Variation in Voice in Indonesian/Malay: Historical and Synchronic Perspectives

MARK DONOHUE

1 Variation in a ‘Language’: The Speed of Diversification

Much has been written about linguistic divergence, in the sense of the gradual historical fissioning of one language into what must eventually be considered distinct linguistic codes (see, for instance, Antilla 1989, Crowley 1987, Lass 1997). More recently a respectable amount of work has been produced about the ‘opposite’ end of the scale of language development, about linguistic convergence, in which languages, often of diverse typological profiles and/or diverse genetic backgrounds, show degrees of mutual assimilation in a defined social or geographic zone (see, for instance, the papers in Bakker and Mous 1994, and also Heath 1978, 1981, Mithun (this volume), Thomason and Kaufman 1988, Thurston 1987, Ross 1999, Weinreich 1963, amongst others).

In Indonesian/Malay we have an instance of a historically-attested single language which has spread, through trade and later bilingualism, over a wide area of insular Southeast Asia in the course of what is for language spread a very short time. The speed and geographic scope of this diffusion represents a spread of a magnitude that is rivaled only in the period of European colonisation. The legacy of that spread is a remarkably rich range of varieties of Indonesian/Malay across the archipelago, varieties which hover on the border of mutual intelligibility. While the overall lexical simi-
larity between different Malay varieties is high, differences centre of the high-frequency functors (aspect markers, negation, and especially pronouns, this last most markedly in the vicinity of New Guinea). This combines with various grammatical differences to lead to the almost or functional unintelligibility that is found when speakers of one variety of Indonesian/Malay encounter speakers from other areas and who do not command the standard languages.

This paper takes voice systems as a diagnostic of variation, and examines the variation in the scope of six different Indonesian/Malay voice systems seen as the (re-)rankings of various universal constraints. Following the description and analysis of these different patterns, I shall explain as far as possible the changes in terms of deviation from a more conservative model, a historically prior system which is still attested in the Philippines, guided by functional, language variety-internal considerations and possible local language interference or influence. The description of variance will be phrased in terms of correlations with the high-ranking constraints in each system, and their relative rise and fall. This offers a model that accounts for the development of the different voice systems while illustrating both the conservatism that Austronesian languages are known for, as well as a tendency to develop from a typologically unusual, and evidently ‘unstable’ system to a variety of more widely attested, and so presumably typologically stable, voice systems.¹

2 Indonesian/Malay

Indonesian/Malay is an Austronesian language originally from what is now Malaysia. The earliest records come from near modern Palembang, in the 7th century AD, from what was then the kingdom of Srivijaya. It spread widely from this base by trade from the 14th century onwards, giving rise to numerous local varieties. Today it is spoken natively by populations in eight countries, and is a national language of four: Malaysia, Singapore, Brunei and Indonesia. Despite sharing the same name and mainly the same history, there is significant variation between the different varieties. Accelerated by the arrival of Arab, and later Portuguese and Dutch, traders in the region, Malay was transplanted to areas far beyond its previous range of use, and has shown local developments not shared in the ‘homeland’, as well as extensive creolisation in some areas, while showing the Austrone-

¹ ‘Unstable’ appears in quotes because, although all evidence of grammaticalization indicates that Austronesian languages are moving away from this model, it is nonetheless true that the Patient-prominent pattern has been stable in many conservative Austronesian languages for over 5,000 years.

3 Overview

The maximal system is attested in Standard Malay/Standard Indonesian. I shall illustrate this system with Standard Indonesian (SI), which has three basic voice choices, as well as a semantically derivational option.²

The three basic voice choices are shown in (1)-(3). Sentence (1) shows the active voice, marked by the presence of a preverbal Agent, and the prefix *maq-. (orthographically meng-, but showing considerable morphophonemic alternation, as described for Austronesian languages by, eg., Newman 1984 and many since) on the verb, in this case tonton ‘watch’.² The passive counterpart is shown in (2); here the preverbal argument is the Patient, and the verb is prefixed with *di-, the non-active prefix. Furthermore, the Agent is marked with the oblique preposition oleh ‘by’. The third basic voice choice is shown in (3), and it shows a blend of the syntactic properties of the active and the passive clauses. The preverbal subject DP is the Patient, and the verb is prefixed with *di-, the same prefix that appears in the passive clause. In (3), however, the Agent is not marked prepositionally,

² Counts show that even the most divergent Malay lects in the east are still approximately 85% cognate with ‘standard’ Malay or Indonesian. We should note, however, that the ±15% non-cognate words tend to be the most high frequency words: pronouns, function words and negation markers. This leads to low intelligibility between varieties.

³ Chung (1978) demonstrates that there is also a morphologically unmarked active voice choice (what she calls ‘stem sentences’). Clauses with this coding option are found predominantly in the informal register, and behave identically to the maq-marked active voice described here. Carter (1984) also provides useful discussion on clause types with unmarked verbs. See Kaswanti Parwo (1988) for some alternative views on some of Chung’s analysis, and of a view of voice in Indonesian that is very different to that presented here, but for which the discourse motivations are not incompatible with mine. Other works on voice in Indonesian include Kraft (1995).

⁴ Some verbs regularly appear without marking for the active voice. For example makan ‘eat’, minum ‘drink’ do not normally take *maq- unless derived: *makan, *minum. This reflects the fact that they include frozen versions of cognitive morphemes: makan < *maq- + *kan, and minum < *(o)m- + *inum. See also note 2, and the discussion of example (13).
and indeed does not show the behaviour of an oblique in any fashion, appearing to remain a core argument of the verb, though not the subject.\footnote{The following abbreviations appear in the examples and discussion: 1, first person; 2, second person; 3, third person; ACC, accidental prefix; ACT, active voice; AGR, agreement; AGL, 'agent'-like argument, = Comrie's A; AGT, non-subject 'agent'; AM, Ambon Malay; APPL, applicative; AV, active voice; C, case; CONJ, conjunction; DAT, dative (case); DV, dative voice; GEN, genitive; INV, inverse; IM, Jakarta Malay; KM, Kupang Malay; L, local (first or second person); LOC, locative; MM, Makasar Malay; NOM, nominative; NONACT, non-active voice; OBJ, object; OBL, oblique; PASS, passive; PAT, 'patient'-like argument, = Comrie's P; PERF, perfective; PHI, Philippine; PL, plural; PM, Papuan Malay; POSS, possessive; PV, 'patient' voice; RED, reduplication; REL, relative; SEm, Semli Malay; SG, singular; SI, Standard Indonesian; S1, sibling; SM, Standard Malay; STRUC, structure; SUI, subject; TR, transitive; functions as an applicative or causative depending on the verb's semantics; VM, voice marking; X, high prominence; x, low prominence; y, younger.}

*Active*

(1) \textit{Dia me-[n]jonton gadis cantik itu.}  
3SG ACT-watch girl beautiful that  
'He watched that beautiful girl.'

*Passive*

(2) \textit{Gadis cantik itu di-tonton (oleh dia).}  
girl beautiful that NONACT-watch by 3SG  
'That beautiful girl was watched (by him).'

*Non-active, non-passive*

(3) \textit{Gadis cantik itu di-tonton-nya.}  
girl beautiful that NONACT-watch-3SG.GEN  
'He watched that beautiful girl.'

In (3) the Agent is also present in the form of a clitic on the verb. A pronominal agent in an active clause may not be cliticised to the verb, and in a passive clause the only clitic option available to the agent is clisis with the preposition, shown in (4).

(4) \textit{Gadis cantik itu di-tonton oleh-nya.}  
girl beautiful that NONACT-watch by-3SG.GEN  
'That beautiful girl was watched by him.'

When the agent is not pronominal, and the voice selected is the non-active, non-passive shown in (2), the agent appears as part of a close constituent with the verb, as in (5).

(5) \textit{Gadis cantik itu di-tonton adik-ku.}  
girl beautiful that NONACT-watch ySi-1SG.GEN  
'My little brother watched that beautiful girl.'

As can be seen, there are two different \textit{di}-constructions: a passive, and a non-active, non-passive (henceforth \textit{inverse}). Morphologically the passive is distinguished from the inverse by the preposition \textit{oleh} marking the (optional) agent in the passive, and by the inability of a passive agent to appear cliticised to the verb. Syntactically, they are distinguished by the fact that locative adverbials may grammatically intrude between a passive verb and its prepositionally marked agent, but they may not appear between an inverse verb and its bare-DP agent, as shown in (6) and (7).\footnote{Occasionally, especially in written Indonesian, the subject P is found post-verbally. If the clause is passive, with an \textit{oleh}-marked A, the P may occur between the verb and the A. If the clause is inverse then this is not possible. Thus the passive clause in (i) is grammatical, while a similar word order with an inverse clause, shown in (ii), is not.}

*Inverse*

(6) \textit{Gadis cantik itu di-tonton (*di kebun) adik-ku.}  
girl beautiful that NONACT-watch LOC garden ySi-1SG.GEN  
'My little brother watched that beautiful girl in the garden.'

*Passive*

(7) \textit{Gadis cantik itu di-tonton (di kebun)}  
girl beautiful that NONACT-watch LOC garden  
(oleh adik-ku). by ySi-1SG.GEN  
'That beautiful girl was watched (by my little brother) in the garden.'
Further proof of the different syntactic status of these *di*-marked clauses comes from reflexive constructions: only the higher argument can bind, and only the passive denotes an agent to be lower than the theme/patient (eg., Arka and Manning 1998). These data indicate that while the Agent in the passive clause must be considered to be 'lower' than the Patient, in terms of a grammatical functions/thematic role hierarchy, the Agent and the Patient in the inverse clause in (9) show the same relationship to each other as they do in the active clause shown in (8): no demotion of either of the thematic roles has occurred, only a reassigning of core grammatical functions to the different arguments.

**Active**

(8) Dia me-maju-kan diri-nya.
3SG ACT-advance-TR self-3SG.GEN

'He promoted himself.'

**Inverse**

(9) Diri-nya di-maju-kan-nya.
self-3SG.GEN NONACT-advance-TR 3SG.GEN

'He promoted himself.'

**Passive**

(10) Dia di-maju-kan oleh diri-nya.
3SG NONACT-advance-TR by self-3SG.GEN

'He was promoted by himself.'

(11) *Diri-nya di-maju-kan oleh-nya.
self-3SG.GEN NONACT-advance-TR by-3SG.GEN

'Himself was promoted by him.'

Just as the reflexive *dirinya* in (11) cannot bind its antecedent, so too we find that the reflexive cannot be the Agent in either an active or an inverse clause: compare *dirinya memajukannya* with (8), and *dia dimajukan dirinya* with (9). Given these reflexive facts, and the data on other grammatical tests not discussed here, we can conclude that the following relations between grammatical functions and syntactic roles hold in the different voices.

---

7 Basically, core > non-core, and within each category participants are ranked by thematic role.
there are some differences to the Standard Indonesian usage (see 4.2 below for these points). Its use can be seen in (16) and (18). Note that, unlike the voice morphology seen so far, the prefix ter- may appear on intransitive predicates; in addition to (16), note the grammaticality of (17), indicating that ter- is not simply fossilised morphology, best treated synchronically as part of the root of this predicate, but must be taken to have a semantic derivational function.

(16)  
Dia ter-jatuh dari atas pohon.  
3SG ACC-fall from above tree  
‘He (accidentally, happened to) fell from up in the tree.’

(17) 
Dia jatuh dari atas pohon.  
3SG fall from above tree  
‘He fell from up in the tree.’

