ROUTLEDGE LANGUAGE FAMILY SERIES

Each volume provides a detailed, reliable account of every member language, or representative languages of a particular family. Each account is a reliable source of data, arranged according to the natural system of classification: phonology, morphology, syntax, lexis, semantics, dialectology and socio-linguistics. Each volume is designed to be the essential source of reference for a particular linguistic community, as well as for linguists working on typology and syntax.

The Austronesian Languages of Asia and Madagascar
Edited by Alexander Adelaar & Nikolaus P. Himmelmann

The Bantu Languages
Edited by Derek Nurse & Gérard Philippson

The Celtic Languages
Edited by Martin Ball & James Fife

The Dravidian Languages
Edited by Sanford B. Steever

The Germanic Languages
Edited by Johan van der Auwera & Ekkehard König

The Indo-Aryan Languages
Edited by George Cardona & Dhanesh Jain

The Indo-European Languages
Edited by Paolo Ramat & Anna Giacalone

The Iranian Languages
Edited by Gernot Windfuhr

The Khoesan Languages
Edited by Rainer Vossen

The Manchu-Tungusic Languages
Edited by Alexander Vovin

The Mongolic Languages
Edited by Juha Janhunen

The Oceanic Languages
Edited by John Lynch, Malcolm Ross & Terry Crowley

The Romance Languages
Edited by Martin Harris & Nigel Vincent

The Semitic Languages
Edited by Robert Hetzron

The Sino-Tibetan Languages
Edited by Graham Thurgood & Randy LaPolla

The Slavonic Languages
Edited by Bernard Comrie & Greville B. Corbett

The Turkic Languages
Edited by Lars Johanson & Eva Csato

The Uralic Languages
Edited by Daniel Abondolo

THE AUSTRONESIAN LANGUAGES OF ASIA AND MADAGASCAR

Edited by Alexander Adelaar and Nikolaus P. Himmelmann

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2005
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group

Cartography by Chandra Jayasuriya, The University of Melbourne

Typeset in Times New Roman by
Newgen Imaging Systems (P) Ltd, Chennai, India
Printed and bound in Great Britain by
MPG Books Ltd, Bodmin

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN 0–7007–1286–0

CONTENTS

CHAPTER FIVE

THE AUSTRONESIAN LANGUAGES OF ASIA AND MADAGASCAR: TYPOLOGICAL CHARACTERISTICS

Nikolaus P. Himmelmann

1 SOME PRELIMINARY DIVISIONS AND DEFINITIONS

When discussing typological characteristics of a genetically coherent group of languages, there are two points of view, an external and an internal one. Externally, the focus is on features which characterize the group as a whole vis-à-vis other language families or linguistic areas. Internally, the focus is on features which characterize one (typological) subgroup as opposed to another. The two points of view are obviously interrelated in that a complex internal typology – i.e. a group of languages differs with regard to a large number of features – usually makes it difficult, if not impossible, to reach significant external generalizations.

The Austronesian languages of Asia and Madagascar provide a paradigmatic example of this difficulty. They are typologically much more variegated than the Oceanic languages (and many other language groups and families). There are only very few features which are sufficiently general and widespread to be considered typological characteristics of the group as a whole, including reduplication (see section 2.2), the distinction between inclusive and exclusive pronouns (see section 3.7) and morphological causatives (see section 4.3).

The complex internal typology of the Austronesian languages of Asia and Madagascar entails a number of expository problems and inconveniences. Most importantly, for almost every generalization one has to give at least a rough indication of the set of languages to which it applies. To this end, occasional reference is made to the genetic groupings discussed in ADELAAR (A HISTORICAL PERSPECTIVE). But two other ways of specifying the relevant set are used much more frequently. First, the relevant set is often specified in geographic terms, referring either to national territories (e.g. *Philippine languages*) or major islands or island groups (e.g. *the languages of Borneo, the Moluccas*, etc.). See section 1.1 for further comments. Second, major use is made of a somewhat rough but still useful distinction between two basic types of languages, i.e. symmetrical voice languages and preposed possessor languages, as discussed further in section 1.2.

Finally, in order to keep the amount of hedging and repetition to an absolute minimum, it will be useful to identify right at the outset the most important exception to almost any

typological generalization regarding the Austronesian languages of Asia and Madagascar. The Chamic and Moken-Moklen languages spoken in Vietnam, Cambodia, Myanmar (Burma) and Thailand have long been in intensive contact with non-Austronesian Southeast Asian mainland languages and have taken on their basic structural characteristics. These include regular tonal distinctions, fairly rigid SVO order, lack of affixation, etc, and are discussed further in the chapters on Cham and Moken-Moklen (see also Thurgood 1999 for the most thorough and up-to-date exposition of the contact history in this area). Unless explicitly mentioned, these languages are excluded from generalizations or divisions discussed in this chapter.

1.1 Geographical groupings

It has proved a rather difficult task to find a suitable label for the languages under investigation in this volume. There is a well-established term covering all but two Austronesian languages outside Asia and Madagascar, i.e. *Oceanic Languages*. These are the Austronesian languages of the Pacific with the exclusion of Palauan (Belau) and Chamorro, which geographically belong to Micronesia. However, no well-established term for non-Oceanic Austronesian languages exists, and since *Austronesian languages of Asia and Madagascar (plus Palauan and Chamorro)* is a rather clumsy way to refer to them, the rather loose geographical expression *western Austronesian* will be used here instead. A possible problem with this expression is that it has been used in the literature in various, and often imprecise, ways. In the present chapter, *western Austronesian languages* is strictly equivalent to *non-Oceanic Austronesian languages*. It is not to be confused with the term *Western Malayo-Polynesian* which is used in genetic classifications (cf. ADELAAR, A HISTORICAL PERSPECTIVE) and which covers only some of the languages dealt with here.

Apart from *western Austronesian*, there are a few other geographical references which have been used in different ways in the literature and thus require brief comment.

