognate with Proto-Atayal **cV**, which is ancestrally a preposition, PAN must have had a preposition **cV**, but the Atayal **tV** could be a local innovation. On the other hand, if all the **tV** PCMs in MP languages were cognate with Proto-Atayal **tV**, which is ancestrally a preposition, PAN must have had a preposition **tV**, but the Atayal **cV** could be a local innovation. But it turns out that Formosan languages show both **tV** casemarkers and space-prepositions, and **cV** prepositions and casemarkers. There is no reason to believe that either ancestral line died out in Proto MP. So, till proven otherwise, the reasonable belief is that the ancestral **tV** and **cV** prepositions or casemarkers merge in Proto MP.

11. The PAN *tV is reconstructed because there are **t**- forms, notably Wulai **te** allative preposition, that are not accounted for by PAN *cV. So the *tV is likely as a PAN preposition.

12. In accordance with universal diachronic principles, the PAN PCM *cV is found in (1) locative preposition **ci** in Wulai Atayal; (2) locative preposition **ca** in Mayrinax Atayal; (3) accusative (and dative) casemarker **cu** in Mayrinax; (4) very

likely Wulai's **sa**, the PCM for the "dative" or "locative" case in Wulai Atayal. Tsou's puzzling "accusative" marker **n-ca** (Starosta 1974), or **nin-ca** (Szakos 1994:83) could conceivably be a CV PCM. But the order of the morphemes suggests that **n** or **nin** is a PCM and **ca** a Determiner (possibly a vowel variant of the **cu** demonstrative that is **tu, to,** and **ta** in Western MP languages and is an article **ta, te, ti** in Micronesian and Polynesian languages). Further, (5) the Tsou **ci** has so thoroughly moved from locative to genitive that (like Tagalog and Old Javanese **ng** < **nV**) it is used essentially as a "linker" (a concept I'll deal with elsewhere).

13. The **CV** PCMs with **k, s, d, c,** possibly **t,** and zero consonant seem to occur both with Nominative-like and with Case 3 syntax (locative, allative, accusative), each perhaps for its own reason and with its own history.

14. A zero-consonant form, simple **V**, was ancestrally a locative preposition which glosses "at", "in", "on" in English.

15. The zero-consonant locative occurs in AN languages more often as **i** than as **a** or **u**, even though canonical locatives are places and not persons. That is why we must consider the likelihood that its earliest form was **i** for all classes of objects and that only later in PAN was it assimilated analogically into the **i, a, u** paradigm.

16. In language groups as far apart as East Polynesian and Rukai, the locative preposition **i** became the accusative casemarker. In doing so, it passed through what is the "look at" or "count on" construction in English, which has been called the PP Term. Samoan Polynesian shows that phrase approaching accusative semantics but still syntactically a PP and not as the accusative object it has become in East PN languages such as Hawaiian.

17. The most widespread use of preposition **kV** throughout the AN languages, and probably its earliest use, in PAN, was as an allative preposition, "to", "toward", which also had comitative "with" uses. In languages widespread in AN, it developed into a dative or accusative casemarker. Both those developments occur in other language families throughout the world. The **kV** that we shall see marked specifics or topics, and later nominatives, may be of another origin.

18. Two phonologically different particles, in different time frames, went through a development from topic markers or markers of specificity (or both) to nominative or nominative-like use. The first to do so was the quasi-nominative *****kV**. The other was the quasi-nominative *****zero+V**. The latter has somewhat more often combined with demonstratives.

19. Both Li (1995) and Huang (1994:109) show a phonological distinction in Mayrinax Atayal between nominative-like, topic-like '**i**' (initial glottal) and locative **i**' (no initial glottal). But in all other languages the quasi-nominative has zero initial, and even in Mayrinax the "nominative" form for nonspecific common nouns is simple **a** with zero initial. Nobody has proposed phonologically different PAN ancestries for the pair. The Mayrinax glottal may be only prosodic in origin.

20. A paradigm found in Mayrinax and in Old Javanese shows that at some time in their common ancestry the **kV** particle was used (right of a PCM) to show specificity in any case form. Topics are characteristically specific, and one possibility is that **kV** was a marker of specificity before it was a topic marker. It may be more plausible that the Mayrinax and Old Javanese specificity marker included the marking of topics than that the topic marker came to be used to mark specificity for all syntactic cases.

21. Topics with either **kV** or zero**V** form occurred, as a topic should, at the left edge of the sentence. For the **kV** form, the evidence shows that in the beginning the marker's place was right of the topic, marking the border between the topic and the sentence. Later, a marker was also put at the left edge of the topic, and the topic could be said to be "framed" between the two. It was not necessarily the same marker. At least one language today (Bunun) can put a **kV** marker at either side and a zero**V** to the left of the topic. The zero**V** markers can be casemarked demonstratives of the form **i**+[demonst]; or can be a non-demonstrative determiner (article).