(18)  
Gadis cantik itu ter-lihat (oleh adik-ku).  
girl beautiful that ACC-watch by 3SG GEN  
‘That beautiful girl was seen (by my little brother).’  
‘That beautiful girl was visible (to my little brother).’

We have seen that there are three voice types in Standard Indonesian, which can be distinguished on the basis of order in the clause, marking on the verb, and marking on DPs.

• Order in the clause:
  The Subject is preverbal; Objects and Obliques are postverbal; Obliques may follow locational adjuncts, while Objects must precede any such PPs.

• Verbal marking:
  An Agent Subject is shown by the prefix meng- on a bivalent verb root; a Patient Subject is shown by the prefix di- on a bivalent verb root.

• DP marking:
  A Subject or Object appears in a bare DP with no case marking; Obliques are marked; oblique Agents are marked with the preposition oleh.

This combination of features allows us to construct three different voices, as summarised in (12), even though there are not three dedicated morphological paradigms. For clauses in which the subject is a local person, either a first or second person (singular or plural) pronoun or an address term used to refer to the speaker or the hearer, we do not find the passive coding option, only the active and the inverse; this was demonstrated in (13)-(15). Additionally, as seen in the contrast between (5) and (14), the elitic form of an Agent in an inverse clause differs depending on whether it is a local Agent or a non-local Agent. Local agents require enclisis to the verb, while the 3SG elitic is enclisis.

Having established the relevant morphosyntax for a Standard variety of Indonesian, we shall now investigate how voice is realised in non-standard varieties.

4 Other Varieties of Indonesian/Malay

There are many named varieties of Indonesian/Malay, of which only a few, geographically and typologically distinct varieties are examined here. In addition to these named varieties there are a very large number of nameless varieties of the language, typically known in their area simply by a name something like bahasa pasar ‘market language’.⁸

Proceeding from West to East, the varieties that are reported on here in some detail are: Standard Indonesian; Jakarta Malay; Makasar Malay; Kupang Malay; Ambon Malay; Serui Malay; Papua Malay. In addition to these, supporting data from Perak Malay, Kelantan Malay and Standard Malay will be introduced where appropriate and where available. The location of these different language varieties can be seen on the map.

The parameters defining voice that shall be examined here are:

• The kind of voice (active, passive, inverse, ...) represented by a particular morphosyntactic configuration (for example, the combination of DP belget + Verb-en (by DP) in English marks a passive construction);

• The number of voices distinguished in the language variety (English, and most languages, have a two-way voice distinction, but varieties of Malay show up to three different voices);

⁸ A partial list of some of the excluded named varieties of Malay/Indonesian would include: Pattani Malay, Baba Malay, Bazar Malay / 'Colloquial Indonesian', North Matuku Malay, Sarawak Malay, Brunei Malay, Luyataka Malay, Dobol Malay, Merauke Malay, Cocos Malay, and Sri Lanka Malay. They have not been discussed here because not enough description is available on their syntax.
Any restrictions on voice selection that apply when there are local persons in the clause (we have seen that Standard Indonesian does not permit a passive construction with a first or second person agent);

- The morphological form of affixing (affixal or analytical marking, and the form of the marking);

- The morphological form of the by-phrase marker, and any other functions it has in the language.

The next sections present this information for eight varieties of Indonesian/Malay, describing the voice system with Standard Indonesian, as described at the beginning of section 3, as a reference point.

4.1 Kelantan Malay

The Kelantan Malay voice system is radically different to SI. The passive is marked with the independent word/auxiliary apo (no etymology is available for this word), and marking by-phrases with a choice of two non-unique coding strategies: while SI employs the dedicated by-phrase preposition oleh, in Kelantan Malay a by-phrase is marked with either gemination (used in this variety to case mark non-subjects, including objects, obliques and the objects of prepositions), or an overt preposition, either di(1) ‘locative’ or kə ‘allative’. The active voice is unmarked, as seen in (19), making decisions about relative markedness easy to resolve.

(19) Kucî maké iké
cat eat fish
‘The cat ate the fish.’

(20) Ke pois maké k-kucî.
fish PASS eat NON.SBJ-cat
‘The fish was eaten by the cat.’

(21) Duyiye tu pois ame? di amo taidi.
durian that PASS take LOC ISG earlier
‘That durian was taken by me earlier.’

There is no voice corresponding to the inverse voice of SI; Kelantan Malay shows a simpler two-voice alternation between active and passive alone. The word order of the major constituents in Kelantan Malay is the same as SI, with the addition of the ‘auxiliary’ apo. Further details on the constraints operating in this kind of system can be found in the subsequent discussion of Ambon Malay, which has a similar system.

4.2 Standard Malay

The Standard Malay voice system follows SI in all respects except for the accidental prefix, which (in addition to the uses described for SI above) is also used in active sentences to mark an unintentional event.

(22) Adik ter-makan lauk yang pedas itu.
ySi ACC-eat side.dish REL hot that
‘(My) little brother ate some of that hot side dish (by accident).’

While there are many differences between Standard Malay and Ulu Muar Malay, this use of the accidental prefix in active clauses is also reported in that latter Malay variety (and is also found in Riau and Singapore Malay, indicating that it has some geographic spread in the western part of the Malay area). In Standard Indonesian, however, this is not a grammatical use of ter-.

Because of the clearly semantic-derivational nature of this active ter-, and its irrelevance to an analysis of voice (compare (22) to (1) earlier; in both cases the Agent is SUBJECT and the Patient is OBJECT), Standard Malay will not be discussed further as an entity distinct from Standard Indonesian.

4.3 Jakarta Malay

The voice system of Jakarta Malay (also known as Betawi, though this term is ambiguous) initially appears to be very similar to SI, except that there are no restrictions on the coding choices available for local agents; compare (25) below with (12) and (13) in section 3. There are differences: the language also makes much more frequent, and sometimes lexicalised, use of applicatives than does SI or Standard Malay, and this is reflected in constraints that affect the voice system as well (described in 4.3). One instance of the applicative can be seen in the predicate hit ‘see’, which requires the applicative -in order to overtly express the Patient as an object. The verb inherently specifies (SUBJ, OBL), and cannot code the Patient as a term without the applicative morphology.

In terms of differences in the voice system, the possibility of an oblique local person is illustrated in (24); as seen in the ungrammaticality earlier in (12), this is not possible in Standard Indonesian/Malay.
4.4 Makasar Malay

The voice system in Makasar Malay shows the same lack of distinction in coding possibilities between local and non-local person that we saw in Jakarta Malay, but additionally lacks any passive constructions. There are voice alternations, but only between active and inverse clauses. As with Jakarta Malay, applicatives are more prominent in the grammar of Makasar Malay than they are in Standard Indonesian, and a constraint found in languages with productive use of dynamic applicative alternations (of the form *Obl/Ex, specifying that prominent participants should not be encoded as non-terms — see section 5 for more discussion) is also prominent in the list of constraints governing the operation of the voice system in Makasar Malay. The active and the inverse voices both show two coding possibilities, one with prefixal agreement (which has full pronominal status) and one without; if there is prefixal agreement, then this takes the place of the voice-indicating morphology. Note that the agreement prefixes, which index the agent, are different in the active and the inverse clauses, for the 1SG at least. Apart from the choice of pronominal marking on the verb the word order of the active and the inverse clauses will also be different, with the active clause showing an AVP or VPA order, and the inverse a PVA or PAV order.

(26) Saya me-[m]ukol dorang ~ (27) Say-pukol dorang
1SG ACT-hit 3PL
'I hit them.'

3PL INV-hit-1SG 3PL
'They were hit by me.'
4.6 Serui Malay

The voice system of Serui Malay is similar to that of Ambon Malay, but not completely identical. The verb that marks the passive is the same as Ambon Malay, *dapa* 'get', but the marker for the oblique agent is the comitative *deng*, rather than the ablative *dari*. More significantly we find that there is a restriction on the coding of local-person agents: they cannot be coded obliquely. This means that while the translation equivalent of (32) is acceptable in Serui Malay (as *Adi dapa pukol deng dorang*), there is no passive variant of a clause with a local agent. We can see the difference in the ungrammatical attempt to form a passive with a local agent, as in (37). In Ambon Malay the equivalent sentence is acceptable, as AM: *Adi dapa pukol dari beta.*

(36)  
*Adi dapa pukol (deng dorang).*  
YSi get hit with 3PL.  
'(My) younger brother was hit (by them).'

(37)  
*Adi dapa pukol deng kita*  
YSi get hit with 1SG  
'(My) younger brother was hit by me.'

When both participants are third person, the passive may be used freely, following pragmatic constraints.

(38)  
*Bini tu pukol adi,*  
woman that hit YSi  
'That woman hit (my) younger brother.'

(39)  
*Adi dapa pukol deng bini tu,*  
YSi get hit with woman that  
'(My) younger brother was hit by that woman.'

No positive information on active uses of *dapa* in active clauses is available for Serui Malay. Sentences analogous to (34) and (35) have not been observed, but also have not been ruled out as ungrammatical.

4.7 Kupang Malay

No system of voice alternations has been reported for the Malay of Kupang, though the accidental prefix is attested. In contrast to its use in more western varieties, no oblique agents have been reported with this construction. An example can be seen in (40).

(40)  
*Su-ta-robek-ta-robek dia puŋ sisi.*  
PERF-ACC-tear-RED 3SG POSS side  
'It's side was already torn in various places.'

The fact that nearby local languages (Manggarai and Palu’e of Flores, for instance) have been shown to have functional voice systems without the presence of any verbal morphology or auxiliary, or morphological byphrase marking, relying purely on a word order contrast to code the difference, means that the contrast seen in (41) and (42) calls for further investigation for its possible status as a voice alternation. The AVP clause in (41) is a typical transitive clause, and the PAV clause in (42) has been analysed as being a variant of (41) with a topicalised P, but it might equally well be analysed as a passive variant of (41). While topicalisation is a regular feature of all Indonesian/Malay varieties (and most other languages of Southeast Asia), and is definitely found in the languages of the Kupang area, this does not exclude the possibility of a passive interpretation of (42).

(41)  
*Beta makan nasi-tu.*  
1SG eat rice-that  
'I ate rice.'

(42)  
*Nasi-tu beta makan.*  
rice-that 1SG eat  
'That rice, I ate.' ~ 'That rice was eaten by me.'

Donohue (2005) discusses the morphologically unmarked voice system of Palu’e, which might be analogous to the alternation shown here in (41) and (42). If a similar analysis applied to the local Malay varieties we would have an additional, and typologically quite distinct, voice system.

4.8 Papuan Malay

There is no voice system in the Malay of north-east Papua, around Jayapura. All clauses are strictly active, regardless of the status of the agent and the patient (younger, more urban, speakers now allow an Ambon-style system of voice alternations, but the description here is true for more con-
servative varieties). In a sense, then, this easternmost Malay variety represents the simplest voice system from among those varieties surveyed here, since it allows for no alternations, and thus avoids participation in any of the secondary questions about the organisation of the voice system.

(43) Dorang pukol kita pu adi.
   3PL hit 1SG GEN ySi
   "They hit my younger brother."

(44) *Adi (dapa/kena) pukol
   ySi get/etc. hit
   '(My) younger brother was hit.'

Although there is no voice alternation there are productive and frequent uses of topicalisations, some of them constrained by the relative animacy of the participants, and in some varieties there is a switch reference system between coordinated clauses. While beyond the scope of the current paper, this information shows that Papuan Malay has the grammatical resources to make as many distinctions as do the more western varieties examined here.