- *Philippine languages*: Philippine languages are the Austronesian languages spoken in the Republic of the Philippines. The Sama varieties spoken in the Sulu Archipelago in the southern Philippines (see Map 13.1), although spoken within the Republic of the Philippines, differ morphosyntactically from other Philippine languages and are generally not included in generalizations about Philippine languages. The same would appear to be true of the South Mindanao languages (Blaan, Tboli, Tiruray and possibly Bagobo (or Giangan)) but very little is known about them.
- *Philippine-type languages*: This term, though widely used, has never been precisely defined. It is used to refer to the Philippine languages (as just defined) and languages from neighboring islands which share typical Philippine characteristics. Roughly, they also include the Formosan languages and the languages of northern Borneo and northern Sulawesi. Sometimes Malagasy, Chamorro and Palauan are included too. Obviously an accurate delimitation of Philippine-type languages depends on what is considered a typical Philippine characteristic. As discussed in the next section, a somewhat narrower definition of Philippine-type languages is used here which excludes Malagasy, Chamorro, Palauan and a few Formosan and northern Bornean languages.
- *Languages of western Indonesia*: This refers to all languages spoken on Indonesian territory west of Lombok. This includes Bali, Java, Kalimantan, Sumatra and the minor islands in between them, but excludes the languages of Sulawesi. Most statements regarding the languages of western Indonesia will be sufficiently qualified so as to exclude Acehnese and the Barrier Island languages.

- *Barrier Island languages* are the languages spoken on the island chain west of Sumatra. They include Nias, Enggano, Mentawai, Sichule, and Simeuluë. Morphosyntactically, these languages differ significantly from the languages spoken on Sumatra.
- *Lesser Sunda languages* refers to the languages spoken on the island chain extending from Lombok to Alor, including Sumba and Savu but excluding Wetar and languages further east which belong to the Southwest Moluccan languages. Most languages spoken on Pantar and Alor, the two easternmost islands of the chain, are non-Austronesian and hence excluded here from any generalization referring to Lesser Sunda languages. Although often considered Lesser Sunda languages, the languages spoken on Timor and Roti are also excluded here since they have a somewhat different typological profile (see also section 6).

1.2 Symmetrical voice vs. preposed possessor languages

Even if the Chamic languages and Moken-Moklen are excluded, the structural variety attested by the remaining western Austronesian languages is still considerable. In order to be able to propose any generalizations at all, it will be useful to make a distinction between two basic types of languages found in the area, with a third group of languages not clearly aligning with either type. The first type of languages will be called *symmetrical voice languages*. The defining characteristic of these languages is the presence of at least two voice alternations marked on the verb, neither of which is clearly the basic form. Compare the following Malay examples:

(1) *Anak saya me-lihat orang itu.*
child 1s AV-see person DIST
'My child saw that person.'

(2) *Orang itu di-lihat anak saya.*
person DIST PV-see child 1s
'My child saw that person.'

The voice alternation illustrated by these two examples is symmetrical in that both the actor voice form *melihat* and the undergoer voice form *dilihat* are morphologically overtly marked by a prefix (*me-* and *di-* respectively). Thus, morphologically speaking, there is no unmarked or basic form from which the other form is derived. Furthermore, both examples appear to be syntactically equivalent in that both involve two nominal arguments (*anak saya* and *orang itu*), one preceding and the other following the verb without further overt marking by a preposition or case marker. The only obvious syntactic difference between the two voices pertains to the fact that in actor voice the experiencer precedes the verb and the stimulus follows it, while in undergoer voice the order is reversed. In contrast, in an English passive construction such as *That person was seen by my child* the experiencer does not only occur in postverbal position but is also marked as oblique, hence rendering the overall construction intransitive.

Both the above examples and the notion of a *symmetrical voice alternation* require further discussion and elaboration which will be found below in sections 3.3, 3.6 and 4.2. Here it will be sufficient to note that the question to what extent the alternation illustrated by the two examples is in fact symmetrical continues to be a matter of controversy. But this controversy is only of marginal relevance to present purposes. It is widely agreed that many western Austronesian languages show voice alternations which differ significantly in form and function from the passive alternation in, for example, European languages

(the fact that example (2) is not given a passive translation is intended to reflect this difference). For much of the exposition of this chapter, it does not matter whether or not one agrees with a symmetrical voice analysis of these alternations. What does matter is the (uncontroversial) observation that these alternations are found in only some western Austronesian languages, thus providing a criterion for dividing these languages into two sets. Symmetrical voice languages in the sense intended here include the Austronesian languages of Taiwan, the Philippines, Malaysia, Madagascar, western Indonesia (with the exception of Acehnese and the Barrier Island languages) and the northern half of Sulawesi (Saluan (but not Banggai), Kaili-Pamona, Tomini-Tolitoli, Gorontalo-Mongondow, Minahasan and Sangiric).

The *Philippine-type languages* are a subset of the symmetrical voice languages. As mentioned above, the defining characteristics for this subset have never been explicitly defined (see also Himmelmann 2002a/b). Here the following preliminary definition is used. Philippine-type languages are symmetrical voice languages which have

(a) at least two formally and semantically different *undergoer* voices (see section 4.2.2);
(b) at least one non-local phrase marking clitic for nominal expressions (e.g. Tagalog genitive *ng*; see further section 3.5);
(c) pronominal second position clitics (see section 3.2).

This definition of Philippine-type languages excludes Malagasy, Chamorro, Palauan and the Austronesian languages of Brunei and Sarawak as well as Tomini-Tolitoli, Gorontalo-Mongondic, Sama-Bajau, and South Mindanao languages, all of which have occasionally been called Philippine-type languages. Among the Formosan languages, the Tsouic and Rukai languages are clearly not Philippine-type according to this definition, and the status of others (e.g. Bunun) needs further investigation. Similarly, the status of the Sangiric and Minahasan languages in northern Sulawesi needs further investigation since it is not clear to what extent the pronominal clitics in these languages are in fact second position clitics.