22. So long as the marked topic remained sentence-initial it was outside the clause and not related syntactically to the clause nor to anything in it. That extra-clausal slot was used not only for topics, but also for two other kinds of non-clausals: vocatives and exclamations. Those also were framable with identical zero+**V** markers. (In most Polynesian languages today, vocatives can still be framed by identical case-marking, **e** < *$*$**i**, left and right.)

23. It became possible to move the topic to the right edge of the sentence, where perhaps at first it served as a clarifying afterthought and then became integrated into the clause as its subject (Spec,IP). That made VOS word order. (One language today, Nataoran Amis, gives a choice of having the ancestral topic either sentence-initial, marked as Topic case with zero+**V**, or sentence-final, marked as Nominative case with **kV**.)

24. Most languages have abolished the right-marking of topics, especially when they are sentence-final. But at least two

languages today keep at least optionally the ancient practice of framing topics left and right with **CVCV** casemarked demonstratives (determiners) whose left half is **i-**. One is Malagasy, where the topic is obligatorily final and is the Subject [Spec, IP], and its right marker is optional in some constructions. The other is Nataoran Amis, where the right framer (also deletable) is an ancient proximal demonstrative and is fossilized for nothing but framing, while the left framer can be any other demonstrative or determiner that conveys specificity.

25. The movement of the (former) topic to sentence-final as afterthought, and its incorporation into the clause as Nominative or Absolutive subject, create an anomaly. Languages in general seem to have a strong tendency to put syntactic subjects, especially agent-subjects, to the left of objects and obliques (S before O). Most MP languages moved the Case 1 subject out of sentence-final position and became SVO. There are languages such as Fijian that prefer VOS.

26. One pattern, found in Wulai Atayal as well as Tagalog and other Central Philippine languages, has Case 1 (Nominative or Absolutive), Case 2 (Genitive or Ergative), and Case 3 (Oblique Dative, Locative). For AF (Agent Focus, accusative-like sentences with no Accusative case as such) Wulai puts the non-focus patient in Case 3 if it is a pronoun and not a common noun, but Tagalog always in Case 2 (that puts non-focus agent and non-focus patient in the same **n-** case form even when they occur in the same clause, with locative or other element in the Focus case.

27. One set of **sV** arose, at first **si** for third person singular human beings only, as the Determiner in Topic / Nominative / Focus case. Some languages came to have agent and patient marked alike, with **si**, for personal pronouns and names of persons. This tendency may have been related to the fact that nominatives and accusatives are substantives (DetP), in contrast to PP. Reid's impeccable work on the Central Cordilleran languages highlights the oblique **s**-marker. The **sa** Case 3 PCM must be PMP, as it occurs broadly in PM languages (even a Central Pacific one). Its occurrence in Paiwan as **tu, tua**, (though not ***ta**) shows that it must have occurred in PAN as **su + a > sua > sa**. So the Topic-Nominative and the Oblique **sV** forms may have arisen independently of each other and their occurrence may be coincidental. Yet more work needs to be done on the **si** particle to check the possibility that it might be **su + i > sui > si**.

28. The marker **dV** began in PAN as a preposition with locative use. In Philippine languages it competed with **kV** and with **sa**

as locative (Case 3) prepositional casemarker. In Indonesian-Malayan it served as the left preposition in embedded PP ("on top of"), as i did in Oceanic; and an "emphatic" nominative use arose somehow, possibly from honorific use of a plural form.
29. The analysis given here poses a very basic question. Are there any primary branches of AN whose ancestors did not have left topics with framing? If so, all the supposed primary branches that have such framing, or ancestrally had it, must constitute a single primary branch of AN.

Interaction of PCMs with Demonstratives

Here's how I (and some others) have heretofore viewed PCMs and demonstrative/determiners. My views have changed a bit with further study of the evidence.

PCM consonants:	PCM vowels:	Demonst/Det:
zero (at, in)	i (person,	ni, di (this, near me)
zero, k (topic)	proper name)	a (neutral)
n (from, of, by)	u (unmarked,	tu (neutral)
k (to, accompanying)	neutral)	na (that)
s (not Gen/Erg)	a (nonspecific	da (that, out of sight)
d (not Gen/Erg)	common noun)	
t (not Gen/Erg)		

It is curious that PMP had two sets of single-mora CV particles, and the two impinged on each other. One set was the prepositions and topic markers (if we may lump them together as potential casemarkers); the other was the demonstrative pronouns.

It is specially curious that the demonstratives and the potential casemarkers shared phonological restrictions and preferences. Neither set has a member that is labial (**p, b, m, w**). Both sets have members with initial zero consonant, and both are strongly represented by alveolars (**nV, dV, tV, cV**), the last being a "church" affricate. The casemarker set also includes **kV**, but not ***gV** nor a velar nasal. The casemarker set has an alveolar fricative or affricate (**sV**). (A common assumption is that PAN *S was the "ship" and PAN *s the "sip" consonant, though I think *S was [s] and *s was [ts].)

In conservative AN languages (as in Romance languages) prepositions or casemarkers have tended to amalgamate with determiners (weakened demonstratives), making CVCV words in AN, and then have undergone further lenition, while phrases of PCM plus Noun stay separate words.