4.9 Overview

The morphological variation can be summarised as in table 1. Here the different varieties are shown in a roughly west-east arrangement from the top of the table to the bottom, which also approximately reflects the direction of language migration from west to east. The voice-marking morphology is shown in the left three columns; the convention [ ] is used to indicate an analytical, rather than morphological, form. Clearly the accidental marker is the most robust, or stable, being found in minimally-changed form in the greatest range of varieties (and, probably not coincidentally, being present in a large number of substratal languages in the areas in which Malay varieties have propagated). There is a perfect symmetry in the distribution of the morphological active and inverse voice prefixes, with no variety existing that has one and not the other. If only one voice is overtly marked, it is the passive and it is marked peripheristically, not morphologically.

<table>
<thead>
<tr>
<th>Voice marking</th>
<th>Oblique agent</th>
<th>Object agent</th>
<th>Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>non-Active</td>
<td>Accidental</td>
<td>x-V V-x</td>
</tr>
<tr>
<td>SI/SM</td>
<td>map</td>
<td>di-</td>
<td>ter</td>
</tr>
<tr>
<td>JM</td>
<td>ay</td>
<td>di-</td>
<td>te(r)</td>
</tr>
<tr>
<td>MM (map')</td>
<td>di-</td>
<td>ter-</td>
<td>-</td>
</tr>
<tr>
<td>KeM</td>
<td>-</td>
<td>[aza]</td>
<td>?</td>
</tr>
<tr>
<td>PeM</td>
<td>p</td>
<td>di-</td>
<td>ta</td>
</tr>
<tr>
<td>AM</td>
<td>-</td>
<td>[dapa]</td>
<td>ta</td>
</tr>
<tr>
<td>SBM</td>
<td>[kena]</td>
<td>ta-</td>
<td>✓</td>
</tr>
<tr>
<td>'BM'</td>
<td>-</td>
<td>[kena]</td>
<td>ter-</td>
</tr>
<tr>
<td>SeM</td>
<td>-</td>
<td>[dapa]</td>
<td>ta</td>
</tr>
<tr>
<td>KM?</td>
<td>-</td>
<td>-</td>
<td>ta</td>
</tr>
<tr>
<td>PM</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>


We have noted above that the presence of a morphological passive/inverse implies the presence of a marked active as well. This would appear to argue that there is no markedness relationship between the active and the nonactive voices. Since, however, there are also local agential object variants in exactly these languages, such as were seen in (12), in which there is no morphological marking of the voice selection, simply agreement for the agent, we can state that it is the active voice that is more often more morphologically marked (though see Chung 1978 for a discussion of some complications for this simplistic analysis).

Implicationally we can state that the presence of an object agent correlates with the presence of morphological voice marking. Furthermore, the presence of local person enclitics marking the agent on inverse verbs, such as are found in Jakarta Malay and Makasar Malay, implies the presence of third person enclitics in the same function. When the voice selected is passive, there is usually the possibility of coding the agent in a by-phrase with a proposition; some varieties, such as many Bazaar Malay varieties, do not allow an oblique agent with their kena-marked passive. A wide variety of by-phrase markers are used, from the completely specialised oleh of Stan-
dard Malay to the more general prepositional markers that are found in other varieties. Most possibilities are covered, as can be seen in table 2.

*Table 2. Marking an oblique agent*

<table>
<thead>
<tr>
<th>by-phrase</th>
<th>Other uses</th>
<th>Attested in:</th>
<th>SI/SM:</th>
</tr>
</thead>
<tbody>
<tr>
<td>oleh</td>
<td></td>
<td>SM, SI</td>
<td>oleh</td>
</tr>
<tr>
<td>dari</td>
<td>‘from’</td>
<td>AM</td>
<td>dari</td>
</tr>
<tr>
<td>di</td>
<td>‘at’</td>
<td>KeM</td>
<td>di</td>
</tr>
<tr>
<td>de'</td>
<td>‘at’</td>
<td>PeM, UMM</td>
<td>di</td>
</tr>
<tr>
<td>C-</td>
<td>(non-subject)</td>
<td>KeM</td>
<td>(all non-subjects)</td>
</tr>
<tr>
<td>ko</td>
<td>‘to(ward)’</td>
<td>KeM</td>
<td>ke(pada)</td>
</tr>
<tr>
<td>ame</td>
<td>‘with’</td>
<td>JM</td>
<td>sama</td>
</tr>
<tr>
<td>deng</td>
<td>‘with’</td>
<td>SrM, SBM</td>
<td>dengan</td>
</tr>
</tbody>
</table>

The abbreviations for language names used here are the same as those used in Table 1.

5 Modelling the Variation

The different voice systems described above can be modelled with reference to constraints on the appearance of different argument types in different grammatical functions in a sentence. The model followed here assumes the background (and many of the formal mechanisms) that have been introduced in recent work such as Legendre, Raymond and Smolensky (1993), Aissen (1999), and Sellis (2001). The following constraints on associating grammatical functions with different syntactic roles, persons, or arguments of different pragmatic statuses will be relevant to the exposition that follows, though in most cases only three or four of the most highly-ranked of these constraints are necessary to correctly predict the possibilities that emerge. In these cases only those constraints relevant to the exposition will be listed in the subsections that follow.

*Obl/Agt* Do not assign the grammatical function object to arguments bearing the syntactic role of Agent.

This constraint requires that an Agent must be coded as either subject or oblique.

Note that, in contrast to the languages discussed in Aissen (1999), this constraint is not universally ranked highly, and invariably, in all of the Austronesian languages under discussion. The constraint clearly rises in prominence as one heads east through the archipelago (towards less conservative languages), but is equally clearly violated in the western, more conservative, languages.

*Obl/x* Do not assign the grammatical function object to arguments which are low in (discourse) prominence.

This constraint prescribes against objects being low in prominence.

*Obl/L/Agt* Do not assign the grammatical function oblique to arguments bearing the syntactic role of Agent.

This constraint rules out passive codings of events, since they would assign an oblique grammatical function to the Agent.

*Obl/L/X* Do not assign the grammatical function oblique to participants which represent local (first or second person) persons if they are agents.

This constraint rules out passives with local persons. I express it as *Obl/L/Agt*, rather than the simpler *Obl/L*, because there are instances of verbs that select for a subject and an oblique argument that allow local obliques. An example is Dia cinta pada saya ‘She loves me.’

*Obl/X* Do not assign the grammatical function oblique to arguments which are high in (discourse) prominence.

This constraint, the mirror of both *Obl/x* and *Subj/x*, calls for any prominent participants to be coded as core arguments, either subject or object.
In English only active and passive encodings are available, but in many of the languages there are additional possibilities, as we have seen in section 3. Rather than having all of the possibilities in the table illustrated in tableaux, only the one indicated in bold font will be exemplified; the other encodings follow from the same set of constraint rankings. Thus from table 3 only the tableaux for the sentence I was seen by her/him would be shown; in this case, the higher topicality of the Patient means that it is coded as the subject (the constraint rankings for English are the same as those for Ambon Malay, discussed below in 4.5).10

| (45) |
|---|---|---|---|
| V (Agt/3/x, Pat/L/X) | *Obl/Agt | *Subj/x | *Subj/Pat |
| Active A/Obj > P/OBJ | | | |
| Inverse A/OBJ > P/Subj | *! | | |
| Passive A/OBL > P/Subj | | | |

I shall assume that, as a guiding heuristic, the number of constraint violations a particular construction produces should give a rough idea of the relative frequencies of the different choices in natural discourse. That is, a construction with no violations should be more frequent than one that produces a large number of violations. Thus in (45) we can see that the coding choice I was seen by her/him results in one violation. One the other hand He/She saw me, if the Patient me is not more highly prominent than the Agent, would represent an active voice selection that shows no violations at all: all of *Obj/Agt, *Subj/x and *Subj/Pat are satisfied in the evaluation of the coding candidates. From the relative number of constraints violated we should be able to predict that active sentence codings such as Her/She saw me will be more frequent than sentences such as I was seen by her/him, and this is in fact borne out by any textual counts. The same assumptions can be made about the relative frequency of the candidates in the different languages described below.

I shall start the modelling by examining the active, inverse and passive alternations in Standard Indonesian.

10 In the exemplification here for English I have only shown tableaux for one of the eight sentences shown in table 3. This is true of the exemplification of the other languages in this section, and is a practice that has been followed partly for reasons of space, and partly to avoid repetition. I have exemplified the same functional position (Agt/L > Pat/F) in each case, supplemented by anything that is particularly comment-worthy.
5.1 Standard Indonesian/Malay

In Standard Indonesian/Malay the main consideration, as noted earlier in section 2, is that the passive voice is not available for local person Agents. Given Agents and Patients of local and non-local persons, the coding possibilities that are found are illustrated in table 4, using the clause ‘if(s)he saw me/you/them.’ For this variety we find active clauses in (i), (iii), (v) and (viii), inverse clauses in (ii), (iv), (vi) and (ix), and passives in (vii) and (x). The active clause in (viii) below is equivalent to sentence (1) earlier, while (ix) and (x) correspond in terms of morphosyntactic structure to sentences (3) and (2). Note that the first and second persons allow a more general "proclitisation" in inverse clauses; as an alternative to the clitic forms, the free pronouns may be used in this position: Engkau saya lihat is grammatical for (ii), for instance, using a free form of the first person singular pronoun. Similarly, Engkau Budi lihat or Engkau guru lihat is grammatical if the speaker’s name is Budi or occupation is teacher (guru). Similar constructions with names or occupations in place of a second person pronoun are also grammatical. This sort of proclitisation does not apply to third person Agents, and it does not affect the grammar of the clauses.11

Table 4. Coding options in Standard Indonesian/Malay

<table>
<thead>
<tr>
<th>Agent</th>
<th>Patient</th>
<th>Local</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>(i) Saya me-lihat engkau</td>
<td>(iii) Saya me-lihat dia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ii) Engkau ku-lihat</td>
<td>(iv) Dia ku-lihat</td>
<td></td>
</tr>
<tr>
<td>third</td>
<td>(v) Dia me-lihat saya</td>
<td>(vii) Dia me-lihat dia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(vi) Saya di-lihat-nya</td>
<td>(ix) Dia di-lihat-nya</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(vii) Saya di-lihat oleh-nya</td>
<td>(x) Dia di-lihat oleh-nya</td>
<td></td>
</tr>
</tbody>
</table>

The constraint ranking that I assume to be the determining factors for this data set are shown in (46).

In Standard Indonesian/Malay the most salient constraint is *Obl/L/Agt; there is an absolute ban on oblique local arguments. This means that while inverse clauses can appear with all kinds of Agent, local or non-local, passive clauses are only grammatical with third person Agents. This bans passive constructions with first or second person agents; it does not, however, ban voice alternations, since a non-subject 1st or 2nd person Agents can be coded as objects, due to the lack of a high-ranked *Obl/Agt constraint. *Obl/X plays a role, though not as prominent a role as it does in Philippine-type languages; this constraint operates in tandem with *Subj/X, a general constraint that prescribes against subjects that are low in prominence. Unusually, from a cross-linguistic perspective, *Subj/Agt is a prominent constraint in the language. While it is not unusual in the western Austronesian languages, it is unusual from the point of view of other language families (Aissen 1999; though see section 7).

For the exposition of Standard Indonesian/Malay I shall present a complete explication of the tableaux that result in all of the grammatical sentences shown in table 4, involving all combinations of local and non-local agent, with prominent and non-prominent participants. For the subsequent sketches of voice systems I shall only illustrate a small selection of the different grammatical sentences, enough to show the differences in candidate selection, and shall only discuss the four highest-ranking constraints, since these constraints determine all (or at least the majority) of the decisions on candidates. I shall indicate any relevant violations of lower-ranking constraints.