The advantage of this preliminary definition of Philippine-type languages is that the group of languages thus delimited would appear to be quite homogeneous and to share a number of further morphosyntactic features in addition to the three defining features (for example, infixes, aspect-mood marking, a strong preference for predicate-initial clause structure, etc.). The exact extent of their commonalities, however, needs further testing.

The second basic type of languages found in the western Austronesian area will be called *preposed possessor languages*. In these languages, possessors regularly precede the possessum as in the following Ambai examples: *Yani ne munu* (Yani POSS house) 'Yani's house', *ne-mu tarai* (POSS-2s.POSS body) 'your body'). Once again, this definition is in need of further elaboration and discussion which will be found in section 3.11. Here it is important to note that the criterion refers to the most common or unmarked order found in possessive constructions. That is, it is not required that all possessive constructions in a preposed possessor language show the order POSSESSOR-POSSESSUM, and conversely, non-preposed possessor languages may optionally allow a POSSESSOR-POSSESSUM order. Preposed possessor languages in this sense are the non-Oceanic Austronesian languages of Timor, the Moluccas and West Papua as well as the Pidgin-Derived Malay varieties (see ADELAAR, MALAYIC VARIETIES).

The two parameters *preposed possessor* and *symmetrical voice alternation* tend to correlate with other features (which makes them typologically useful). Most importantly, they tend to correlate negatively with each other in that languages with symmetrical voice alternations generally show postposed possessors, and languages with preposed

possessors either do not show any grammaticized voice alternations at all or the voice alternations are clearly asymmetrical. Importantly, the reverse implications do not hold: languages with postposed possessors are not necessarily symmetrical voice languages, and languages with asymmetrical voice alternations are not necessarily preposed possessor languages, as the transitional languages discussed below show in particular. For further correlations, see section 6.

The distinction between symmetrical voice and preposed possessor languages is not new in the sense that the occurrence of preposed possessors was noted as a major feature of Moluccan languages as early as Van Hoëvell (1877:15f). This feature has continued to play a major role in discussions of the subgrouping of the Austronesian languages of eastern Indonesia (see Collins 1983:27–29 and Grimes 2000:3–6 for succinct discussion and references). In this regard, it is important to note that preposed possessor order is used here as a parameter in a *typological* classification, not a genetic one (its value for genetic subgrouping is rather doubtful, as already noted by Jonker (1914)).

In the present typological classification it is possible that a language is neither a preposed possessor language nor a symmetrical voice language. Such languages are found in the southern half of Sulawesi (e.g. Bugis, Makassar, Banggai, Mori Bawah, Muna, Tukang Besi), in the Lesser Sunda islands (e.g. Keo, Kambera) and in northwest Sumatra (Barrier Island languages and Acehnese). Most of these languages do not have verbal voice morphology. If they do (e.g. Makassar, Mori Bawah), it involves asymmetric voice alternations. Possessors are generally postposed (a major exception is Banggai in eastern central Sulawesi). With regard to other parameters, these languages differ quite substantially (many have elaborate person marking systems, but others do not; only a few have infixes; etc.). That is, they do not adhere to a common typological core profile. For expository purposes, it will be convenient occasionally to refer to these languages as *transitional languages*, reflecting the fact that most of them (except the northwest Sumatran languages) are geographically located in between the symmetrical voice and the preposed possessor languages. However, it should be emphasized that this is *not* a typological category. It is thus of a very different nature than *symmetrical voice languages* and *preposed possessor languages*.

In terms of numbers of languages, roughly 60% of the ca. 800 western Austronesian languages are symmetrical voice languages and 25% are preposed possessor languages, while the remaining 15% do not clearly align with either type and hence are called transitional here.

Note that there is also the converse of a transitional languages as just defined, i.e. languages that are both symmetrical voice languages and preposed possessor languages. Such languages appear to be extremely rare, but there is at least one example, i.e. the Formosan language Pazeh. Apart from having preposed possessor order in the basic possessive construction, this language is very similar to its neighboring symmetrical voice languages and should therefore actually be considered a symmetrical voice language (see also section 3.11).

1.3 Sources and conventions

This typological overview partly draws on the language sketches found in the remainder of this book. The names of these languages are given in bold in the remainder of the present chapter. The sketch chapters will usually contain further information, bibliographic references and examples for the issue under discussion. With very few exceptions, this information is not repeated here.

To broaden the typological variety documented in the language sketch chapters, data from languages with somewhat different typological profiles are also included. Unless indicated otherwise, data on these languages are from the following sources: Atayal (Taiwan, Rau 1997), Pazeh (Taiwan, Li and Tsuchida 2001, Blust 1999), Rukai (Taiwan, Li 1973), Paiwan (Taiwan, Egli 1990), Bontok (northern Philippines, Reid 1970, 1992), Cebuano (central Philippines, Wolff 1966, 1972), Ratahan (northern Sulawesi, Himmelmann and Wolff 1999), Lauje, Totoli and other Tomini-Tolitoli languages (central Sulawesi, Himmelmann 2001 and field notes), Urak Lawoi' (Malayic variety/southern Thailand, Hogan 1999), Standard Indonesian (Malayic variety, Sneddon 1996, Musgrave 2001), Acehnese (northern Sumatra, Durie 1985), Madurese (Madura/Java, Davies 1999a), Sundanese (Java, Müller-Gotama 2001), Balinese (Bali, Clynes 1995, Arka 1998, Pastika 1999), Muna (southeastern Sulawesi, Van den Berg 1989), Tukang Besi (southeastern Sulawesi, Donohue 1999), Keo (Flores/Lesser Sunda islands, Baird 2002), Kedang (Lembata/Lesser Sunda islands, Samely 1991), Waima'a (East Timor, Hull 2002, author's field notes), Tetun Dili (East Timor, Hull and Eccles 2001, Williams-van Klinken, Hajek, and Nordlinger 2002), Selaru (southeastern Moluccas, Coward and Coward 2000), Dobel (Aru/southeastern Moluccas, Hughes 2000), Buru (central Moluccas, Grimes 1991), Alune (Seram/central Moluccas, Florey 2001), Ambai (West Papua, Silzer 1983).

