Pronouns, Clitics, Orders and Grammaticalization in Tukang Besi

Mark Donohue

Department of Linguistics, Research School of Pacific and Asian Studies
Australian National University

There are two bound pronominal sets on Tukang Besi verbs. One set, that which marks Ps, is more transparently related to the independent pronominal forms and is a clitic, not an affix. Despite this, I argue that it is in fact the more grammaticalized of the two bound pronominal sets. Just as the notion of ‘markedness’ can be shown to be decomposed into mutually contradictory defining conditions, I argue that the widely used notion of ‘grammaticalization’ is similarly conflicted.

1. Pronominal sets in Tukang Besi

Tukang Besi is a Malayo-Polynesian language from Southeast Sulawesi in central Indonesia, on the very edge of the ‘Central Malayo-Polynesian’ zone (Donohue 1999, 2004a). It is a conservative Austronesian language, being verb-initial and employing case-marking on all NPs.

There are five (and a half) bound pronominal sets in Tukang Besi; in this paper I shall concentrate on the first three, which are used to cross-reference arguments on main clause verbs, and shall also compare these with the independent set, shown in the last column of Table 1.

Where’s Tukang Besi from?

There are five (and a half) bound pronominal sets in Tukang Besi; in this paper I shall concentrate on the first three, which are used to cross-reference arguments on main clause verbs, and shall also compare these with the independent set, shown in the last column of Table 1.
Table 1. Pronominal forms in Tukang Besi

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Role:</td>
<td>S,A</td>
<td>P</td>
<td>Dative P</td>
<td>Intra-directive</td>
<td>POS'SR*</td>
</tr>
<tr>
<td>(mood):</td>
<td>R</td>
<td>I</td>
<td></td>
<td></td>
<td>(any)</td>
</tr>
<tr>
<td>1sg</td>
<td>ku-</td>
<td>ku-</td>
<td>=aku</td>
<td>=naku</td>
<td>=su</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2sg</td>
<td>'u-/nu-</td>
<td>ko-</td>
<td>= ko</td>
<td>= nso</td>
<td>= 'u</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3sg</td>
<td>no-/o-</td>
<td>na-/a-</td>
<td>= 'e</td>
<td>= nes</td>
<td>= no</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1pa</td>
<td>ko-</td>
<td>ka-</td>
<td>= kami</td>
<td>= nsami</td>
<td>= nto</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1pl</td>
<td>to-</td>
<td>ta-</td>
<td>= kita</td>
<td>= nggita</td>
<td>= mami</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2pl</td>
<td>i-</td>
<td>ki-</td>
<td>= komiu</td>
<td>= ngkomiu</td>
<td>= miu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3pl</td>
<td>no- / o-</td>
<td>na-/a-</td>
<td>= 'e</td>
<td>= ne</td>
<td>= amai</td>
</tr>
</tbody>
</table>

* The forms in this column are nominal possessor / subordinate terms.

We can note that the P enclitics are transparently related to the independent forms (in most cases simply by removing the i-). The S,A prefixes are less transparently related, and show more variation in form and more evidence of paradigmatic intervention from other grammatical categories (like modality), in that there are two sets, realis and irrealis.

The order of bound elements on the verb is fixed; for all main verbs S,A prefixes are obligatory, while P enclitics are optional. In (1) and (2) the prefix ku- and the clitic = 'e are shown in the only possible order in which they may appear.

(1) Ku- 'ita = 'e na 'obu.
   1sg-see =3p NOM dog
   ‘I saw the dog.’

(2) Ku- 'ita te 'obu.
   1sg-see CORE dog
   ‘I saw the dog.’

The position of the pronominal clitics is not affected by the presence of negation or auxiliaries in the clause. Interestingly, aspectual clitics are attracted to auxiliaries (though this does not apply to negation, where the placement of aspect on the negator or on the verb implies different scope).

(3) a. Mbeaka ku- 'ita te 'obu.
   neg 1sg-see CORE dog
   ‘I didn’t see the dog.’

b. Ku-mbeaka 'ita te 'obu.
When a non-bound pronominal occurs, there is no strict ordering of this with respect to any nominals (or other pronominals) in the clause. This is in sharp contrast to many other Austronesian languages (see, for examples, Billings and Kaufman 2004, Lee and Billings 2005).