Separate Origin of C- and -V
in Single-Mora Casemarkers

Figure 1 was developed for my paper at the 1998 SEALS meeting in Kuala Lumpur. It puts PAN demonstratives **ni**, **na**, **di**, and **da** in a simple and transparent logical order. No one language conforms a hundred per cent with the paradigm, but some come very close. (Cebuano Bisayan is one.) All the attested uses can readily arise from it. When we examine each row and each column separately, a surprisingly simple way to define the relation is that the vowel shows first-person distality (**-i** for "this, near me", and **-a** for "that, not near me") while the consonant shows second-person distality (**n-** near you, **d-** not near you). That analysis explains why **ni** is near us both, and **da** is far from us both; **di** is near me but not near you; and **na** is near you but not near me.

Figure 1

PAN reconstructions proposed for four demonstratives:

	near me	not near me
near you	**ni** this	**na** that
not near you	**di** this	**da** that

For PAN it is also possible to reconstruct other demonstratives, including *a, *Cu, *ta, *ti, and possibly others. Their semantic meaning in terms of distance or other dimensions (perhaps visibility) has not been clearly established.

How can such a situation come about? Obviously, one hypothesis to be considered is that each of these single-mora particles was earlier a pair of particles. The right particle began with zero consonant or one that was lost. The left particle lost its

vowel in amalgamating with the right particle. If that were not true, an alternate hypothesis could be that the correspondence of semantics and phonology came about through some sort of analogy.

As this approach is reasonable for the demonstratives, let's see if it can also work for the casemarkers and their ancestors, the PCMs.

When we check the more conservative languages in various primary divisions of AN, it is striking that a particular casemarker (or topic marker or preposition) occurs with vowels varying among **a**, **i**, and **u**. The vowels vary among languages, and also vary within a given language. When they vary within a language, the variation correlates with classification of the noun that is governed thereby.

If I may summarize the occurrences, it seems to be that vowel -**a** occurs with nonspecific common nouns (at least those not denoting persons), while vowel -**i** occurs with names of persons and with personal pronouns; and that vowel -**u** is the unmarked category (default category), occurring where neither the **a** rule nor the **i** rule applies.

The **Ca** particles (category 1) marked, as a minimum, nonspecific common nouns not denoting persons. The **Ci** particles (category 3) marked, as a minimum, personal pronouns denoting persons and proper names of persons. Noun phrases (DetP) not falling into either of those categories were marked (category 2) **Cu**. The **a** category could expand to include all common nouns not denoting persons, or all nonspecific common nouns, or even all common nouns. Or, the **i** category could expand to include kinship terms; or to include all nouns denoting persons. Or, any two adjacent categories (**Ca** and **Cu**; or **Cu** and **Ci**) could merge, selecting one of the two applicable vowels. Here is some evidence from conservative AN languages that suggests that generalization and justifies it.

Languages seem to vary in the extent to which the specific category markers -**a** and -**i** expand the scope of their use. Oceanic and other Eastern MP languages tend to generalize the use of vowel -**i** in their prepositions, topic markers and casemarkers.

Therefore, with the PCM set, just as with the demonstrative set, the most reasonable belief is that the ancestral monomoraic forms arise by amalgamation of pairs of monomoraic forms.

Specifically, I propose that the ancestral forms of the PCMs (ancestral topic markers and prepositions) were of form **Cu** (where C can be zero); and that their variants with other vowels were formed by amalgamation with particle *i for proper names of persons and particle *a for nonspecific common nouns. Put more specifically, **Cu** plus a became **Cua** and then **Ca**; and **Cu** plus i became **Cui** and then **Ci**.

The added particles **a** and **i** may have been former articles or noun-classifiers. If they were articles, they may have descended from demonstratives at a still earlier time.

As Zeitoun has noticed (manuscript 1996:9) the casemarker must be **ki** before proper names of persons in two Rukai languages, Tanan and Maga. That fits the generalization that the vowel -i was the one required for PCMs for this class of substantive. What doesn't fit is that it is for accusative in Tanan but nominative in Maga. With my hyphens inserted, her examples are:

Tanan:	ko-a-ni	ababay	wadamək	**ki**	maləŋa
	that	woman	beat	Acc	NAME

Maga:	ustiti	ŋkua	**ki**	toto
	beat	1sFREE OBJ	Nom	NAME

"Toto beat me."

Now it is quite plausible to have an accusative marker in Central-Eastern MP resemble an nominative marker in Western MP, because the older nominative-absolutive (former topic) case in Western MP has been reanalyzed as the accusative in CEMP. But that's not plausible within the Rukai group. In Tanan a broader view of non-demonstrative determiners shows (Li 1973:87):

	Nom	Acc
-pers	ka	ka, sa
+pers	ku	ki

At least it confirms the generalization we drew about u / i / a.

Though in most Western MP languages, and presumably in PMP, the nominative role is marked by zero-consonant **V**, markers of form **kV** also take nominative or quasi-nominative roles in languages spread throughout the AN family. Such uses *are found in Rukai, Mayrinax, Seediq, Amis, Bunun, and* Central Pacific.