Despite this relatively broad range of languages that have been consulted, it will soon become obvious that the present survey is somewhat biased towards symmetrical voice languages. One reason for this bias is the fact that the author is most familiar with these languages. There are also far fewer publications about preposed possessor languages although, fortunately, descriptive work on these languages has increased noticeably in recent times (see references above, Grimes (ed.) 2000, and also Hull 2001 for the languages of Timor).

Examples in the phonology section are sometimes written with IPA symbols rather than in the established practical or standard orthography. Otherwise, the practical or standard orthography is used. Note that in these orthographies glottal stop is often represented as <'> or as <q>, or it remains unrepresented. Stress is only indicated if it does not fall on the penultimate syllable.

The orthographic representation of clitics as orthographically 'free' or 'bound' forms also follows the standard orthography. Note in particular that clitics in Meso-Philippine languages, including **Tagalog**, are generally represented as orthographically independent words.

Examples without source references are from the author's own corpora. In examples from other sources, the glosses have been adapted to the conventions followed throughout this book.

2 PHONOLOGY AND MORPHONOLOGY

Viewed crosslinguistically, Austronesian languages tend to be fairly inconspicuous with regard to basic phonological features. Segment inventories typically consist of three to five vowels and 16–20 consonants, with up to four nasals matching corresponding voiced and voiceless stop series and only few fricatives. The most common syllable structures are (C)V and (C)VC, frequently with restrictions on the consonants allowed in coda position. If consonant-clusters are allowed syllable-internally, they are typically restricted to onset position and usually consist of either nasal plus obstruent or obstruent plus glide or liquid. Consonant clusters across syllable boundaries are also fairly restricted. Often only

clusters consisting of an obstruent preceded by a homorganic nasal (so-called NC-clusters) are allowed for. Stress is usually non-distinctive and occurs on the penultimate syllable. Lexical bases tend to be disyllabic and there is also a widespread tendency to reduce affixed words with more than two syllables to disyllabic words. In disyllabic lexical bases of the structure $C_1V(C)C_2V(C)$ there tend to be numerous co-occurrence restrictions on C_1 and C_2 and on the two vowels (this has been documented in detail for **Javanese**, Balinese, Malay and Muna, inter alia). It is very common to insert glides in between a high vowel and a following non-high vowel (hence /ia/ → [ija], /ua/ → [uwa], etc.), to palatalize dental or alveolar stops and fricatives before /i/, and to leave final stops unreleased.

Less commonly occurring consonants include implosives (in fact, these are fairly widespread in central and southeastern Sulawesi, northern Borneo and the Lesser Sunda islands, but also found in **Tsou**), interdental fricatives (e.g. some Formosan languages such as Thao and Rukai), retroflex stops (in, for example, **Javanese**, Madurese and Rukai) and uvular stops in addition to velar stops (in, for example, Formosan languages). The allophones of the alveolar lateral approximant which occurs in most segment inventories may include lateral fricatives, retroflex lateral flaps or palatal lateral approximants (e.g. in Formosan languages, **Buol**), a few languages actually having two lateral phonemes in addition to an alveolar trill or flap which is also part of the standard inventory (e.g. Rukai, Thao). **Nias** is unique among western Austronesian languages in having a bilabial trill but note that bilabial fricatives are widely attested as phonemes or allophones throughout the area. In the Lesser Sunda islands and the southern half of Sulawesi prenasalized consonant series frequently occur.

Some languages with larger consonant inventories have a three-way distinction between voiceless aspirated, voiceless unaspirated and voiced stops (e.g. Madurese, Urak Lawoi'). Waima'a, a language of East Timor, appears to have a significantly larger consonant inventory, with aspiration and glottalization providing for additional phonemic contrasts (Hull 2002, Hajek and Bowden 2002). Preglottalized consonants have also been reported for other languages, including Keo (spoken on Flores) and some Formosan languages (**Tsou**, Thao). **Biak** represents languages with somewhat smaller consonant inventories which appear to be more common in the southeastern Moluccas and West Papua (e.g. Selaru, Dobel, Ambai).

With regard to vowels, even closely related languages vary as to whether or not schwa is part of the basic inventory. In some languages, an unrounded back vowel [ɯ] or [ɤ] occurs, either in place of its rounded counterpart (e.g. Tukang Besi) or in addition to it (e.g., Sundanese, Madurese, Land Dayak varieties). Diphthongs are found in many descriptions but here one has to be careful to separate substantial phonetic or phonological differences from differences in descriptive practice. More often than not, what is described as a diphthong by one author is described as vowel plus glide sequence by another (e.g. /ay/ or /aw/). As Clynes (1997) convincingly argues (with regard to the root-final diphthongs of Proto-Austronesian, but the argument also applies to many synchronic descriptions), the vowel-plus-glide analysis is to be preferred on phonological grounds in most instances (possible exceptions include **Moken-Moklen** and Acehnese). Another widespread lack of agreement in descriptive practice (and occasionally also in actual fact) pertains to the phonological analysis of the glides themselves which sometimes are considered allophones of vowels and sometimes as consonant phonemes.

More complex vowel inventories are found in the Southeast Asian mainland languages (**Chamic**, **Moken-Moklen**), on Java, and in some parts of Sumatra and the Lesser Sunda islands. Acehnese, in fact, is said to have ten oral monophthongs plus seven nasal monophthongs and ten diphthongs (Durie 1985:16f, but see also Stokhof 1988:329–331). Kedang is described as distinguishing a neutral vowel series from one articulated with an advanced tongue root (Samely 1991:13–35).