The preceding examples have shown, briefly, that bound pronominals are strictly positioned on the verb, with S,A agreement forms always appearing prefixal to the verb root, and P agreement forms always following as enclitics. These agreement markers are strictly hosted by the verb, even in the presence of other possible hosts (such as auxiliary verbs, which do attract aspect marking) in the clause. Furthermore, the independent pronominals do not display any clitic-like behavior, behaving rather just as do DPs in the language generally. In short, we have seen no exceptional behavior from the pronominals at all. The fact that the S,A cross-referencers have been described as affixes, and the P cross-referencers as clitics requires some explanation.

2. PS: clitic, or prefix?

The reasons for this terminological distinction are somewhat complicated for Tukang Besi. For instance, both S,A agreement markers and P agreement markers are part of the stress domain of the word, as can be seen in (21). This is shown in (6).

Despite their both appearing in the same stress domain, there are differences between the two agreement markers. The S,A prefixes are outermost in the verb, while the P enclitics may be followed by aspectual clitics, as in (7).

This might appear to be a minor difference, but is relevant for the discussion of noun incorporation. Unlike the case for the various prefixes and suffixes of Tukang Besi verbal morphology, an incorporated nominal is not part of the stress domain of the verb.
Basic clause

(8)  \( \text{No-manga=} 'e \ na \ kaluku \ te \ amai. } \\
3\text{R-eat}=3_p \ \ \text{NOM coconut CORE 3pl.} \\
\text{‘They ate the coconut.’}

(8’) [\text{noma’yəʔenaka’lukuˈte}a’ma’i]

NI clause

(9)  \( \text{No-manga kaluku} \ na \ amai. } \\
3\text{R-eat coconut nom 3pl.} \\
\text{‘They eat coconuts.’}

(9’) [\text{no’maŋaka’luku}naʔa’ma’i], *[\text{nomaŋaka’luku}naʔa’ma’i]

The structure involved in these NI constructions is most likely more similar to those described by Asudeh and Ball (2005) than the more familiar NI structures that involve the complete integration of the N into the V. The relevance of NI constructions for the analysis of the pronominal cross-referencing is that in this construction we find that the enclitics can be separated from the verb stem by an incorporated nominal. This is most simple when the clitic is an aspectual clitic, as in (10) and (11). (With a non-NI clause this is not possible; the aspectual clitic must be hosted by the V: \text{Nomangamo te kaluku, *Nomanga te kalukumo} ‘They have eaten coconuts.’) Note that the incorporated nominal can appear following the V-final aspect clitics (cf. Massam 2001).

(10)  \( \text{No-manga=} mo \ kaluku \ na \ amai. } \\
3\text{R-eat}=\text{PF coconut nom 3pl.} \\
\text{‘They now eat coconut.’}

(10’) [\text{noma’yəmoka’luku}naʔa’ma’i]

(11)  \( \text{No-manga kaluku=} mo \ na \ amai. } \\
3\text{R-eat coconut}=\text{PF nom 3pl.} \\
\text{‘They now eat coconut.’}

(11’) [\text{no’magaka’luk’umona}ʔa’ma’i]

When a transitive verb has been causativized, it is the causee, not the causand, which may be cross-referenced on the verb. Despite this, it is the causand, and not the causee, which may display NI (it is still a distinct domain for the purposes of stress assignment).
Causativized version of (8)

(12) * No-pa-manga = 'e te kaluku (na ana) te ina.
3R-CAUS-eat = 3p  CORE coconut NOM child CORE mother

'Mother made the kids eat the coconuts.'

(12') * No-pa-manga = 'e te ana na kaluku te ina.

NI with a causativized clause

(13) No-pa-manga kaluku te ana na ina.
3R-CAUS-eat coconut CORE child NOM mother

'Mother would make them eat coconuts.'

(13') * No-pa-manga ana te kaluku na ina.

It is possible for a single complex predicate to display both NI and cross-referencing for the causee. The cross-referencing must appear following the incorporated nominal, and not preceding it. This shows that the P agreement can be attached to something that is phonologically not part of the same word as the V, which is less like an affix, and more like a clitic, in terms of behavior.

NI and P agreement on a causativized clause

(14) No-pa-manga kaluku = 'e (na ana) te ina.
3R-CAUS-eat coconut = 3p  NOM child CORE mother

'Mother would make the children eat coconuts.'

(15) * No-pa-manga = 'e kaluku (na ana) te ina.

In section 3 I present a number of facts that complicate the discussion of the bound pronouns. Before that, however, I shall reiterate a number of facts that are usually assumed about the process of grammaticalization, particularly as it concerns the development of agreement systems.