The point here is to lead to the proposal that not only for CV demonstratives but also for CV PCMs the PAN consonant distinguishes categories on one semantic dimension and the PAN vowel marks categories on another semantic dimension.

So this section advances three propositions. (1) Every PCM in PAN was phonologically simple CV. (2) The consonant bears the syntactic/semantic role, and the vowel, which is **u**, **a**, or **i**, shows the type of the substantive. (3) The basic unmarked form of the vowel is **u**, and the observed **i** < **ui** < **u i**, and the observed **a** < **ua** < **u a**, where the added particles **a** and **i** were former articles or noun-classifiers.

Further Confirmation: Survival of Intermediate Stages

The preceding section can stand alone. But the following findings may strengthen the conclusion.

Here is evidence in three languages that gives further support to the hypothesis that **i** < **ui** < **u i**, and **a** < **ua** < **u a** in the CV prepositions or casemarkers. Those languages are Chamorro, Kanakanavu, and Paiwan.

In Chamorro, Topping (1973:135) defines **nu** as an article, noting that another scholar defined it as a preposition. He states. "**Ni** is probably a contracted form of **nu i**, as is shown in the following examples:

> Lini'e'   si Pete **ni**   patgon.
> Lini'e'   si Pete **nu i** patgon.
> "Pete was seen by the child."

For Kanakanavu, Li (1997:353) says: "Kanakanavu has the following two (or three) sets of case markers, as based on Tsuchida (1976:36-17) and Mei (1982): Nominative: **sua**, **sa**, si; Oblique: **sua**, **sa**; Locative: na."

While Chamorro supports hypothesis 2 for **nV** and Kanakanavu supports it for **sV**, Paiwan supports it for both. In Ferrell (1982:182) is found:

> **nu**   belonging to, of
> **nu-a**   belonging to, of

and on page 13:

> **nua**   vavaian   a   alak
> CMgen   female   CM=   child
> "the woman's child".
> "the child (who) belongs to the woman"

In the following evidence for **sa** < **sua** < **su a**, bear in mind that Paiwan **t** is PAN ***s**. (That's one reason for my proposal that PAN ***s** was [ts].) On page 12, Ferrell defines **tua** as the casemarker "showing that the relationship is neither equational nor genitive". This makes PAN *su and *sua accusative casemarkers as well as prepositions of space relations and for all oblique adjuncts. And on page 285, "construction marker preceding noun phrase which is neither in Focus nor Agent" with examples:

k/m/an **tu-a** kan-en   "eat food"
sa-linga **tu** ma-ngetjez   "to wish for (someone's) coming"

So we have shown that the choice of vowels in **nV** and **sV** correlated with semantic classes of nouns in earliest PAN. We have mentioned the possibility that some of the PCMs with other consonants may have had lexically-determined vowels, and may have been assimilated only later to the **a/u/i** paradigm.

My suggestion is that the intermediate stages cited here are survivals of expressions from several thousand years ago. An alternate hypothesis might be that in what Sapir called "drift", related languages separately changed in the same ways. A proven example of drift is that in both English and German we say "brown house" (the Germans spell it better), though in the common ancestor of the two languages the phrase rhymed with "spoon loose". But, in the **su-a**, **nu-a**, **nu-i** matter, coincidental survival of ancestral forms seems more plausible.

Evidence that "k plus V" preceded
"zero + V" as Marker of Topic

As we shall see, Mayrinax has some "nominatives" with initial zero consonant and others with initial **k**. We shall see that it also has forms in which **ku** marks common nouns that are specific as opposed to nonspecific in reference.

But the Nataoran Amis language, as described by Chen (1987:135) has both a "Topic" case and a "Nominative". The Topic, as we'd expect from its name, comes at the left edge of the sentence. It begins with any of a set of determiners (ancestral demonstratives) each of which is casemarked with zero + V, and it ends with either a pause or a particular zero-casemarked determiner **iri**, which is specialized for that purpose. It comes down from PAN's left topic marker, **i**, plus demonstrative **di**, "this". In contrast, the Nominative (or

Absolutive) comes right of the verb and is marked by a determiner beginning with **k-**, such as **kia** (the).

So Amis gives the choice of expressing the topic/nominative either at the left of the sentence, in Topic case, or right of the verb, marked in **kV** Nominative case. That shows that the **kV** preceded the simple **V** (plus or minus demonstrative) in changing from Topic marker to Nominative marker function.

Chen's work on Amis was her dissertation under Starosta, and used Lexicase. Here is the list of casemarkers for both -demonst and +demonst:

	Topic	Nom	Gen	Acc	Loc
-demonst	o	ko	no	to	i
+demonst	ia	kia	nia	tia	i-tia

The **o** vowels, of course, are ancestral **u**. Note the unique vowel **i** for the locative where all the other cases show vowel **o** < ***u**. That's an example of the evidence leading me to think that the locative did not arise from the **u / a / i** vowel variants as the other PCMs die.