Hardly any restrictions on the coda in CVC-syllables occur in Formosan and Philippine languages which also allow a wide range of consonant clusters across syllable boundaries. **Tsou**, in fact, is crosslinguistically remarkable for its broad range of consonant clusters. **Tsou** is also among the languages which allow a range of syllable-initial consonant clusters. Such languages are found in Taiwan (another example is Thao) and in Halmahera and the Moluccas (e.g. **Taba**, **Leti**). While Formosan and Philippine languages generally do not allow sequences of like consonants (or geminate consonants), such sequences are a prominent feature of South Sulawesi languages (**Makassar**, Bugis, etc.). More sporadically and mostly as a result of morphonological processes, they are also found elsewhere in, for example, **Leti**, Dobel, **Taba**, Tukang Besi, Totoli, Bontok, Madurese, Toba Batak and various **Malayic varieties**; see also Blust 1995).

A fair number of languages, including Dobel and many Philippine languages (e.g. **Iloko**, **Tagalog**), have mandatory onsets, i.e. there are no vowel-initial syllables (and hence no vowel-initial lexical bases). This fact is often overlooked or misinterpreted because the practical orthographies in use for the Philippine languages do not include a regular representation for syllable-initial glottal stops. There are both phonological and phonetic reasons to assume that these syllable-initial glottal stops are phonemic and not just automatically added in the absence of an onset. For instance, base-initial glottal stops, just like any other consonant, are not omitted when a prefix is added. Thus when Tagalog *mag-* is combined with *ʔingáy* 'noise' the result is [mag.ʔi.ŋáj] and not *[ma.gi.ŋáj]. Consequently, there is no need to develop complicated accounts for infixation and reduplication of supposedly vowel-initial bases in these languages (cf. also Halle 2001:156).

The Philippine languages provide the most significant exception to the generalization that stress is non-distinctive. In most of these languages, stress placement is not predictable and may occur on either the penultimate or the ultimate syllable. Small classes of lexical bases with unpredictable final stress are also said to occur in a few other languages, including Atayal, Thao (Blust 2001:327), Toba Batak and Dobel. In a number of other languages (e.g. Pazeh, Acehnese, Balinese, Iban, Kendayan, Land Dayak), stress regularly occurs on the ultimate syllable rather than the penultimate syllable.

The tonal distinctions found in **Cham** and **Moken/Moklen** are part of their Southeast Asian mainland typological profile. Otherwise, western Austronesian languages generally do not have lexical tones, a possible exception being Ma'ya, a language spoken on the Raja Ampat Islands in eastern Indonesia (cf. Van der Leeden 1997 and Remijsen 2002).

Sundanese is well known for extensive nasal spreading (Robins 1957, Cohn 1990). To a somewhat lesser extent, this also occurs in Madurese, Balinese, and a number of **Malayic varieties**, particularly in Sumatra and Borneo (e.g. Iban, Kendayan). In Borneo languages, one additionally finds nasal preplosion and related phenomena (Blust 1997a) which are also found in neighboring non-Austronesian languages on the Southeast Asian mainland.

A widespread phenomenon superficially contravening the preference for disyllabic lexical bases is the addition of paragogic vowels to consonant-final bases, a phenomenon particularly widespread in Sulawesi (**Buol**, Sneddon 1993) but also found in many other languages throughout the area (e.g. Atayal, **Tsou**, **Kambera**, Buru). This additional vowel is extra-metrical in that it is not counted for stress assignment and usually

disappears in compounding and when suffixes are added to the base. Thus, for example, the Lauje word *luba?e* 'hair', which consists of the base *luba?* and paragogic *e*, is regularly stressed on the antepenultimate syllable [lúba?e]. The paragogic *e* disappears when the third person possessive suffix *-(o)nye* is added to the base as in *luba?onye* 'his/her hair'. See also the 'weak final syllables' in **Malagasy**.

Turning to morpho(pho)nology (or morphophonemics), the basic generalization would appear to be that Austronesian languages are morphonologically simple in that the relation between basic and derived words is formally transparent, often involving a straightforward concatenation of formatives. There are significant exceptions to this generalization, as briefly shown in section 2.3 below. But inasmuch as it is true, it results from the fact that most languages show only a very limited set of morphonological alternations. It would be wrong, however, to conclude from this that the inventory of morphonological alternations across the whole group is equally limited. On the contrary, there is hardly a morphonological process that is not attested in at least one western Austronesian language. **Leti** and Meto (Dawanese) have complex and pervasive metathesis rules (limited to a few affixes, metathesis occurs throughout the area, e.g. **Biak**, **Taba**, Buru, Cebuano, **Tagalog**, Bontok, Atayal). In central Sulawesi and northern Borneo, some affixations involve vowel harmony (e.g. Lauje, Totoli, **Kimaragang**). Ablaut is found in a number of Borneo languages including **Belait** (Blust 1997b). Umlaut-like alternations are attested in **Kambera**. Consonant mutation has been described for **Nias** and **Kambera**. This is not to mention various other kinds of alternations such as assimilation, dissimilation, deletion and addition of segments which are well-attested throughout the area. The two most pervasive and characteristic alternations, however, are reduplication and nasal assimilation and substitution, which will be discussed in more detail below.

2.1 Nasal assimilation and substitution (*N* and *-um-*)

Many symmetric voice languages have a prefix or a set of prefixes which end in a nasal (*CVN-*) or simply consist of a nasal (*N-*) which assimilates to, and sometimes also modifies, the initial segment of the base in various ways. A notable exception is Formosan languages where only fossilized reflexes of prefixes with *N* are encountered (Zobel, to appear). Such fossilized reflexes are also found throughout the rest of the Austronesian family, and it is in this sense that nasal assimilation and substitution can be considered a highly characteristic feature of the family even though it is no longer productive in half of the languages.