Examining the sentence shown as (i)-(iv) in table 4, involving a local Agent and either a local or a non-local Patient, we find that the use of the active voice is restricted to instances in which the Agent is more prominent than the Patient. When this condition is not met, the inverse voice forms shown in (ii) and (iv) (and repeated below as (47) and (48)) will be used. Different results hold when the Agent is not more prominent than the Patient: an active voice selection, shown in (i) and (iii) and repeated below as (49) and (50), is found.12

11 Some speakers, for instance the speech represented in Kaswandt Purwo (1988), allow third persons to appear in a clitic-like position preceding the verb. This is not a feature of all (formal) Indonesian varieties.

12 This is because the only constraint that is relevant to the identity of a participant as local or non-local is the highly-ranked *Obl/L constraint. This applies equally for local Agents acting on Patients of any description.
Inverse coding

(47) Engkau ku-lihat.
   2SG 1SG.INV.AGT-see
   'I saw you.'

(48) Dia ku-lihat.
   3SG 1SG.INV.AGT-see
   'I saw her/him.'

Active coding

(49) Saya me-lihat engkau.
   1SG ACT-see 2SG
   'I saw you.'

(50) Saya me-lihat dia.
   1SG ACT-see 3SG
   'I saw her/him.'

The tableaux that will generate these two sentences are shown as (51)-(54). These tableaux show local persons acting on local or third persons, modelling sentences (47) and (49) above. Substituting the third person Patients for local patients will produce sentences (48) and (50) without any change in the patterns of constraint violations, since person of the Patient does not play a role in determining a candidate’s fitness.

Tableaux with local agents

(51) These tableaux select sentences (47) and (48).

<table>
<thead>
<tr>
<th>V(Agt/L/x, Pat/L/x)</th>
<th>*OBL/L/Ag</th>
<th>*OBL/L/X</th>
<th>*SUBJ</th>
<th>*SUBJ/Ag</th>
<th>*SUBJ/X</th>
<th>*OBL/Ag</th>
<th>*OBL/X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(52) These tableaux select sentences (47) and (48).

<table>
<thead>
<tr>
<th>V(Agt/L/x, Pat/L/x)</th>
<th>*OBL/L/Ag</th>
<th>*OBL/L/X</th>
<th>*SUBJ</th>
<th>*SUBJ/Ag</th>
<th>*SUBJ/X</th>
<th>*OBL/Ag</th>
<th>*OBL/X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(53) These tableaux select sentences (47) and (48).

<table>
<thead>
<tr>
<th>V(Agt/L/x, Pat/L/x)</th>
<th>*OBL/L/Ag</th>
<th>*OBL/L/X</th>
<th>*SUBJ</th>
<th>*SUBJ/Ag</th>
<th>*SUBJ/X</th>
<th>*OBL/Ag</th>
<th>*OBL/X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(54) These tableaux select sentences (49) and (50).

<table>
<thead>
<tr>
<th>V(Agt/L/x, Pat/L/x)</th>
<th>*OBL/L/Ag</th>
<th>*OBL/L/X</th>
<th>*SUBJ</th>
<th>*SUBJ/Ag</th>
<th>*SUBJ/X</th>
<th>*OBL/Ag</th>
<th>*OBL/X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When the Agent is third person, matching the examples shown as (v)-(x) in table 4 we find a change in the patterns of constraint violations, due to the accessibility of passive coding options now that the *OBL/L/Ag constraint does not create any violations. Additionally, the fact that there are no third person proclitics, only enclitics, changes the appearance of the clause, though not the assignment of grammatical functions (the morphological realisation of voice is discussed in section 7). The three possibilities we find are shown in (55)-(60), showing variants with both local and non-local Pa-
tients, which behave identically for the purposes of candidate evaluation. The tableaux shown as (61)-(64) illustrate the selection of the clauses in (v)-
(vii) from table 4 (and repeated as (55), (57) and (59) below); again, the
sentences shown as (viii)-(x) (corresponding to (56), (58) and (60)) will be
evaluated with an identical set of constraints and violations, with (viii) cor-
responding to (v) in terms of the outcomes, (ix) corresponding to (vi), and
(x) corresponding to (vii) (thus, similar resolutions are found for (55) and
(56), (57) and (58) and (59) and (60)).

Active coding

(55)  Dia me-lihat saya.
3SG ACT-see 1SG
'S/he saw me.'

(56)  Dia me-lihat dia.
3SG ACT-see 3SG
'S/he saw her/him.'

Inverse coding

(57)  Saya di-lihat-nya.
1SG NONACT-see-3SG
'S/he saw me.'

(58)  Dia di-lihat-nya.
3SG NONACT-see-3SG
'S/he saw her/him.'

Passive coding

(59)  Saya di-lihat (oleh-nya).
1SG NONACT-see by-3SG
'I was seen by her/him.'

(60)  Dia di-lihat (oleh-nya).
3SG NONACT-see by-3SG
'She/he was seen by her/him.'

<table>
<thead>
<tr>
<th>Tableaux with non-local agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>(61) These tableaux select sentences (59) and (60).</td>
</tr>
<tr>
<td>V(Agt/3/3, Pat/L/X)</td>
</tr>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
</tr>
</tbody>
</table>

(62) These tableaux select sentences (59) and (60).

| V(Agt/3/3, Pat/L/X) | *OBL/Ag|  *Subj/Ag|  *Obj/Ag|  *Ob/Ag|  *Subj/Pat|  *Ob/Ag|  *Obl/Ag|
| Active A/SUBJ > P/OBJ | * | | | | | |
| Inverse A/OBJ > P/SUBJ | * | | | | | |
| Passive A/OBL > P/SUBJ | * | | | | | |

(63) These tableaux select sentences (57) and (58).

| V(Agt/3/3, Pat/L/X) | *OBL/Ag|  *Subj/Ag|  *Obj/Ag|  *Ob/Ag|  *Subj/Pat|  *Ob/Ag|  *Obl/Ag|
| Active A/SUBJ > P/OBJ | | | | | | |
| Inverse A/OBJ > P/SUBJ | | | | | | |
| Passive A/OBL > P/SUBJ | | | | | | |
These tableaux select sentences (55) and (56).

<table>
<thead>
<tr>
<th>V(Agt/3/X, Pat/L/X)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(64)</td>
<td>Active A/RE &gt; P/Obj</td>
<td>Inverse A/Obj &gt; P/Subj</td>
<td>Passive A/Obl &gt; P/Subj</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td>!</td>
<td>!</td>
<td></td>
</tr>
</tbody>
</table>

The constraint interactions from the different tableaux that have been presented as (51)-(54) and (61)-(64) have been summarised in table 5. Here we can see for each of the sixteen tableaux which of the constraints was the one that led to a particular voice candidate being rejected. In only four cases is the constraints that determines the final successful candidate not one of the four highest-ranking constraints; in these cases *Obj/x determines the choice of passive rather than inverse voice when a non-local Agent is low in prominence. These instances have been shown in boldface in table 5.

We can see that a small subset of the total number of applicable constraints is sufficient to model 75% of the variation we find, with only the selection of passive voices (rather than inverse) requiring reference to any but the four most highly-ranked constraints. For the sake of brevity I shall present tableaux for the following language varieties showing only the interaction of the four most highly-ranked constraints; where it is significant, I shall indicate which of the lower-ranked constraints is responsible for the selection of the successful candidate, or at least indicate which of the lower-ranked constraints are violated by the successful candidate.

I next show the modelling of a northern Philippine language, Tagalog, as an example of the kind of morphosyntactic system that the Indonesian/Malay ones are believed to have developed from. This will be followed by a sample of the better-known Malay varieties, proceeding cast from the Standard Indonesian variety presented in this section.

| Table 5. Constraints leading to the selection of different voice choices |
|---------------------------|---|---|---|---|
| i | L > L | x > x | *SUBJ/Agt | √ | *OBL/L/Agt | Inverse |
| ii | x > X | *SUBJ/x | √ | *OBL/L/Agt | Inverse |
| iii | X > X | *SUBJ/Agt | √ | *OBL/L/Agt | Inverse |
| iv | X > x | √ | *SUBJ/x | *OBL/L/Agt | Active |
| v | L > 3 | x > x | *SUBJ/Agt | √ | *OBL/L/Agt | Inverse |
| vi | x > X | *SUBJ/x | √ | *OBL/L/Agt | Active |
| vii | X > X | *SUBJ/Agt | √ | *OBL/L/Agt | Inverse |
| viii | X > x | √ | *SUBJ/x | *OBL/L/Agt | Active |
| ix | 3 > X | x > x | *SUBJ/Agt | *OBL/x | √ | Passive |
| x | x > X | *SUBJ/x | *OBL/x | √ | Passive |
| xi | X > X | *SUBJ/Agt | √ | *OBL/x | Inverse |
| xii | X > x | √ | *SUBJ/x | *SUBJ/x | Active |
| xiii | 3 > X | x > x | *SUBJ/Agt | *OBL/x | √ | Passive |
| xiv | x > X | *SUBJ/x | *OBL/x | √ | Passive |
| xv | X > X | *SUBJ/Agt | √ | *OBL/x | Inverse |
| xvi | X > x | √ | *SUBJ/x | *SUBJ/x | Active |

5.2 The Source? Philippine-type Languages
Most reconstructions of Malayo-Polynesian languages assume that the parent language had a structure similar to that attested today in most languages of the Philippines, and for the sake of simplicity this view shall be assumed here as well. Note that an alternative analysis, in which a reconstruction of the syntax of proto Malayo-Polynesian reveals a more Chamorro- or Nias-like morphosyntax (a fact suggested by the distribution of these, and other, three-voice languages in the Austronesian world), would have a different view on the original system, one that is not so divergent from the Indonesian/Malay model (for reference materials underlying this proposal, see materials on Chamorro [Topping 1973, Gibson 1992], Nias [Brown 2001, Sundermann 1905, 1913], or Tukang Besi [Donohue 1999]).

The salient facts about Philippine-type languages, and to a lesser extent their relatives in (western) Indonesia, that diverge from general typological norms, as far as this paper is concerned, are summarised in the following points. Philippine languages:
typically have many voice options, beyond simply allowing either Agent and Patient as SUBJ, as is found in English
this correlates with the fact that prominent participants of any role are preferentially coded as subject. In Indonesian languages that lack a multiple voice system there is heavy and productive use of applicatives.

- do not have a demotional voice system non-subject agents are objects, not obliques. There is no passive voice in the northern languages, but Indonesian languages typically do allow a passive.

- show a textual and lexical preference for non-active voice(s) some lexemes cannot appear in main clauses with an active voice, but must code their Patient as SUBJ. Textually, approximately 75% of clauses are not in active voice (see, for instance, Bresnahan 1991 for Tagalog), but appear with a non-Agent as SUBJ.

Note that there are no constraints on the assignment of grammatical functions to particular persons: any of first, second or third persons may be subject, or object, as either Agent or Patient. In Table 6 we can see the different coding options for local and non-local (pronominal) agents, and local or non-local (pronominal) patients. The verb used to exemplify the system is kita 'see', with the voice choices na- 'inverse' and naka- 'active'. In all cases the verb is bivalent, with both a subject and an object, and in all cells in the table we find that both the active and the inverse option are allowed.