Here is the list contrasting the Topic and the Nominative for determiners of three distalities:

-Dem	‖		+Demonstrative	
		proximal	neutral	remote
o Topic		i-na	i-a	i-ra
ko Nom		ki-na	ki-a	ki-ra
"the"		"this"	"that"	"that"

Evidence from Mayrinax Atayal
on Marking of Topic, Specificity,
and Nominative or Absolutive

Mayrinax is one of the most conservative AN languages. Huang (1994:109) lists her own table of casemarkers (some of them actually prepositions) and also Mei's table which was published only in Chinese. I mark the lines N (common nouns implying nonspecific Det), S (common nouns implying specific Det), and P (personal pronouns and proper names of persons). Mei's are below Huang's. I mark the

columns for "cases" 0 (Neutral), N (Nominative), A (Accusative), D (Dative), G (Genitive), I (Instrumental), B (Benefactive), C (Comitative), and L (Locative). Genitive includes agent that is not subject (not focus) and, of course, is the one that most anciently was Ablative "from". For Mei's system, C is his Comitative-Allative, and G is his Genitive-Oblique.

It is easily seen that the column don't conform to what Starosta calls Case Forms (morphologically defined and having syntactic role). They come somewhat closer to what he calls Case Relations (defined in terms of semantic role).

	0	N	A	D	G	I	B	C	L
N		a	cu		na	na			i
S		ku	cku	cku	nku	nku	nku		cku
P	i	i	i	i		ni	ni	ki	ki
N		a	cu	cku	na	na		ki	ca
S		ku	cu	cku	nku			(ki?)	(ca?)
P		i	i	ni	ni			ki	i

Huang's charting of Chen's system is not clear on some points.

Notice that specificity (which Huang, after Givon, calls referentiality) occurs both in column N, where it is marked by zero vowel, and in line S, where it is marked by adding a suffix **ku**. The intersection of the two is anomalous. As in other bisyllables in the language, the unstressed vowel in the left syllable is obscured (to something like a schwa) and is not written. From the entries in the N line we may suppose that "cku" is for "cuku" and that "nku" is for "naku".

The specificity shown in column N is consistent with the inherent specificity of topics and the proposal that the nominative markers are old topic markers.

But Mayrinax deviates from the pattern found in some other languages that distinguish specific from nonspecific common nouns by adding vowel **a** for the nonspecifics. What we might expect would be vowel **a** for all entries in row N and vowel **u** retained in row S. If it happens at all (by the most liberal criterion) it could hold only for the Genitive, the Instrumental, and possibly the Benefactive and Nominative.

What seems to happen in Mayrinax is this. Either **kV** or zero + **V** should appear in the N x N cell with vowel **a** (hence **ka** or **a**) and in the N x S cell with vowel **u** (hence **ku** or **u**). In fact, it happens as predicted, though Nonspecific common

nouns keep the zero-vowel form and the Specific ones keep the
k- vowel form, maximizing the distinguishability thereof.

So by examining Mayrinax alone we should conclude
that the use of **ku** as a suffix adding specificity to common
nouns is a peculiarity of the language. It is a striking discovery,
therefore, to find that the same thing (though with suffix **ka**)
was done in Old Javanese. Becker and Hunter (1988) give this
table of "deictic prepositions" in Old Javanese [square brackets
are my comments and hyphens are mine]:

	[Case 1] [Nominative] direct	[Case 2] [Genitive] oblique	[Case 3] [Dative-Loc] directional
indefinite, nonspecific	i	ni	ri [<*di]
definite and	i-ng	ni-ng	ri-ng definite
specific	i-ka-ng	ni-ka-ng	ri-ka-ng

This table is the only one based on attestation alone that
is completely logically consistent (though see Teselkin 1972).
Possibly the ancient AN languages were more analytic and
agglutinative than those observed in the past two centuries.
Those authors' concept of "definite" differs from the usual;
theirs equates to the suffixation of the "ng" linker found also in
Central Philippine.

The Wulai Atayal language, as reported by Rau (1992),
also uses **ku** and **ka** (with regular phonological change) to mark
specificity. The following sentence shows it with the **u** vowel
for the woman and the **a** vowel for the demonstrative: The
vowel variation follows our **a, u, i** paradigm.

```
kyap -un  nya'    qu'   knerin qa-ni
catch UNA 3SGen. Spec. woman this
          agent   syntactic Nom/Absolutive
                  semantic patient
```
"He caught this woman."
"This woman was caught by him."

Evolution of kV and zero-V
Topic Markers from PAN to PMP

One promising hypothesis is that the topic marker **ka**
arose as a conjunction separating a topic from the body of the

sentence. If so, it could be of the same origin as the conjunction **ka** ("and", "but") found in some AN languages; which, in turn, especially when given the vowel **i** for persons, could be the ancestral source of the allative preposition **ki**.

If so, it is not clear whether this use as a general marker of specificity (perhaps definiteness) grew out of or into the use as a topic marker, though it is agreed that topics are characteristically specific.