The homorganic nasal, as it is conventionally called, is typically realized as a velar nasal (/ŋ/) before vowels. It remains unrealized before nasals (and often also before liquids and glides), and before other consonants it assimilates to the place of articulation of the following consonant. Voiceless consonants other than /h/ are usually deleted after the homorganic nasal. The Lauje examples in Table 5.1 illustrate these general regularities. See Newman (1984) and Pater (2001) for a broader range of examples and a review of the theoretical issues involved.

As the examples in Table 5.1 make clear, the assimilation to the place of articulation of the following consonant is only an approximate one in several instances. While in the case of the glottal stop and fricative (/ʔ/ and /h/) the non-existence of an appropriate nasal 'explains' the lack of a precise match, this is less clearly so in the case of the alveolar fricative /s/ which is frequently replaced by palatal /ɲ/ rather than expected alveolar /n/. When /s/ is not replaced, but only preceded by *N*, the homorganic nasal is always /n/ as in Lauje *monsau* 'rub (strongly)' (< *sau*).

TABLE 5.1: NASAL ASSIMILATION AND SUBSTITUTION IN LAUJE

Assimilation	Substitution	Examples involving the actor voice prefix *moN-* (given in practical orthography)
N → (ŋ/ __V	–	*mongupi* 'dream' < *upi*
N → 0/__l/r/m/n/ŋ	–	*molandas* 'pull' < *landas*, *morapang* 'give a speech' < *rapang*, *mompoyonga?* 'shut (eyes)' < *mpoyong*, *montuul* 'lie' < *ntuul*, *mongkelung* 'lie down' < *ngkelung*
N → m/ __b	–	*mombanit* 'bite off' < *banit*
N → m/ __p	+	*momangang* 'chew betelnut' < *pangang*
N → n/ __d/ɟ	–	*mondangoy* 'bake sago' < *dangoy*, *monja?ang* 'cook' < *ja?ang*
N → n/ __t	+	*monapa* 'roast' < *tapa*
N → ɲ/ __s	+	*monyunsut* 'suck' < *sunsut*
N → ŋ/ __g	–	*monggeges* 'rub' < *geges*
N → ŋ/ __k/ʔ	+	*mongikib* 'gnaw' < *kikib*, *mongoyab* 'fan' < *ʔoyab*

Not covered by the regularities stated so far are stems beginning with fricatives other than /s/ and /h/. Such stems with an initial bilabial (/ß/, /Φ/), labio-dental (/v/, /f/), or velar fricative (/ɣ/, /x/) are found in a number of languages throughout the area (they seem to be particularly common in Taiwan and Sulawesi). Unlike /s/ and /h/, these fricatives are often strengthened to a stop when co-occurring with *N* (but see **Malagasy** where initial /f/ and /v/ are replaced by /m/). The bilabial and labio-dental ones generally become /b/ as in the Tomini-Tolitoli languages Taje *mombaβa* 'carry' (< *βaβa*) or Tajio *mombeen* 'give' (< *Φeen*), while the velar ones become /g/ as in Ratahan *munggoreng* 'fry' (< *xoreng*), regardless of voicing.

The preceding remarks already indicate that there is a lot of variation regarding the morphonology of *N*, both within and across languages. Substantial variation also occurs, for example, with regard to substitution where just about everything is possible, except that nasals, liquids and glides appear never to be replaced. Voiced obstruents may also be replaced, e.g. Lauje *momambal* 'report, inform' (< *bambal*). And conversely, voiceless obstruents may remain in stem-initial position after *N*, e.g. Lauje *montanong* 'bury' (< *tanong*). Sometimes, forms both with and without substitution occur, e.g. Lauje *mombentet* or *momentet* 'tear' (< *bentet*). This variation is particularly pervasive with regard to initial /b/ which is replaced sporadically in almost all languages with a productive *N*-prefix, and in quite a number of languages substitution is the rule rather than the exception. Other voiced stops (in particular /d/ and /g/) are much more rarely substituted by *N*. Note that in a few languages, such as **Buol**, Iban, Salako and Balinese, substitution of stem-initial obstruents has been generalized and thus pertains to voiceless and voiced obstruents in the same way (in Balinese for lexical bases of two or more syllables only). In a few other languages, mostly in the central parts of Sulawesi (e.g. **Mori Bawah**, Kaili, Mandar), voiceless consonants regularly become prenasalized rather than being replaced by *N*.

For other examples of variation in the morphonology of *N*, see in particular **Nias**, **Karo Batak**, **Kimaragang**, **Malagasy** and **Makassar**.

In passing, it may be noted that the variation with regard to substitution allows a distinction of different kinds of segmentally homonymous *N*-prefixes. Thus, for example,

in Tagalog there is a stem-forming prefix *paN-* after which base-initial voiceless obstruents are regularly omitted. And there is a second prefix *paN-* used to derive words meaning 'for (use in) X'. With this second prefix, substitution of voiceless obstruents is either impossible or optional, hence *pangkapé* 'for (use in) coffee' (< *kapé*) and *pansukláy* or *panukláy* 'for use in combing' (< *sukláy*). See DeGuzman 1978 and also **Karo Batak**.

While nasal assimilation and substitution is most common and widespread with *N*-prefixes, somewhat similar processes also occur with regard to the infix *-um-*, although on a much more restricted scale. The widely attested infix *-um-* is generally inserted between the first consonant and vowel of the base (i.e. *-um-* + CVCCVC → C<*um*>VCCVC). Before vowel-initial stems, it often simply becomes a prefix, sometimes changing its shape from *um* to *m* (i.e. *-um-* + VCCVC → *um*-VCCVC or *m*-VCCVC). Exceptions to these general regularities pertain to bases with an initial labial or nasal consonant and to derived stems with certain derivational prefixes. Languages differ quite significantly with regard to the number of exceptions.