3. Grammaticization, form and function

Models of grammaticalization assume that the existence of agreement in a language is the result of the cliticization, and later affixation, of once-free pronominal elements. In the process of being grammaticalized these pronominal elements have lost some of their phonological material and independent status—and, in some cases, their pronominal status as well—(Givón 1976, Bresnan and Mchombo 1987, among others). The examples in (16) show, using English morphemes, the supposed route for the development of agreement morphemes in a language. For both examples stage i represents the language with no agreement; stage ii shows the optional appearance of a left- or right-dislocated NP, and a pronoun inside the clause. In stage iii the pronoun is bound to the verb, no longer an independent word with it’s own stress. In the final stage, stage iv, the pronominal is
phonologically reduced and no longer analysable as a pronoun, but simply as an agreement marker on the verb, since the NP is now reintegrated into the basic clause structure.

**Acquiring subject agreement**

(16) a. i. That man ate the apples  
    ii. That man, he ate the apples  
    iii. That man, he-ate the apples  
    iv. That man i-ate the apples

**Acquiring object agreement**

b. i. That man ate the apples  
    ii. That man ate them, the apples  
    iii. That man ate-thǝm, the apples  
    iv. That man ate-ǝm the apples

In general, in discussions of grammaticalization, it is assumed that the more divergent the bound form is from the (co-existing, or putative historical) independent form, the greater the time depth of its binding, and so the less independence the bound form has (Harris and Campbell 1995). Similarly, the more involved in the inflectional semantics of the host a morpheme is, the longer it has been associated with that host. We have seen examples of the free pronoun *them* being realized as *thǝm*, and in its most fully bound form *ǝm*, where it realizes the Old English *hem*, long since lost as an independent pronoun but preserved as a clitic.

(17) Free form > resumptive pronoun > (optional) clitic > bound affix  
    phonologically maximal → phonologically minimal

By this model the prefixal forms in Tukang Besi would count as more highly grammaticalized than the enclitic forms. Some of the data that could be used to support this claim are shown in (18). We can see that the enclitic forms that are used to show agreement for P are more similar to the free forms than are the prefixal forms that mark S or A agreement. Additionally, we note that the prefixes are, for all but the 1sg, fused with the realization of the realis/irrealis distinction in the language, suggesting a long involvement with the inflectional categories that are relevant to verbs.

(18) Free form > (optional) clitic > bound affix  
    iaku =aku ku- 1sg  
    ikita =kita to-, ta- 1pl  
    maximally isolating → portmanteau tendencies  
    phonologically maximal → phonologically minimal  
    not grammaticized → maximally grammaticized

If the verbal agreement morphemes of Tukang Besi were analysed as showing the greatest degree of grammaticalization

The following section presents a number of facts that indicate that we must consider the prefixes to be less tightly bound to the verb than are the enclitics.
4. Arguments against the iconicity of grammaticalization

In this section I shall present data and argumentation that indicate that the S,A prefixes are less tightly bound to the verb than the P enclitics. I will suggest that the apparent freedom of pronominals in (5) is epiphenomenal, relying on the misinterpretation of words such as amai (and iaku, ikami, ikita etc.) as pronouns. A more accurate analysis treats them simply as nominals, with pronominal features.

The evidence for the status of the prefixes as less grammaticized than the clitics is presented in the following sections.

4.1 Prosodic

The prefixes may be separated from the verb they appear with by a hesitation pause; this is not possible for the enclitic, or for other prefixes. Textual examples are given in (19) through (21). In these cases we can see that interruptions can occur between the prefix and the verb it appears with.

(19) Toka eaka ko-, eaka ko-motur(u) i la’a=mo.
    but not  1PA,R- not  1PA,R-sleep OBL only = PF

    ‘But w[e]- di[dn’t], we didn’t sleep at all.’

(20) Sa-anu=no, o-, ’ido-api=’e=mo, te watu, jari sa-to-to’oge.
    when-whatsit = 3GEN 3R- live-APPL = 3P = PF  CORE  stone and.so 1-RED-big

    ‘And then, it, grew on it, the stone, an’ was all big.’

(21) Jari o-, o-po-’awa=mo. O-pogau na Ndokendoke kua ...
    so  3R-  3R-REC-get = PF  3R-say NOM  Monkey COMP

    ‘And so th[ey], they met. Monkey said: “...” ’

In (22) we see that elicited examples involving a break between a verb and its P agreement enclitic are not felicitous. In (14) the examples show that the ability for a prefix to be separated from the verb is a function of the agreement prefixes in particular, and not other verbal prefixes, such as the passive or the causative.