The Topic-separation particle hypothesis is supported by these findings of Zeitoun's (1996:5). In Mantauran Rukai, where PAN ***ka** appears with regular change to glottal, that particle is the topic marker that concludes the topic and signals that the main clause is about to begin:

> tamatama   ʔa   okanə   vələvələ
> father   Top   eat   banana
> "Father ate a banana"
> ["As for father, he ate a banana"?]

> ðonaʔi   vələvələ   ʔa   okanə   tamatama
> that   banana   Top   eat   father
> "That banana, father ate it."

For Mayrinax, Huang (1994:234) shows sentences with initial topics ending with the Topic marker **ga'**. Some are canonical topics: "As for my dog, it is black and big." "As for him, he will buy me cigarettes." Examples in which a whole clause is the topic are "If you cook, I will wash clothes." "When Yumin comes, I will give him money." In both those, the sentence begins with particle **'i'** (identical with Nom. marker for proper names), and so the two particles frame the topicalized clause. A different particle labeled Irrealis begins topicalized conditions contrary to fact.

As Wolff (1994 and earlier) has presented evidence that PAN had no ***g**, it seems likely that Mayrinax's **ga** topic marker is cognate with Mantauran's and with Bunun's **ka.**

Seediq (Starosta 1984:343) shows the simple V particle as the topic marker **o'** separating the left topic from the rest of the sentence. Perhaps that form arose later than the *k or *g form. At any rate it is a vowel variant of the particle **i** which we have seen as the left marker of Topics that sometimes have k**V** on the right.

Lodən oʔ mikiken kaaLi diha laqi
old man TOP word fight with children
"The old man argued with the children."

In Bunun (Jeng 1994) in a tale of 37 sentences are 23 topics, 8 with both topic-initial and topic-final markers, and 15 with topic-final markers only. The initials are all "maqi". The finals are 14 "a", six "ka", two miscellaneous and one zero. The **a** could be the homophonous PAN demonstrative (Dempwolff's 'i(y)a) or more likely the form with a prosodic glottal whose **ʻi** form and **a** form are Mayrinax nominative casemarkers. Bunun has also a sentence-beginning particle **u**, which is a vowel alternant of the **i** that became the left Topic or Nominative marker in Amis and many MP languages. This is not a statement about what its synchronic status may be in Bunun today.

With almost no exception, AN languages have Verb-Object order and (as Greenberg noted) therefore have prepositions, not postpositions.

The PAN preposition **kV** is specially interesting. Its PAN meaning was almost certainly both allative "to, toward" and comitative "with, accompanying", as Mei (1994) found in Mayrinax Atayal.

The topic marker **kV** may not be related directly, if at all, to the homophonous allative preposition. But the topic marker may be related to the **ka** conjunction "and", which may be related to the comitative **kV**, which could be related to the allative.

The Focus case, in conservative languages of MP and of other divisions of AN, is descended from a PAN topic, and some scholars prefer to call it the Topic case. For Malagasy, members of Keenan's group consider it a Topic in another sense, set off from the rest of the sentence, which is the Comment. But it seems odd to express the Comment before the Topic.

In the most conservative AN languages the topic-focus-subject-nominative-absolutive element is required to be specific ("the" or "a certain" or a proper name, not "some" or "any"). That may be a consequence of the inherent prosodic use of true topics, surviving in languages such as Tagalog in which it is no longer a true topic.

So the question remains, when a phonologically-specified element is specific and is a topic, was it first a sign or specificity or of topicness?

In Mayrinax Atayal (Huang 1994:132-133) the focus (and ancestral topic) marked **ku** is the Absolutive or Nominative case and hence the subject-agent in AF (agent-focus, active voice) and is the patient or location in PF or LF sentences which Huang has sometimes called ergative and at other times called passive voice.

Despite a few obvious irregularities, which tend to obscure the analysis, Mayrinax shows interesting features that either pose problems or help to resolve them.

Essentially, for all cases, the casemarker for specific common nouns consists of the casemarker for nonspecific common nouns plus (amalgamated into a single word) the marker **ku**. That paradigm suggests that the primary and perhaps function of **ku** is "specific" and not "topic". Can it be that historically, it began as a specificity marker and only later became a topic marker? Look again. The Nominative case, which presumably was earlier a topic, is shown as marked by **a** for nonspecific common nouns and **'i** for proper nouns. That fits (and is in part the source of) our hypothesis that casemarking vowels are **i** for proper names of persons and **a** for nonspecific common nouns. But then the marker for specific common nouns in the "Nominative" case should be **u**. (Or perhaps **a-ku** or in view of what happens with other cases in this language.) Evidently the speakers of the language have made some adjustments. One reason it's hard to identify the adjustments is that with addition of the -**ku** suffix for specificity the vowel of the left syllable is obscured and not written. Where "accusative" **cu** becomes (specific) **cku** and genitive **na** becomes **nku** we might guess that the underlying specific forms are **cuku** and **naku**, differing in the underlying vowel that is zero-written in the left mora. But how about when **a** becomes specific **ku**? More puzzling, the nonspecific common noun marker for Locative is **i** (despite the fact that most locatives are non-persons). I tend to explain that one by positing that the Locative preposition was always **i** and was only secondarily, much later, assimilated in part to the **u, a, i** paradigm. Mayrinax is not the only language in which **i** is found where the paradigm predicts **a**.