Muna provides one of the more complex examples (Van den Berg 1989:28–31). In addition to showing many standard examples for the infixation of *-um-* such as *s*<*um*>*olo* 'flow', the morphonology of Muna *-um-* includes the following subregularities. Before vowels, *-um-* is regularly *m-* (e.g. *m-ala* 'take'). Voiceless labials (/p/ and /f/) are regularly replaced by *m* (e.g. *mongko* 'kill' < *pongko* and *mutaa* 'laugh' < *futaa*). Bases with initial voiced labials, nasals or prenasalized consonants remain unchanged (*baru* 'happy' may represent either the unmarked base or *-um-* + *baru*). Bases with initial *w* may remain either unchanged (e.g. *wanu* 'get up' representing *-um-* + *wanu*) or *w* is replaced by *m* (e.g. *maa* 'give' < *waa*). Prefixes either abide by these regularities (e.g. causative *fo-* becomes *mo-* when affixed with *-um-*, verb stem deriving *ka-* becomes *kuma-*) or remain unchanged (e.g. detransitivizing *fo-* remains *fo-* in contexts where use of *-um-* is required).

With regard to derivational prefixes, three different scenarios have to be distinguished in all western Austronesian languages. First, *-um-* may be affixed to the prefix according to the same rules as for lexical bases. Second, *-um-* may be affixed to the prefix, but there are special regularities (e.g. the *um*-form of the Tagalog social action prefix *paki-* is *maki-* (not **pumaki-*) while with most *p*-initial lexical bases *-um-* is regularly infixed after the initial consonant as in *p*<*um*>*untá* 'go to'). Finally, the prefix may block affixation of *-um-*, i.e. the stem has to be expanded further before affixation of *-um-* is possible (e.g. the Tagalog causative prefix *pa-* disallows the affixation of *-um-*).

While the morphonological alternations of *N* and *-um-* share some similarities, they must have quite different (historical) origins. The widespread manifestation of *-um-* + *p*-initial prefix as *m* as in the above Tagalog example (*-um-* + *paki-* → *maki-*) is probably due to clipping the first syllable of a trisyllabic prefix (e.g. **pumaki-* → *maki-*) rather than to *-um-* somehow replacing the prefix-initial *p*. In this view, at least some of the very common prefixes of the shape *ma-* may be considered to be reduced forms of (historically as well as synchronically) underlying **kuma-*, given that *ka-* is a very common stem deriving prefix (and despite the fact that *kuma*→ *ma* does not make sense in terms of assimilation and substitution).

Finally, all alternations relating to *N* and *-um-* should be kept distinct from the alternation between non-realis marking *m* and realis marking *n* which is common with prefixes in many Philippine languages (e.g. stative non-realis *ma-* vs. stative realis *na-*). This *m/n*-alternation never involves assimilation or substitution, and it easily creates confusion if it is also represented by a capital N.

2.2 Reduplication

Reduplication is probably the most pervasive morphonological process in western Austronesian languages in that it is a productive process in all of them (a possible exception is Keo). However, Austronesian languages differ greatly in the formal make-up of reduplication patterns and their functions, as will become evident throughout this section and by comparing the language sketches. Here as well as in the language sketches the focus is on formal patterns.

To begin with, it will be useful to recall the distinction between reduplication as a morphonological process of word formation and the mere repetition of words as a means to convey intensity, multiplicity or duration. The latter is a syntactic pattern which is not constrained by morphonological parameters and involves the repetition of whole (phonological) words regardless of their morphological make-up. Such repetitions are possible in all languages (e.g. English *go go go!*). While this distinction is fairly clear on a conceptual level, its application in practice is sometimes not straightforward, as the following example from Standard Indonesian shows.

In Standard Indonesian it is possible to repeat any noun in order to indicate a plurality of referents as in *rumah-rumah* 'houses' or *anjing-anjing* 'dogs'. Such repetition is possible regardless of the morphological complexity of the noun, compare *perubahan-perubahan* 'changes' (< *ubah*). Grammars generally treat this as an instance of reduplication, but it is a borderline case where the analysis as reduplication depends very much on whether the duplication can be shown to convey a grammaticized meaning or function. Full base reduplication can indeed have specific semantic effects in Standard Indonesian in that it allows the derivation of new lexemes (e.g. *guna-guna* 'magic' < *guna* 'use', 'purpose', *mata-mata* 'spy' < *mata* 'eye'), which however only applies to underived lexical bases. Furthermore, there are many instances where full base reduplication only pertains to the lexical base, e.g. *sebesar-besarnya* 'as big as possible' < *besar* 'big'. This also holds for verbs which may be reduplicated to indicate a repeated or ongoing event as in *berjalan-jalan* 'walk about', 'go for a stroll'. Since in the case of morphologically complex bases only the lexical base is repeated, it is plausible to assume that in instances such as *duduk-duduk* 'sit about' we are also dealing with reduplication rather than with simple repetition.

With regard to reduplication, two types of functions or meanings may be distinguished in Austronesian languages. On the one hand, there is the somewhat diffuse and supposedly iconic range of meanings associated with reduplication throughout the world's languages, i.e. distribution, variety, plurality; habitual, repeated or ongoing activity; intensity and emphasis; increase or decrease of size or amount; pejoratives. This range of meanings is often conveyed by full base reduplication (or simple repetition) rather than by the partial reduplication patterns exemplified below.