---

13 A pronoun coded as an object Patient receives dative morphology, as can be seen in Table 6, a feature normally associated with obliques (Donohue 2002 discusses the conditions under which dative case is found). This dative-coded argument is nonetheless demonstrably not an oblique, since it does not allow adjunct fronting (Krueger 1993), which involves the pronominal clitics following it to a preverbal position. Sentence (i), based on (v) in Table 6, is grammatical, while a similar construction is grammatical with true obliques, shown in (ii).

(i) Sa akih=

DAT 3SG.DAT=3SG.NOM AV-see
'He/she saw me.'

(ii) Sa 
palengke=

DAT market=3SG.NOM GO.AV.
'He/she went to the market.'

---

Table 6. Coding options in Tagalog

<table>
<thead>
<tr>
<th>Agent</th>
<th>Patient</th>
<th>Local</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(i) Naka-kita ako sa iyo</td>
<td>(ii) Naka-kita ko siya</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(v) Naka-kita siya sa akih</td>
<td>(vi) Na-kita niya ako</td>
</tr>
</tbody>
</table>

The salient constraints that we need to model these facts, and their relative rankings, are shown in (65).

Primary constraint rankings for Philippine-type languages (specifically Tagalog)

(65) *Obl/Agt » *Subj/x » *Obl/X » *Subj/Agt

The *Obl/X and *Obl/Agt constraints are the most salient in this language in determining the coding choices. The ban on oblique agents is satisfied by not allowing any passive voice options.14 The ban on obliquely-coded participants that are high in discourse prominence is satisfied in Philippine-type languages by having a range of voice choices (sometimes called 'superapplicatives' in the literature) which appear in clauses with a non-subcategorised for participant as subject. Through the use of these 'non-basic' voice options any prominent participant will be preferred as subject.

For instance, in (66) we can see a monovalent verb used with an adjunct; the subject of the clause is the agent siya, and the adjunct 'car' is marked with dative case, governed by the preposition mula 'from'. This same adjunct may, with a change in voice, mark the location as the subject of the clause, as in (67), while the agent appears as a non-subject, non-adjunct.15 The use of the -an suffix in (67) indicates that the ang-marked subject of the clause is a location, and the clitic =mo indicates the Agent of the clause. In (66), on the other hand, the infixed <wa> indicates that the sub-

14 I take *Obl/Agt to prohibit not only overtly expressed oblique agents, but also 'demotion by omission', to characterize agentless passives.

15 We can test the adjunct status of the car with the same adjunct fronting construction used in footnote 3: Mula sa kota= siya lumamon. Note that kota in (67) is marked with a specific preposition, mula, as well as the general dative (loblique) sa.
ject is an Agent (an S or an A), and the oblique ‘car’ is marked with the dative case as.

(66) Lecum>unsad=siya mula sa kotse.
    alight<A/V>=3SG.NOM from DAT car
    'He alighted from the car.'

(67) Lunsar-an=mo a ng kotse.
    alight-DV-2SG.GEN NOM car
    'Alight from the car.'

In (68) and (69) we can see a similar alternation with a bivalent root; here kita ‘see’. When the P is less prominent than the A the active voice must be used to allow the Agent to function as subject. Otherwise the P is coded as subject, as in (69).

(68) Naka-kitan=nko a ng bata.
    AV-see=1SG.NOM GEN child
    'I saw a/the child.'

(69) Na-kita=nko a ng bata.
    PV-see=1SG.GEN NOM child
    'I saw the child.'

The selection of active and inverse voice, such as that in (66), (68) or (69), is shown in (70)-(73), modelling (68) and (69). The inverse voice is selected three times out of four in these tables, reflecting the fact that unless the Agent outranks (and not just matches) the Patient in terms of ‘prominence’, the active voice will not be selected. Of the following tableaux, (70), (71) and (72) will select sentences such as (69), and only (73), with the high-prominence Agent and low-prominence Patient, will select for an active clause such as (68).

(70) Additional low-ranking violations for the winning candidate:
    *Obj/Agt, *Obj/x, *Subj/Pat

<table>
<thead>
<tr>
<th>V(Agt/l/X, Pat/3/X)</th>
<th>*Obl/Agt</th>
<th>*Subj/x</th>
<th>*Obl/x</th>
<th>*Subj/Ag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/ Subj, L/Obl</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Inverse A/ Obj, P/Subj</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/ Obl, P/Subj</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(71) Additional low-ranking violations for the winning candidate:
    *Obl/Agt, *Subj/x, *Obl/Ag |

<table>
<thead>
<tr>
<th>V(Agt/l/X, Pat/3/X)</th>
<th>*Obl/Agt</th>
<th>*Subj/x</th>
<th>*Obl/x</th>
<th>*Subj/Ag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/ Subj, L/Obl</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Inverse A/ Obj, P/Subj</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/ Obl, P/Subj</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(72) Additional low-ranking violations for the winning candidate:
    *Subj/Pat |

<table>
<thead>
<tr>
<th>V(Agt/l/X, Pat/3/X)</th>
<th>*Obl/Agt</th>
<th>*Subj/x</th>
<th>*Obl/x</th>
<th>*Subj/Ag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/ Subj, L/Obl</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Inverse A/ Obj, P/Subj</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/ Obl, P/Subj</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(73) Additional low-ranking violations for the winning candidate:
    *Subj/Agt |

<table>
<thead>
<tr>
<th>V(Agt/l/X, Pat/3/X)</th>
<th>*Obl/Agt</th>
<th>*Subj/x</th>
<th>*Obl/x</th>
<th>*Subj/Ag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/ Subj, L/Obl</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Inverse A/ Obj, P/Subj</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/ Obl, P/Subj</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The selection of the non-active voice in (67) is shown in (74). Even though the location is not a subcategorised element of the predicate, the fact that it is more prominent than the Agent means that it must be coded as subject. I have shown the inverse and passive voice selections in these tableaux for the sake of easy comparison with the other tableaux, even though there is no Patient in the clause, and so these voice selections would be lexically proscribed.

(74) Additional low-ranking violations for the winning candidate:
    *Obj/Agt, *Obj/x |

<table>
<thead>
<tr>
<th>V(Agt/3/X, &lt;Loc/3/X&gt;)</th>
<th>*Obl/Agt</th>
<th>*Subj/x</th>
<th>*Obl/x</th>
<th>*Subj/Ag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/ Subj, L/Obl</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Inverse A/ Obj, P/Subj</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/ Obl, P/Subj</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-voice A/ Obj, D/Subj</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In Tagalog high prominence with any non-Agent participant requires that it be selected as subject, given the rankings of the constraints. The only Agent that can be coded as Subj is one that is more highly prominent than any other participants in the clause, Patient or adjuncts.

While the voice systems of the Indonesian languages do not allow for such a wide range of alternations, the constraint on marking pragmatically salient participants as more than oblique nominals is a pervasive one in the languages that have some form of applicative alternation (see Donohue 2001).

From this beginning, and keeping in mind the list of constraints presented at the beginning of section 5, we can examine the possibilities for the other varieties of Indonesian/Malay under consideration, and how they can be modelled. The same constraints will be examined for each language, though the ranking of the constraints varies from language to language. A summary of the ranking of the most prominent constraints in each language is given in table 12, in section 6.

5.3 Jakarta Malay

Jakarta Malay shows essentially the same system as Standard Malay, except that there are two morphological possibilities for encoding a local Object Agent, shown in (ii) and (iii) (and in (vii) and (viii)), and that it is additionally possible for a local Agent to be coded as an oblique, as in (iv) and (vii).

Table 7. Coding options in Jakarta Malay

<table>
<thead>
<tr>
<th>Agent \ Patient</th>
<th>Local</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>(i) Gue g-liat-in lu</td>
<td>(v) Gue g-liat-in die</td>
</tr>
<tr>
<td></td>
<td>(ii) Lu gue liat-in</td>
<td>(vi) Diegusliat-in</td>
</tr>
<tr>
<td></td>
<td>(iii) Lu di-liat-in gue</td>
<td>(vii) Dedi-liat-in gue</td>
</tr>
<tr>
<td></td>
<td>(iv) Lu di-liat-in ame gue</td>
<td>(viii) Dedi-liat-in ame gue</td>
</tr>
<tr>
<td>third</td>
<td>(ix) Dic g-liat-in gue</td>
<td>(xii) Die g-liat-in die</td>
</tr>
<tr>
<td></td>
<td>(x) Gue di-liat-in die</td>
<td>(xiii) Die di-liat-in die</td>
</tr>
<tr>
<td></td>
<td>(xi) Gue di-liat-in ame die</td>
<td>(xiv) Die di-liat-in ame die</td>
</tr>
</tbody>
</table>

Compared to the constraints seen for Standard Malay we can see that the constraints that model Jakarta Malay have generalised the \(*Obl/Agt \overset{\text{Constraint}}{\rightarrow} *Obl/X \overset{\text{Constraint}}{\rightarrow} *Subj/Agt \overset{\text{Constraint}}{\rightarrow} *Obl/X *Subj/Agt\) constraint to a \(*Obl/X \overset{\text{Constraint}}{\rightarrow} *Obl/X \overset{\text{Constraint}}{\rightarrow} *Obl/X \overset{\text{Constraint}}{\rightarrow} *Obl/X *Subj/Agt\) one, in which prominent arguments are dispreferred in oblique grammatical functions. This, combined with the \(*Obl/X \overset{\text{Constraint}}{\rightarrow} *Obl/X \overset{\text{Constraint}}{\rightarrow} *Obl/X \overset{\text{Constraint}}{\rightarrow} *Obl/X *Subj/Agt\) constraint, are constraints that we would independently need in the language, to drive the highly productive system of applicatives. While applicative marking in Standard Indonesian has historically been a strong factor in the grammar, synchronically the applicative use of the suffix -kan is commonly found only in a small number of verbs (such as kirim 'send', tulis 'write' and masak 'cook'), and in certain set phrases. In Jakarta Malay, on the other hand, the applicative -in is fully productive, and is required on many low-transitive verbs (such as liat, which in Jakarta Malay, but not Standard Indonesian, requires this suffix to code the seen participant as an object).

Primary constraint rankings for Jakarta Malay

(75) \(*Obl/X \overset{\text{Constraint}}{\rightarrow} *Subj/X \overset{\text{Constraint}}{\rightarrow} *Obl/X \overset{\text{Constraint}}{\rightarrow} *Subj/Agt\)

An example of the operation of this set of constraints can be seen in (76), describing the selection of the appropriate coding for the sense 'I saw her/him' with a prominent Patient and a low-prominent Agent.

(76) Additional low-ranking violations for the winning candidate:

\(*Obl/Agt, *Obl/L/Agt, *Subj/Pat\)

<table>
<thead>
<tr>
<th>V(Agent/Patient, Pat/3/X)</th>
<th>(*Obl/X)</th>
<th>(*Subj/X)</th>
<th>(*Obl/X)</th>
<th>(*Subj/Agt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active/Agent&gt; Patient&gt; Object</td>
<td>*\overset{\text{Constraint}}{\rightarrow} \overset{\text{Constraint}}{\rightarrow}</td>
<td>*\overset{\text{Constraint}}{\rightarrow}</td>
<td>*\overset{\text{Constraint}}{\rightarrow}</td>
<td>*\overset{\text{Constraint}}{\rightarrow}</td>
</tr>
<tr>
<td>Inverse/Agent&gt; Patient&gt; Object</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive/Agent&gt; Patient&gt; Object</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(77) Die di-liat-in ame gue.

3SG NONACT-see-APPL by 1SG 'He/She was seen by me.'