Here's an interesting discovery in Nataoran Amis, observable in the report by Chen (1987). Look at marking of true Topics in Nataoran Amis. Being specific or definite,

Topics, if common nouns, are marked at their left, with deictic determiners which are the same as casemarked demonstratives in the "Nominative" case, i.e. **i-na**, **i-a**, or **i-ra**, in that order of distality from "this" to "that". (Distances **ni** and **na** are merged in Nataoran, a process that began in phrases in which PCM **ni** and Det **a** coalesced into **na**.) And they are framed, i.e., also marked on the right with either **i-ri** or a pause. Here is an example (Chen 1987:151).

i-a tamdaw **i-ri**, na   [taes-en haw nomiso] ?
[the] man   TM   PAST   hit-ERG   QM
PN2s(GEN)
"As for the man, did you hit him?"

Chen applied the term TM, topic marker, only to the **i-ri** on the right, and not to the determiner on the left; presumably because determiner is already a category and has other uses, while **i-ri** is unique. "The topic marker belongs to a unit set which has **i-ri** as its only member." True, but what is it diachronically? The answer is that **i-ri** is the specialized and fossilized use of another demonstrative, to wit, PAN *id_1i, Tagalog **i-re**, Malagasy **i-ry**, the most proximal distance of all in Tagalog, "this, very close to me here". As I have shown (1999) a cognate is PPN *re, Samoan **le**, Tongan **e**, specifical article "the" or "a certain". At some time in the past, in Amis, that one was specialized as the right framer of topics and constrained against general use as a deictic.

In a MP language, Malagasy, a phrase often has a demonstrative put both left and right of it. A common one is **ity** phrase **ity**. The word **i-ty** is a typical ancient casemarked demonstrative in which i is the ancient casemarker for the Topic-Nominative case, and **ty** [ti] shows low distality ("this", not "that"). The -y is merely Malagasy's way of writing final -i, perhaps copied from English.

A sentence in Malagasy, in the analysis by Ed Keenan's student, Pearson (1996), consists of two elements. The left element, which typically is longer and more complex, is the predicate or Comment. The right element is the subject or Topic. Here is a simple sentence (ibid.:121) with double bars separating the two elements:

Mihinana   ahitra ǁ ny   omby
eat (ST)   grass   Det ["the"] cow
"Cows eat grass."

The "ST" (subject topic) is what others call "AF" (agent focus). The subject-topic is at the right and bears a mark of being specific-definite-referential, in this case an ancient proximal demonstrative, **ni** ("this"), which has been weakened by further grammaticalization into an article, and marks a specific type of creature. Another way to say it in English is "The cow eats grass."

Now, in contrast, here's a sentence (ibid. 118) whose subject-topic has a Det, **ity**, that still has specific demonstrative denotation:

Mamaky	ny	boky	‖	**ity**	lehilahy	**ity**
read(ST)	Det	book		this	man	this

In a footnote, Pearson explains: Malagasy demonstrative DPs (e.g. ity lehilahy ity) usually contain two copies of the demonstrative element, at the left and right of the subject/topic. These echoing forms are called "framing demonstratives".

Other sentences in the volume show non-framing of subject-topics, but none shows framing in any DetP in the predicate.

In the hypothesis we are examining, framing in Formosan Amis and MP Malagasy was an intermediate step between final-only marking as with **ga** in Atayal, or **'a < ka** in Mantauaran Rukai, and initial-only marking as with "i(+ng)" "a(+ng)" as in Philippine languages.

Implications for Major Grouping of Austronesian

I suggest that a reasonable explanation for all this is that at some time in the past, the common ancestor of Amis and Malayo-Polynesian, possibly PAN, developed a marking for non-clausal phrases that occurred sentence-initially and were not part of any clause. The marking consisted of simple **V** (**i** or **u**), with or without a demonstrative right half. The alternative marker **kV** was perhaps more a marker of specificity as such. Either one went through a stage in which it framed the non-clausal element on its right and on its left. Framing could consist of simple **V** on the left and **kV** on the right. The casemarking element could be followed by a demonstrative or other determiner (**V, dV, nV, tV**), making a bimoraic framer. It became possible to move the whole topic to the right side of the sentence, in which position it could become part of the clause,

and become more Nominative and less of a Topic. Some languages, including Malagasy, have kept the option of leaving it framed. Languages such as Tagalog kept the former topic as *Case 1, the "a-ng" case. In such languages the sentence-final* position, though still allowed for the Focus , has ceased to be its usual position.

Furthermore, a conclusion is that PMP and Amis share a basic syntax that we don't see elsewhere: framing of the (former) Topic with demonstrative determiners right and left.

Using classical phonological methods, Blust (1999) has proposed ten first-order subgroupings within Austronesian: nine within Formosa plus MP.