(22) a. * ku-’ita, = ’e na ’obu b. * no-’ita, = aku
    1SG-see = 3P NOM  dog  3R-see = 1SG.P

    ‘(I saw, the dog)’  ‘(They saw, me)’

(23) a. * no-to- , tu’o b. * no-pa- , kede=’e
    3R-PASS- -chop  3R-CAUS- -sit = 3P

    ‘It was, chopped’  ‘They made, her/him sit’
In sum, the data from prosody suggests that the S,A prefixes are less strictly phonologically bound to the verb than are the enclitics.

4.2 Morphological

Tukang Besi reflects the Austronesian infix -um-. Abstracting away from its several functions (see Donohue 1999), this infix appears before the first vowel of a verb, just as is the case in more conservative languages such as Tagalog. This can be seen in (24), in which the use of the infix is obligatory in this particular subordinate clause construction in both languages.

\[
\begin{array}{lll}
\text{Tagalog} & \text{Tukang Besi} \\
\text{(24) a.} & \text{ang} & \text{lalaki-ng} & t<\text{um}>akbo \\
\text{Nom} & \text{man-LNKR} & \text{run < UM>} & b. \text{te} & \text{mo’ane} & t<\text{um}>inti \\
\end{array}
\]

‘the running man’

In main clauses in Tukang Besi we must show agreement on the verb with S,A prefixes, as described in the previous sections. These prefixes, while part of the prosodic word for the purposes of stress assignment, do not ‘count’ as part of the relevant domain for the purposes of aligning the -um- infix. In (25a) we have a simple inflected verb; (25b) shows the same verb in irrealis mood, an environment in which the use of -um- is quite natural. The infix appears immediately preceding the first vowel of the verb root, not the first vowel of the inflected verb. It is not possible for -um- to appear preceding the first vowel of the prefix.

\[
\begin{array}{lll}
\text{(25) a.} & \text{No-kede} & b. \text{Na-}k<\text{um}>ede. \\
3r-sit=\text{PF} & 3i-sit < UM > & c. * n<\text{um}>a-kede \\
\end{array}
\]

‘They sat.’ ‘They will sit.’

We might argue that the facts in (25) simply reflect an instance of the mirror principle: the prefix applies following all other affixation, and has a higher scope than the infix, so must be realized outside the domain of the alignment of the infix. This analysis will not work, however; examining a greater body of data we see that other prefixes do fall within the domain of -um- infixation. In (26b) the root kede has been derived with the prefix pa-‘causative’; the resulting stem acts as a single domain for the purposes of infixation. As (26c) and (26d) show, it is not possible for the infix to appear aligned to the root alone, or to the root + pronominal prefix.

\[
\begin{array}{lll}
\text{(26) a.} & \text{no-pa-kede} = 'e' & b. \text{na-}p<\text{um}>a-kede = 'e' \\
3r-caus-sit = 3p & 3i-caus < UM > -sit = 3p & c. * n<\text{um}>apakede'e \\
\end{array}
\]

‘They seat them’ ‘They will seat them’

In sum, the S,A prefixes do not count as part of the domain that is relevant for infixation.
4.3 Acquisitional

Children acquire the P agreement clitics before they acquire the S,A prefixes. In (27) and (28) I show approximations of the order of acquisition of verbal inflectional morphology by children. The adult forms are shown in (27a) and (28a), giving verb forms with maximal agreement. As indicated in the b examples, the earliest child language does not inflect verbs, using simply bare verb roots. If an object is required in the sentence, it can be indicated by the use of a (non-case marked) pronominal element. Between this bare-verb stage and the fully-inflecting adult stage there is a period, at about 5-7 years old, when children use both prefixes and enclitics on their verbs; the enclitics correspond to the enclitics used by adults. The prefixes, on the other hand, are invariant, and do not approximate the S,A prefixes employed by adults. Instead, the prefixal position is filled with mo, a semi-frozen adjective-‘deriving’ prefix found on other lexical items, but not an inflectional morpheme (this is also the productive anticausative prefix; see Donohue 2004b).