Note that a sentence such as (77), while grammatical in Jakarta Malay, is not usual. This is because of the inherent prominence associated with local persons: it would be an unusual stretch of discourse that included a non-prominent first or second person pronoun. Given the constraints in (75) we can see that a clause with a first person Agent and a third person Patient, both prominent, would have a different coding outcome, shown in (78).
(78) Additional low-ranking violations for the winning candidate: *OBL/Agt, *SUBJ/Pat

<table>
<thead>
<tr>
<th>V(Agt/LX, Pat/LX)</th>
<th>*OBL/X</th>
<th>*SUBJ/X</th>
<th>*OBL/Ag</th>
<th>*SUBJ/Pat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td></td>
<td></td>
<td>#1</td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td>*1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3SG NONACT-see-APPL 1SG
'I saw her/him.'

(80) Die gue liat-in.
3SG 1SG see-APPL
'I saw her/him.'

This models the grammaticality, but reported rarity, of sentences such as (79).

5.4 Makasar Malay

In Makasar Malay the voice system resembles the Philippine style system described for Tagalog more closely than many of the other languages of the Indonesian archipelago in that there is no passive alternative.

Table 8. Coding options in Makasar Malay

<table>
<thead>
<tr>
<th>Agent \ Patient</th>
<th>Local</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>(i) Saya me-liat kita'</td>
<td>(v) Saya me-liat dia</td>
</tr>
<tr>
<td>(ii) Say-liat kita'</td>
<td>(vi) Say-liat dia</td>
<td></td>
</tr>
<tr>
<td>(iii) Kita' di-liat-ku</td>
<td>(vii) Dia di-liat-ku</td>
<td></td>
</tr>
<tr>
<td>(iv) Kita' ku-liat</td>
<td>(viii) Dia ku-liat</td>
<td></td>
</tr>
<tr>
<td>third</td>
<td>(ix) Dia me-liat saya</td>
<td>(xi) Dia me-liat dia</td>
</tr>
<tr>
<td>(x) De-liat saya</td>
<td>(ix) De-liat dia</td>
<td></td>
</tr>
<tr>
<td>(x) Saya di-liat-nya</td>
<td>(xii) Dia di-liat-nya</td>
<td></td>
</tr>
</tbody>
</table>

While there is an active-inverse alternation in transitive clauses, and each of these has a proclitic-agent option (only inverse voices with third person agents lack the proclitic possibility), there are no coding options in which the agent is coded obliquely. As with Tagalog, there is no passive. Al- though there is no extensive system of morphology that allows various non-terms to be coded as subject, the use of applicatives is extensive in Makasar Malay, allowing a range of semantic roles to participate in the voice system (as with Jakarta Malay).

Notice that there are many more forms in table 8 than have been seen in previous tables. This is a reflection of the fact that there are two possible coding choices for active clauses, and for inverse clauses involving local Agents. There are proclitics for all persons in active constructions, but only for local person in the inverse (see (29) for details).

(81) Dia me-liat laki.
3SG ACT-see man
'He/She CONTRASTIVE,FOCUS saw the man.'

(82) De-liat laki.
3SG.ACT,NOM-see man
'He/She GIVEN,KNOWN saw the man.'

(83) Laki di-liat-nya.
man NONACT-see-3SG,GEN
'He/She GIVEN,BACKGROUNDED saw the man.'

The constraints that determine the operation of the voice system in Makasar Malay are shown in (84).

Primary constraint rankings for Makasar Malay

(84) *OBL/Agt > *SUBJ/X > *OBL/X > *SUBJ/Pat

As with the Tagalog explication seen in 5.1, the high-ranked constraint *OBL/Agt bans passive constructions. This is the salient constraint behind the organisation of the Makasar voice system. The constraints against agentive subjects or prominent obliques, which were important in the more western languages examined so far, are not present in Makasar Malay, where the (cross-linguistically more 'normal') constraint *SUBJ/Pat appears, favouring active constructions over inverse or passive ones. The presence of a *OBL/X constraint drives the operation of applicative constructions, though not as strongly as the *OBL/X constraint proposed for Tagalog or Standard Indonesian, and certainly not as strongly as the combination of the two constraints in Jakarta Malay.
Additional low-ranking violations for the winning candidate: *Obl/Agt

<table>
<thead>
<tr>
<th>V(Agt/Lx, Pat/3/X)</th>
<th>*Obl/Agt</th>
<th>*SUBJ/x</th>
<th>*Obl/x</th>
<th>*SUBJ/Pat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td></td>
<td>!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td></td>
<td>!</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

(86) Ku-liat laki.
1SG.ACT: NOM-see man
'I saw the man.'

(87) Laki di-liat-ku.
man NONACT-see-1SG.GEN
'I saw the man.'

As mentioned above, the difference between the two coding choices shown in (86) and (87) depends on the status of the Agent in discourse. This provides further support for the idea that we need to allow for more than simply two levels of prominence or topicality (here X 'highly topical' and x 'less topical') in the analysis, as Sells (2001) proposes.

5.5 Ambon Malay

In Ambon Malay we see a system with no inverse option, only an active-passive distinction, a pattern that was seen in Kelantan Malay in 4.1 and that continues in Malays spoken further east. Correlating with the lack of an inverse system is the absence of morphological coding for the passive, a separate auxiliary being used instead.

Table 9. Coding options in Ambon Malay

<table>
<thead>
<tr>
<th>Agent \ Patient</th>
<th>Local</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i) Beta lia ose</td>
<td>(iii) Beta lia dia</td>
</tr>
<tr>
<td></td>
<td>(ii) Ose dapa lia dari beta</td>
<td>(iv) De dapa lia dari beta</td>
</tr>
<tr>
<td></td>
<td>(v) De lia beta</td>
<td>(vii) De lia de</td>
</tr>
<tr>
<td></td>
<td>(vi) Beta dapa lia dari dia</td>
<td>(viii) De dapa lia dari dia</td>
</tr>
</tbody>
</table>

(88) *Obl/Agt > *SUBJ/x > *SUBJ/Pat > ... > *GF/Pers

The constraint rankings for Ambon Malay are shown in (85), and they basically correspond to the description of English passives in Alissen (1999), the only addition being the overt presence of *Obl/Agt as a highly ranked constraint. This has been included for the purposes of comparison with the other Malay varieties discussed here, which do not feature this constraint highly. The constraint *Obl/Agt bans the inverse constructions that feature in the western languages.

Tableau for two different passive coding choices are shown in (89) and (91). While the outcomes are identical for identical reasons, they are useful for comparison with the Serui Malay data in the following section.

Additional low-ranking violations for the winning candidate:
*Obl/Agt, *Obl/L/Agt

<table>
<thead>
<tr>
<th>V(Agt/Lx, Pat/3/X)</th>
<th>*Obl/Agt</th>
<th>*SUBJ/x</th>
<th>*SUBJ/Pat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td></td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td>!</td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td></td>
<td>!</td>
<td>*</td>
</tr>
</tbody>
</table>

(90) De dapa lia dari beta.
3SG get see from 1SG
'He/She was seen by me.'

(91) Additional low-ranking violations for the winning candidate:
*Obl/Agt

<table>
<thead>
<tr>
<th>V(Agt/3/X, Pat/L/X)</th>
<th>*Obl/Agt</th>
<th>*SUBJ/x</th>
<th>*SUBJ/Pat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/SUBJ &gt; P/OBJ</td>
<td></td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>Inverse A/OBJ &gt; P/SUBJ</td>
<td></td>
<td></td>
<td>!</td>
</tr>
<tr>
<td>Passive A/OBL &gt; P/SUBJ</td>
<td></td>
<td>!</td>
<td>*</td>
</tr>
</tbody>
</table>

(92) Beta dapa lia dari dia.
1SG get see from 3SG
'I was seen by her/him.'

As with English an active clause will always be selected if the Agent is prominent, or if the Patient is non-prominent. Nevertheless, passives are grammatical with all combinations of person and number.
5.6 Serui Malay

The Serui Malay system uses the same dapa auxiliary that is found in Ambon Malay, and differs formally only in that the oblique Agent is marked with the comitative preposition deng, not the ablative dari. Furthermore, there is a high-ranking constraint on the coding of bivalent clauses in Serui Malay that forbids Agents being coded as oblique arguments, and so works against the appearance of forms such as 'They were seen by me', while allowing 'I was seen by them.'

Table 10. Coding options in Serui Mala

<table>
<thead>
<tr>
<th>Agent \ Patient</th>
<th>Local</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>(i) Kita lia ko</td>
<td>(ii) Kita lia dia</td>
</tr>
<tr>
<td>third</td>
<td>(iii) De lia kita</td>
<td>(v) De lia dia</td>
</tr>
<tr>
<td></td>
<td>(iv) Kita dapa lia deng dia</td>
<td>(vi) De dapa lia deng dia</td>
</tr>
</tbody>
</table>

The constraints that model the voice system in Serui Malay are the same as those for Ambon Malay, with the addition of a constraint forbidding local persons from being coded as obliques. This affects the ban on passive constructions with first or second person agents, the defining characteristic of Serui Malay and one that it shares with Standard Indonesian. The difference between the two systems is that Standard Indonesian has an additional non-active coding choice, the inverse, that is not available in Serui Malay. The constraints are shown in (93).

Primary constraint rankings for Serui Malay

(93) *Obi/Agt » *obi/» *SUBI/x » *SubI/» *SubI/Pat » » *GF/Pers

Serui Malay, as modelled here, works the same as Lushootseed in Aissen's analysis. There is a voice alternation, but the constraints that govern the voice alternation are outranked by the need to keep local persons as arguments of the verb. The tableau in (94) show a highly prominent third person Patient in a clause with a local Agent, which must be coded as active regardless of this prominence relationship. The passive clause shown in (96) is ungrammatical. (97) shows that with a third person Agent and otherwise the same conditions a passive will be selected.

(94) Additional low-ranking violations for the winning candidate:

<table>
<thead>
<tr>
<th>V(Agt/»X, Pat/3/X)</th>
<th>*Obi/Agt</th>
<th>*obi/»Agt</th>
<th>*SUBI/x</th>
<th>*SubI/»</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/Subi &gt; p/Obl</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse A/Obl &gt; p/Subi</td>
<td>*1</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Passive A/Obl &gt; p/Subi</td>
<td>*1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(95) Kita lia dia. 1sg see 3sg
'I saw her/him.'

(96) *dia dapa lia deng kita 3sg get see with 1sg
'He/She was seen by me.'

(97) Additional low-ranking violations for the winning candidate:

<table>
<thead>
<tr>
<th>V(Agt/3/X, Pat/3/X)</th>
<th>*Obi/Agt</th>
<th>*obi/»Agt</th>
<th>*SUBI/x</th>
<th>*SubI/»</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active A/Subi &gt; p/Obl</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse A/Obl &gt; p/Subi</td>
<td>*1</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Passive A/Obl &gt; p/Subi</td>
<td>*1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(98) De dapa lia deng dia. 3sg get see with 3sg
'He/She was seen by her/him.'

Apart from the addition of the constraint against passives with local Agents, the same remarks apply to Serui Malay as were given for Ambon Malay.