Our discovery of an ancient specificity marker **ka / ku** in Mayrinax and Old Javanese need not challenge Blust's system. It doesn't seem implausible to me that PAN had the specificity system and that it was lost in most branches of AN. A group ancestral to Old Javanese was able to add the linker **-ng** < **nV** between the Det and the noun, the same way it is added after other Det in Old Javanese and some Philippine languages.

On the other hand, our discovery of the left and right framing of ancient topics with demonstratives seems more likely to complicate the system of primary branches of AN. Both Mantauran Rukai and Mayrinax Atayal show certain archaic features that would be hard to reconcile with a proposal that at some time in their ancestry they had the topics framed by demonstratives right and left. It appears that PMP and Amis must share an ancestry not shared with Rukai nor with Atayal.

References

Becker, Alton L. and Thomas Hunter. 1988. Cohesion in Old Javanese prose. In Richard McGinn (Ed.), Studies in Austronesian Linguistics. Ohio University Press, with distributor Chicago: University of Chicago Press. Pp. 475-488.

Blust, Robert A. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Selected Papers from the Eighth International Conference on Austronesian Linguistics. Elizabeth Zeitoun and Paul Jen-kuei Li, Editors. Number 1 in the Symposium Series of the Institute of Linguistics, Academia Sinica. Taipei: Academia Sinica. Pp. 1-94.

Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. The Evolution of Grammar. Chicago: University of Chicago Press.

Chen, Teresa M. 1987. Verbal constructions and verbal classification in Nataoran-Amis. Pacific Linguistics C-91.

Ferrell, Raleigh. 1981. Paiwan Dictionary. Pacific Linguistics C-73.

Finney, Joseph C. 1997. Casemarking and the *ia demonstratives. In: Proceedings of the Seventh International Conference on Austronesian Linguistics (Leiden University 1994). Leiden University. Cecilia Ode and Wim Stokhof, eds. Leiden, Netherlands. Pp. 551-574.

Finney, Joseph C. 1999 (in press). Two studies on Polynesian. In: Proceedings of the 4th Conference On Oceanic Linguistics. Wolfgang Sperling and Steven Fischer, Eds. Auckland: Leo Pacifica.

Foley, William A. and Robert D. Van Valin, Jr. 1984. Functional Syntax and Universal Grammar. Cambridge University Press.

Huang, Lillian M. 1993. A Study of [Wulai] Atayal syntax. Taipei: Crane.

Huang, Lillian M., 1994. A Study of Mayrinax syntax. Taipei: Crane.

Jeng, Heng-hsiung. 1977. Topic and Focus in Bunun. Special Publication 72. Taipei: Institute of History and Philology, Academia Sinica.

Jeng, Heng-hsiung. 1994. Two discourse models for analyzing a Bunun story. Read at the Seventh International Conference on Austronesian Languages, University of Leiden, Netherlands.

Li, Paul Jen-kuei. 1973. Rukai Structure. Special Publication #64, Institute of History and Philology.

Li, Paul Jen-kuei. 1995. The case-marking system in Mayrinax Atayal. Bulletin of the Institute of History and Philology, Academia Sinica, Vol. 66, Part 1, 23-52.

Mei, Kuang. 1994. Word order, case markers and theta agreement in Mayrinax (in Chinese). (not seen; quoted in Huang, 1995:287-289).

Mosel, Ulrike, and Even Hovdhaugen. 1992. Samoan Reference Grammar. Oslo: Scandinavian University Press.

Pearson, Matthew. 1996. Domain phrases and topic arguments in Malagasy existentials. In Matthew Pearson and Ileana

Paul (eds.), The Structure of Malagasy. UCLA Occasional Papers in Linguistics, 17. Los Angeles: University of California. Pp. 113-141.

Rau, Der-Hwa Victoria. 1992. A Grammar of Atayal. Taipei: Crane.

Reid, Lawrence A. 1974. The Central Cordilleran subgroup of Philippine languages. Oceanic Linguistics 13: 511-560.

Ross, Malcolm D. 1992. The sound of Proto-Austronesian. Oceanic Linguistics 31:23-64.

Starosta, Stanley. 1974. Causative verbs in Formosan languages. Oceanic Linguistics 13:279-369.

Szakos, Joseph. 1994. Die Sprache der Cou (The Tsou Language, Ph.D. dissertation). Bonn: Rheinischen Friedrich-Wilhelms University.

Teselkin, A. S. 1972. Old Javanese (Kawi). Translated and edited by John M. Echols. Translation Series, Modern Indonesia Project, Southeast Asia Program. Ithaca, N.Y.: Cornell University.

Topping, Donald M. 1973. Chamorro Reference Grammar. Honolulu: University of Hawai'i Press.

Wolff, John. 1995. The position of the Austronesian languages of Taiwan within the Austronesian group. In Paul J.-k. Li et al. (eds.), Austronesian Studies Relating to Taiwan, Number 3, Symposium Series of the Institute of History and Philology. Taipei: Academia Sinica. Pp. 521-583.