<table>
<thead>
<tr>
<th>Adult</th>
<th>Early child</th>
<th>Later child</th>
</tr>
</thead>
<tbody>
<tr>
<td>(27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. no-kede</td>
<td>b. kede</td>
<td>c. mo-kede</td>
</tr>
<tr>
<td>3r-sit</td>
<td>sit</td>
<td>mo-sit</td>
</tr>
<tr>
<td>‘They sit’</td>
<td>‘(They) sit’</td>
<td>‘(They) sit’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(28)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. no-lita = aku</td>
<td>b. lita yaku</td>
<td>c. mo-lita = 'aku</td>
</tr>
<tr>
<td>3r-see = 1sg.p</td>
<td>see 1sg</td>
<td>mo-see = 1sg.p</td>
</tr>
<tr>
<td>‘They see me’</td>
<td>‘(They) see (me)’</td>
<td>‘(They) see me’</td>
</tr>
</tbody>
</table>

It is clear that if acquisition of linguistic forms corresponds to the ‘basicness’ of fusion of those elements to the stems to which they attach, then the P clitics must be counted as being more basic than the S,A prefixes.

4.4 Semantic

The prefixes mark the realis/irrealis distinction (cf. van den Berg 1996), a grammatical feature normally associated with high scope. The enclitics, on the other hand, are purely local in scope. If this correlates with structural position, it does not support (8) and (9).

5. Therefore …ordering pronouns?

The question of the lack of any ordering restrictions for the ‘independent pronouns’, as seen in (5), might well be an epiphenomenal one: there are good reasons to believe that there are no independent pronouns, and that the forms in the rightmost column of Table 1 are simply Ns. For instance, examine (29) and (30) (Donohue 1999:305–306). Here we have a ‘pronoun’ which is modified by an adjective and by a demonstrative.
Mark Donohue

(29) \[ DP \text{ Te iaku mo'aro ana} \quad \text{ku-melu te i-manga.} \]
\quad \text{CORE 1SG hungry this 1SG-plead CORE PP-eat} \\

'I who am hungry beg (you) for some food.'

(30) \[ Te \text{ ikomiu l<um>} \text{aha te ika atu,} \]
\quad \text{CORE 2PL search <UM> CORE fish that} \\

bar(a) \quad \text{i-gugudu} \quad \text{awan(a) atu la!} \\
\text{don’t 2PL.r-make.noise manner that MASC} \\

'You boys who are looking for fish, don’t make such a noise!

A more empirical analysis would treat the ‘independent pronominals’ as discourse-variable nouns. The behavior of these nominals is identical to that of ‘normal’ nominals (other than being able to be possessed, something that can easily be accounted for on semantic grounds). Furthermore, there are no morphosyntactic privileges that accrue to ‘pronouns’, and not to nouns; note that independent ‘pronouns’ require the same overt case marking in all positions, as well as allow any plausible modification, including by demonstratives (even combinations like \text{te iaku atu} \quad \text{CORE 1SG that ‘that (other) me’ are grammatical, showing that the ‘pronouns’ lack even inherent deictic specification). The reason for the lack of any special ordering characteristics for the ‘independent pronouns’ is simply that there are no independent pronouns, and so they don’t need to be ordered about any more than nouns do.

We have seen that, in contrast, the bound pronominals are locked into position; but the prefixes appear to be less bound than the clitics:

(31) \[ IP \text{ s,a- [VP V(=p) ]} \]

This despite the fact that all the standard criteria for grammaticalization, namely portmanteau behavior and phonological reduction, would lead us to conclude that the S,A prefixes were more grammaticalized, presumably meaning that they had spent longer in a non-independent form.

The form of a morpheme, and its degree of apparent grammaticalization, cannot be assumed to necessarily correlate with the degree to which that morpheme is bound to a stem. The bound forms refuse to be ordered about.
Abbreviations

Abbreviations conform to the Leipzig Glossing Rules (www.eva.mpg.de/lingua/pdf/LGR08_09_12.pdf) with the exception of the following:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, S, and P</td>
<td>following Comrie (1978)</td>
</tr>
<tr>
<td>PA</td>
<td>paucal</td>
</tr>
<tr>
<td>CORE</td>
<td>an argument that is not</td>
</tr>
<tr>
<td>PA</td>
<td>nominative, but still a term</td>
</tr>
<tr>
<td>PF</td>
<td>perfective</td>
</tr>
<tr>
<td>P</td>
<td>P-prefix</td>
</tr>
<tr>
<td>R</td>
<td>realis</td>
</tr>
<tr>
<td>I</td>
<td>irrealis</td>
</tr>
<tr>
<td>REC</td>
<td>reciprocal</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative, the most</td>
</tr>
<tr>
<td>RED</td>
<td>reduplication</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SS</td>
<td>Studies in Philippine Languages and Cultures, Volume 17 (2008)</td>
</tr>
</tbody>
</table>

References