5.7 Papuan Malay

The voice system of Papuan Malay is the simplest of any described here, and indeed is empirically the simplest possible system that can be found in any language. This is because there are are no choices in the coding of arguments with respect to voice selection. All Papuan Malay clauses are active, there being no passive or inverse coding options. This then means that we have the simplest array of clauses to account for, shown in table 11.
Table 11. Coding options in Papuan Malay

<table>
<thead>
<tr>
<th>Agent \ Patient</th>
<th>Local</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>(i) Kita lia ko</td>
<td>(ii) Kita lia dia</td>
</tr>
<tr>
<td>third</td>
<td>(iii) De lia kita</td>
<td>(iv) De lia dia</td>
</tr>
</tbody>
</table>

The relevant constraints that derive this invariant active selection are shown in (99). The constraint *Obj/Agt works against the selection of inverse constructions, and in combination with *SUBJ/Pat, which bans passive constructions, these constraints rule out any voice alternation, regardless of the relative prominence of the two participants.

Primary constraint rankings for Papua Malay

(99)  
*Obj/Agt » *SUBJ/Pat » ... » [*GF/Pers, *GF/Patient]

The following tableaux shows a local agent and a non-local Patient in the same clause. While prominence has been marked, it is in fact irrelevant for the analysis.

(100) Additional low-ranking violations for the winning candidate:

*SUBJ/Agt, *SUBJ/Pat

<table>
<thead>
<tr>
<th>V(Agent, Patient)</th>
<th>*Obj/Agt</th>
<th>*SUBJ/Pat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Agent &gt; Patient</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Inverse Agent &gt; Object</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Passive Agent &gt; Object</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

(101) Kita lia dia.
1SG see 3SG
'I saw her/him.'

Papua Malay has no voice construction, and so no voice alternations. It serves simply as an example of the most extreme adaptation possible away from the three-voice systems of western Indonesia.

6 Change through Time and Space

Table 12 shows the different constraints presented for the seven languages arranged as closely as possible according to their cross-linguistic prominence. The table is arranged to have the more conservative (and, geographi-
cally, the more northern/western) languages at the top, and the more innovative/eastern languages towards the bottom (compare with the map included at the end of this article). In this way it serves as a guide to the diachronic development of voice systems in Indonesian/Malay varieties.

Table 12. Comparing constraints and rankings

<table>
<thead>
<tr>
<th></th>
<th>*Obj/</th>
<th>*SUBJ/</th>
<th>*Obj/X</th>
<th>*SUB/J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ph</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SeM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ph: Philippines; SI: Standard Indonesian (Standard Malay); JM: Jakarta Malay; MM: Makasar Malay; AM: Ambon Malay; SeM: Serui Malay; PM: Papuan Malay. Only the most highly-ranked constraints in each language have been shown.

A number of trends are apparent. We can see that the less conservative languages show a degree of 'normalisation', in cross-linguistic terms, towards the presence of a high-ranking *Obj/Agt constraint, banning the inverse construction. This corresponds almost perfectly with the lowering in rank of the *SUBJ/Agt constraint which is present as a highly-ranked constraint in the western, and northern, Austronesian languages.

There is a strong correlation of the presence of a high-ranking *Obj/X constraint with the presence of applicatives (or multiple voices, such as are found in Philippine-type languages; the non-Agent, non-Patient voices have been termed by some 'superapplicatives').
The constraint *SUBJ/x is universal; it is apparently absent from the high rankings in Papuan Malay, but this is only because it is redundant in a grammar without any voice alternations.

We can group the sampled varieties in terms of the constraints that dominate their voice systems. This is shown in figure 1, where the different languages are shown in an approximation to their geographic positions, and the lines show the extent of the individual constraints (again, compare with the map at the end). The shading in the bottom right (= south-east) of the figure indicates the decline of voice alternations in Malayic languages. The Philippine-type system, that does not allow a passive construction through the high-ranking *OBL/Agt constraint, is found in the south only in Makasar Malay, in Sulawesi, while the *SUBJ/Agt condition is found as a high-ranking constraint in all the western varieties.

![Figure 1. Varieties grouped by prominent constraints](image)

The connection between Standard Malay and Ambon Malay in terms of allowing active accidental forms is probably artefactual; as descriptions of local varieties of Malay improve and go beyond the lexical and morphological we are quite likely to attest more varieties in which there are active accidental constructions. We should also keep in mind the fact that the form of the constructions, the prefix ter- in Standard Malay and the independent verb dapa in Amboon Malay, are not cognate. Nonetheless, the theme of this paper being that we can successfully trace patterns in the development of

voice systems in closely related languages even in the absence of related morphology, the distribution of active accidental forms deserved further investigation.

Indonesian/Malay varieties in this area:

- Tagalog
- AM
- Ser.M
- Pap.M
- Std. Ind.
- Jkt. Mal.
- Mak. Mal.
- MP: Malayo-Polynesian; WMP: Western Malayo-Polynesian;
- CEMP: Central-Eastern Malayo-Polynesian; CMP: Central Malayo-Polynesian;
- EMP: Eastern Malayo-Polynesian; SHWNG: South Halmahera-West New Guinea

![Figure 2. The Austronesian family tree, extra-Formosan (approximate)](image)
began to be an important language in the last 400-500 years. While Ambon Malay, the variety for which we have the best historical documentation, is the sole language of many hundreds of thousands of people in central Maluku, it was built on the linguistic competence of the speakers of a large number of Central Malayo-Polynesian languages (Malay and the western Indonesian languages are Western Malayo-Polynesian). Similarly Makasar Malay is spoken in an area traditionally dominated by Makasar and Bugis, Western Malayo-Polynesian languages of the South Sulawesi group, significantly distinct to Malay, and sharing many syntactic characteristics that have more in common with the languages of the Philippines than with those of western Indonesia. Senui Malay is spoken in an area dominated by speakers of Austronesian languages from the South Halmahera-West New Guinea subgroup, and Papuan Malay in an area where the majority of the population did not, traditionally, speak an Austronesian language at all, or else an Oceanic one. The relationship of these different branches of Austronesian is shown in figure 2.

Given this genetic tree we can draw some obvious correlations with the constraints that determine the voice system of the languages. The Central Malayo-Polynesian language area shows a lack of inverse systems, except for some evidence in the far south-west of the region, while the Austronesian languages of Yapen island, where Senui is found, show complex systems of diathesis that clearly evolved as a result of hierarchically-based constraints on the operation of an earlier voice system. The only variety for which Austronesian substratal influence is not useful explanatory is Papuan Malay, from the modern Jayapura area, and this is because the major substrate is not Austronesian. Rather, the varied Papuan languages of the area are likely to have passed on most of their own morphosyntactic traits to the Malay spoken in the region today, including the lack of a voice opposition.

7 The Morphological Realisation of Voice Choices

The discussion above has centred on the realisation of coding options in the clause, through the selection of a particular voice from the inventory allowed by that language, as dictated by the relevant constraints and their rankings. In this section I shall sketch an account of the conditions constraining the morphological realisation of voice, in terms of two independent parameters, case marking on the DP (the by-phrase) and verbal marking. The challenge appears to lie in the by-phrase case marking, since the Indonesian/Malay languages typically do not mark case (see examples (1) and (2) for examples); some varieties do have case marking options, such as the optional appearance of akang as an accusative marker in Ambon Malay, and the complex system of Sri Lankan Malay, modelled on the Tamil case categories. In fact, the verbal morphology turns out to be the more complicated system to model, after determining the relevant constraints and their ranking. I shall adopt the constraints proposed by Aissen (1999), and the account that follows is very much in the spirit of that work.

In terms of case marking, we can dictate the obligatory marking of the by-phrase in terms of a constraints calling for case marking on oblique arguments; the lack of case marking in Malayic languages is true only of the core arguments, and non-core participants are regularly marked with prepositional or serial verb codings.

The relevant constraints proposed here are:

*STRUC Avoid morphological or syntactic structure.

*∅VM Avoid zero-coding of voice.

*∅C Avoid coding (syntactic) case without morphology

To illustrate the function of these constraints in a typology of voice marking, we can take the case of English DP marking options and the passive by-phrase. While an Agent and a Patient are unmarked for morphological case (if we exclude pronouns from the discussion for the present), an oblique Agent is overtly marked with the preposition by. We can assume that *STRUC dominates any constraints calling for case on subjects or objects, as shown in (102).

Do not mark arguments

(102)  *STRUC > *∅C & Subj, *∅C & Obj

The union of the two constraints, avoiding unmarked case and avoiding the coding of agents as obliques (see section 5), outranks *STRUC, which would call for there to be no case marking. *STRUC does, as seen in (102), outrank any constraints that dictate case marking on core elements. This applies to most modern varieties of Indonesian/Malay (certainly all of those discussed here), as well as English.
Mark by-phrases:  
(103) *∅C & Obi/Agt* *Struc

The different prepositions used to introduce oblique Agents have already been listed in Table 2, in Section 4. Figure 3 shows the geographic distribution of these different prepositions in the languages discussed in this paper. The figure arranges the languages according to their relative locations, though distance is not accurately portrayed. In terms of semantic extensions of the by-phrase marking prepositions there are four areas. In the west we find the locative-marking area, to which Ulu Muar Malay, Kelantan Malay and Perak Malay all belong. These languages all employ the same morphological device for locations as for passive agents; these are the varieties native to the Malayana peninsular, and the forms they use are cognate, reflecting developments of the pan-Malaya locative preposition *dl*. Jakarta Malay and the widely separated Serui Malay both use comitative prepositions to mark oblique Agents, but we cannot posit any common development here, as the forms used are different, *ame* (related to Indonesian/Malay *ama*) in Jakarta and *deng* (related to Indonesian/Malay *dengan*) in Serui. The fact that comitative and instrumental marking is commonly used, cross-linguistically, to mark passive agents as well as their primary functions (e.g., Keenan 1985) means that we do not need to search for any connection between Jakarta Malay and Serui Malay, but may simply and plausibly posit parallel developments. Ambon Malay shows *dari*, the ablative preposition both in that language and in other varieties of Indonesian/Malay, and another common choice, cross-linguistically, for passive marking. Only in Standard Malay and Standard Indonesian is there a dedicated oblique Agent preposition, *oleh*, which has no other prepositional function in the language. The root *oleh* is also found in a small number of other lexical items (*oleh* ‘because of’, *oleh-oleh* ‘souvenir present’, *beroleh* ‘obtain, receive’, *b-oleh* ‘can, may’, *per-oleh* ‘give permission’, *peroleh-an* ‘result, achievement’), but appears to have originally been a literary invention, which has not spread into any of the non-standard varieties. Kupang Malay and Papuan Malay lack by-phrases, and the Philippine-type languages and Makasar Malay lack passives (and also by-phrases), and so do not play a prominent role in this figure.

---

16 As noted in 4.1, Kelantan Malay, the most northerly Malay language in Malaysia, also allows for the possibility of an allative-marked by-phrase, or a by-phrase marked with the general oblique gerundation associated with obliques in general. The conditions under which one or the other of these options appear are not known.

---

For the realisation of verbal morphology to mark the voicing choices found in a language we need to consider not just the grammatical functions, as in (102), but the syntactic roles present in those grammatical function, as in (103). The verbal morphology system of English, and of the overwhelmingly large number of languages with active-passive systems, can be modelled with a set of constraints similar to those shown in (104).

**Constraints leading to marked passives (or inverses)**

(104) *∅VM & ∅Subj/Patient* *Struc*

Through these constraints voice-marking morphology will not be realised on a predicate except when the subject is a Patient, in which case there is overt marking. In English the passive involves an auxiliary verb, plus a particular verb form. We have seen that the passive constructions in Kelantan Malay, Ambon Malay, and others match this description very closely.

A radical divergence from this set of rankings, but one that is easily modelled, can be seen in the set of constraints that we can posit as operating for languages of the Philippine type, such as Tagalog as seen here in 3.1. The constraint against unrealised voice morphology outranks the ban on extra structural material in all cases, leading to a voice system that marks all voice choices, regardless of the syntactic or semantic role of the subject.

**Philippine-type**

(105) *∅VM* *Struc*

Another radical divergence from the standard voice model is that found in languages that have no overt voice morphology, either nominal or verbal,
but do show a voice alternation. While rare, they are found in a part of the Indonesian area, being found in languages such as Pali (Donohue 2005), from Flores in southern Indonesia. Outside Indonesia this voice system is attested in languages such as Lango (Noonan 1992), though it is rare. These grammars are easily modelled by assuming that *STRUC outranks any constraints that would call for verbal morphology. This results in a system in which there is no morphological marking of either an active or a passive voice.\(^{17}\)

**Flores-type**

(106) \[ *\text{STRUC} \succ *\text{VM} \]

Turning to a more complicated example, the morphology of standard Indonesian overtly marks voice choices except in the case of a non-active clause with a local agent. This is a highly specific constraint, but one that can be easily modelled, if we take into account the local conjunction of \[ *\text{VM} \] and a specification of an oblique local agent (*OBJ/L/AGT), as well as a general principle of avoiding unmarked voice choices (*\text{VM})*. This is shown in (107), and has the effect of morphologically marking both active and non-active voice types, except in the case of an inverse construction with a local agent.

**Indonesian-type:**

(107) \[ *\text{VM} \succ *\text{STRUC} \succ *\text{VM} \& *\text{OBJ/L/AGT} \]

We can logically hypothesise the existence of another language type, one not attested in this sample (and, as far as I am aware, not definitively attested anywhere in the world), in which voice is marked only in instances where the agent is coded as a subject. This would appear to be a model of an antipassive, rather than a passive, except that there are no attested instances of languages in which a thorough-going Patient (as opposed to Agent) orientation is attested.\(^{18}\)

\(^{17}\) While this lack of morphological form for a diathesis might seem highly unusual, the parallels between this type of voice system and a more familiar one such as English or the Philippine languages shown in (105) matches that between the morphologically unmarked dative shift in English with the more common overtly marked applicative constructions found around the world.

\(^{18}\) Partly this is terminological, since 'Patient' is being used here to refer to the role played by the most patient-like argument of a bivalent verb, and all languages accord special privileges to the single argument of a monovalent verb, if any. In part, however, it reflects the fact that in those languages which do show constructions with an absolutive pivot there are inevitably other constructions with a differently-grouped (typically nominative-accusative) pivot. This begs the question of the identification of 'subject' in a mixed-pivot language, a non-trivial task.

\begin{center}
\begin{tikzpicture}
\node (KeM) at (0,0) {KeM};
\node (Phl) at (1,0) {Phl};
\node (PM) at (2,0) {PM};
\node (SeM) at (1,-1) {SeM};
\node (MM) at (0,-1) {MM};
\node (UMM) at (-1,-1) {UMM};
\node (RI) at (-1,-2) {RI};
\node (SI) at (0,-2) {SI};
\node (AM) at (1,-2) {AM};
\node (JM) at (2,-2) {JM};
\node (RM) at (3,-2) {RM};
\node (KM) at (4,-2) {KM};
\draw [->] (KeM) -- (SeM);
\draw [->] (SeM) -- (PM);
\draw [->] (PM) -- (Phl);
\draw [->] (Phl) -- (MM);
\draw [->] (MM) -- (UMM);
\draw [->] (UMM) -- (RI);
\draw [->] (RI) -- (SI);
\draw [->] (SI) -- (AM);
\draw [->] (AM) -- (JM);
\draw [->] (JM) -- (RM);
\draw [->] (RM) -- (KM);
\end{tikzpicture}
\end{center}

**Figure 4. Varieties grouped by morphological realisation of voice**

The difference between morphological passives and analytical passives is simply accounted for by separating *STRUC into *STRUC\textunderscore Syntax and *STRUC\textunderscore Morphology, with the relevant constraint dominant in each language. The distribution of morphological and analytical voice marking is shown in figure 4.

Here three areas can be distinguished. Those languages with morphological marking for non-active and, in most cases, active voice as well are shown in the shaded area, while languages with analytical voice systems for the passive are shown unshaded. The band on the top and right encompassing Philippine languages and Papuan Malay indicates the range in which the accidental prefix *ter- (realised variously as tar-, ta- and ta-) see table 1) is not found productively, though it may appear on a few isolated lexical items, frozen as part of the root. Note that the area in which there is no means of expressing by-phrases (see figure 3) includes Papuan Malay, as would be expected, since it has no productive voice system, but also Kupang Malay, in which there is an accidental prefix ta-. This means that the accidental ta- in Kupang is only ever used without an expressed Agent, and indeed the cognitive ta- in Ambon Malay and Sorei Malay is almost never accompanied by an Agent, the morpheme apparently being interpreted more as a marker of stativity in the more eastern Malay varieties, including Makassar Malay.
8 Some Remaining Issues

I have so far uncritically discussed these different 'varieties' of Indonesian/Malay without assessing their status with respect to each other. Leaving aside the immensely complicated sociolinguistic issues, we need to ask if they are really the same language. In terms of voice typology, it is clear that the answer is no. Yet, as mentioned earlier, lexically the different varieties are extremely similar, showing 85+% cognacy in basic vocabulary, and in general very regular and very predictable sound changes. We also know that the time period in which they have diverged is very short. This raises the question of how much grammatical divergence is possible in the one language, and how much linguists assume about the speed of grammatical convergence or divergence is based on unusually conservative cases.

The analysis has assumed, and confirmed, the following hierarchies:

1. **Subj/X > OBL/X > OBL/X**  
   More highly prominent material should be coded in a grammatically more salient position; conversely, grammatically salient positions, such as subject and object, should not be filled with unimportant information.

2. **OBL/L/Agt > OBL/L > Subj/L**  
   Local participants should not be coded in grammatically unimportant positions, but should remain central to the developing discourse.

This hierarchy can better be described in terms of the inherent prominence of pronominal arguments, and particularly local pronominal arguments. Since there is a degree of inherent prominence associate with participants, roughly equal to their position on the animacy hierarchy, this will work to ensure that local persons are not normally coded in oblique positions.

3. **Subj/Pat > (?) OBL/Pat > OBL/Pat**

4. **OBL/Agt > OBL/Agt > Subj/Agt**  
   Somewhat unexpectedly the preferential grammatical positions for coding Agents and Patients do not follow the Subj > OBL > OBL hierarchy that we have seen earlier, but rather place subject and object at opposite ends of the continuum. The extreme rarity, cross-linguistically, of object Agents means that the ranking of the Function/Agent hierarchy is definitely justified.

The rankings for Patient are more controversial, and hinge on the interpretation of conative and conative-like alternations in languages. Here the alternation in the coding of arguments such as *the boys in She hit the boys and She hit at the boys apparently shows an object-oblique relationship, but in many languages, including English, the argument can be shown to be a morphological oblique, but nonetheless remains syntactically an object.

I would like to finish with some remarks on further developments in voice morphology that have not fallen under the scope of a discussion of voice constructions. This can be shown by examining data from Riau Indonesian (Gil 2002), which shows a development of the voice morphology (N- and di-, related to Standard Indonesian *meng- and di-) into 'generalised agreement' markers, apparently indexing argument structure positions regardless of their grammatical function. Gil cites the following example; other non-standard varieties of Malay show similar patterns.20 Clearly we cannot simply assume that N- marks and active voice and that *di- the nonactive. To judge from Gil's discussion of this and similar examples such marking only appears when both the Agent and the Patient are prominent in discourse; the prefixes appear to function as a kind of pronominal agreement pattern.

**Riau Indonesian**

(109)  

<table>
<thead>
<tr>
<th>Baju-nya</th>
<th>di-m-injam.</th>
</tr>
</thead>
<tbody>
<tr>
<td>baju-nya</td>
<td>di-N-pinjam</td>
</tr>
</tbody>
</table>

garment-ASSOC AGR.Pat-AGR.Agt-borrow

'She borrowed his clothes.'

---

19 Keep in mind the fact that here 'Agent' is not being used to refer to the semantic role 'agent', including co-agents in actions, but to the most agent-like argument of a primary transitive verb, and other most agent-like arguments of two or three place verbs that show similar behaviour. We can easily construct sentences with objects that must be construed as agents in the semantic role sense, that are nonetheless Patients in the syntactic role sense. An example is *She helped us to make the tortillas.*

20 With a restricted set of lexical items this applies to Standard Indonesian as well. Words which can only be used verbally with *meng-* retain this prefix even when marked for nonactive voice; examples include *di-[meng-erit] 'understood', di-[mu-kan] 'eaten'. While the fusion of *mu-onto kan must be ancient, since there is no independent form *kan (< proto-Austronesian *kaan), erit is still found with other affixes (berarit 'that means', pengerait 'understanding'), and so the *meng-* on it must be considered at least partly productive (though see Donohue 2004b for a discussion of some of the problems we face with the term 'productivity' synchronically and diachronically).
This pronominal analysis ties in with what Sells (2000) has hypothesised about the Austronesian voice system, and developments seen elsewhere in other less conservative Austronesian languages, such as Tukan Besi and languages further east (Donohue 2004a). Alternatively, the use of *di-* here might be part of what has been observed across Indonesian in the use of *di-* as a marker of formality, leaving only the *N-* as a pronominal element (though I consider this an unlikely analysis). The development of *di-* as a marker of formality presumably originated in speakers of Indonesian/Malay varieties that do not employ *di-* as a voice marker, but who were aware of the appearance of *di-* (as well as *meng-*) in formal, standard uses of Indonesian (thus revealing the sociolinguistic attitudes that treat the various varieties discussed here as belonging to the same 'language' in terms of speaker identification). In some formal registers *di-* can now be heard with verbs that cannot easily be construed as non-active, though it is hard to classify the language to which these registers belongs, since it depends on geographic, sociolinguistic, and situational variables. Examples of the use include the following sentences.

(110) (on the phone, after asking to speak to a particular person at an office)

Bapak **di-nanggu ya** ...
sir **Di-wait TAG**

‘Just wait a moment please sir . . .’

(111) (purchasing cassettes in a shop)

**Dua ini ya pak mau di-beli?**
two this AG sir want Di-buy

‘So it’s these two that you want to buy, is it?’

These examples, while inconclusive, might show the beginnings of the complete loss of voice constructional meaning associated with the *di-* prefix, and its transformation to a marker of register. To what extent this development will affect or be affected by existing uses of *di-* as a voice marker in different varieties of Indonesian/Malay is not yet known, but indicates that synchronic studies too have a significant role in eliciting the diversity that we might have otherwise ascribed to historical change or influence.

Appendix: Map

The Malay and Indonesian varieties referred to:

The varieties discussed in particular detail have been shown on the map with circles, and are shown in bold in the list below.

1. Perak Malay  
2. Kelantan Malay  
3. Standard Malay  
4. Ulu Muar Malay  
5. Baba Malay  
6. Riau Malay/Indonesian  
7. Jakarta Malay  
8. Standard Indonesian  
9. Makasar Malay  
10. Kupang Malay  
11. Ambon Malay  
12. North Maluku Malay  
13. Serui Malay  
14. Papua Malay  
15. Merua Malay  
16. Pattani Malay  
17. Cocos Malay  
18. Laranuka Malay  
19. Dobel Malay  
20. Sarawak Malay  
21. Brunei Malay  
22. Sri Lanka Malay

References