On the other hand, reduplication patterns in western Austronesian are part of a large number of clearly circumscribed derivations, often in combination with further affixes. Thus, for example, reduplication occurs quite regularly in nominalizations such as the formation of instrument nouns (e.g. Thao *ca-capu* 'broom' < *capu* 'sweep' (Chang 1998:282); see also Blust 1998). But it also serves more purely morphosyntactic functions such as marking adnominal modifiers (including the verb in relative clauses) in Dobel (see also **Leti** and Carpenter 1996):

(3) *ʔamu* *sertáy* ***m-maray*** (vs: *ʔamu sertáy maray* 'your clothes are dry')
2S.POSS clothes RDP-dry
'your dry clothes' (Hughes 2000:172)

Among the most frequent and widespread uses of reduplication is the derivation of numerals where it forms restrictives (e.g. **Iloko** *dù-duá* 'only two'), ordinals (e.g. Siraya *ka-ra-ruha* 'second' (with prefix *ka-* preceding *Ca*-reduplication), grouping (Malay *dua-dua* 'two at a time'), etc. Somewhat unusual is its use when counting humans as opposed to other entities. Thus Thao *tusha* 'two' occurs in serial counting or when counting animals or inanimate objects while *ta-tusha* necessarily refers to human referents (Blust 2001:332f). Perhaps even more remarkable in this is Siraya (Adelaar 2000:41,48) where two different kinds of reduplication apply when counting non-humans (CV-reduplication) and humans (Ca-reduplication).

It has occasionally been suggested that a possible third type of function for reduplication consists in the formation of lexical bases (sometimes called *inherent reduplication*). As is well known, many disyllabic lexical bases in western Austronesian languages have the shape of duplicated monosyllables as in Balinese *gigi* 'tooth', *agag* 'wide open' or *cakcak* 'chop up'. These are generally deemed to be fossilized forms, reflecting earlier productive reduplication patterns (see Blust 1988). But Clynes (1995:166–170) argues that, at least in the case of Balinese, there is evidence to suggest that such bases are the result of synchronic base formation via reduplication. Thus, for example, Balinese duplicated monosyllables are stressed on both syllables (hence *gígí*) just like productive full-base reduplications (e.g. *lúh-lúh* 'female (plural)' < *luh*) and unlike other disyllabic lexical bases, which are generally stressed on the final syllable.

Apart from the full base-reduplication illustrated above, the following partial reduplication patterns are attested among western Austronesian languages:

- *C-reduplication*: the first consonant of the stressed syllable is directly prefixed to it as in Dobel ***m***-mata 'raw' or *ʔa-**d**-dém* 'she does' (<*ʔa-dém*) (Hughes 2000:168; syllable onsets are always filled in Dobel) or **Iloko** *lalláki* 'males' (< *laláki* 'male'). This form of reduplication is quite rare, often being restricted to a small set of items as in the case of Iloko where it occurs in animate and kin plurals only (see also Thurgood (1997:144) on Bontok). Dobel and its closest relatives are possibly the only languages where this is the main productive reduplication pattern.
- *Ca-reduplication*: The first consonant is copied and prefixed to the base with a default vowel intervening, e.g. Thao ***fa**-finshiq* 'seed for planting', ***sha**-shishi* 'shake repeatedly' (Chang 1998:282) or Balinese ***də**-daar-an* 'food' (< *daar* 'eat'; Clynes 1995:152). The default vowel is often *a* (e.g. Paiwan, Ratahan, **Tetun Fehan**), which is why this pattern is referred to as *Ca*-reduplication. Other vowels are also possible, as shown by Balinese and **Javanese** which have schwa, or Buru which has *e* (Grimes 1991:77). In **Taba** a default vowel *a* also occurs in more complex reduplication patterns (e.g. *CaC*-). Clynes (1995:154f) suggests that Ca-reduplication is actually a variant of C-reduplication in those languages which do not allow geminate consonant clusters. See Blust (1998) for a more comprehensive survey. Note that the difference between Ca-reduplication and CV-reduplication is automatically lost in languages with antepenultimate vowel neutralization (e.g. Atayal, **Seediq**, Malay, **Javanese**, **Malagasy**, Madurese).
- *CV-reduplication* or *monosyllabic reduplication* (also *syllable reduplication*): The reduplicant consists of a syllable prefixed to the base. This syllable can be either light (monomoraic) or heavy (bimoraic), and some authors would restrict the term *CV-reduplication* to instances of light syllable reduplication as in **Tagalog** *pag-**bi**-bilí* 'selling' and *pag-**la**-lakbáy* 'travelling', where the first syllable of the base is heavy *lak* but the reduplicant is light *la*. In heavy syllable reduplication, either the vowel in the reduplicant is stressed or lengthened (e.g. **Tagalog** ***bí**-bilí* 'will buy') or the syllable is obligatorily closed, compare Bontok *ʔik-ʔik.kan* 'is doing' where the base-initial syllable is already heavy (*ʔik*) with ***lap**-la.pú* 'is going first' where it is not (Thurgood 1997:137). In particular in Philippine languages, it is common to have both heavy and light syllable reduplication in clearly different functions. While the segments in the reduplicant are usually a copy of the base-initial segments, there are also examples where the reduplicant regularly consists of a copy of the base-final syllable (e.g. Madurese *les-toles* 'write (more than once)', *ku-buku* 'books' (Davies 1999a:13) or Urak Lawoi' *jiʔ-bajiʔ* 'well' (Hogan 1999:24)).
- *CV(C)CV-reduplication* or *disyllabic reduplication* (also *foot reduplication*): The reduplicant consists of a combination of two syllables (i.e. a foot) prefixed to the base. The second syllable is usually open, regardless of the shape of the second syllable of the base, e.g. Lauje *e-**ʔinde**-ʔinde* 'nod indiscriminately', *me-**ito**-itong* 'rather black', *ma-**ale**-alenda* 'rather long'. In the case of disyllabic bases with an open final syllable such as *ʔinde*, this kind of reduplication is indistinguishable from full base reduplication, which may be the reason why it is occasionally also called *full reduplication*. But as the preceding examples show, it is not a literally full duplication of the base. Note that in many languages, foot reduplication occurs in addition to full base reduplication as in Lauje *mong-**ontong**-ontong* 'watch (for some time)' (where the base-final *ng* is part of the reduplicant as opposed to *me-ito-itong* where it is not).

Recently, it has been proposed that some western Austronesian languages also allow triplication, i.e. adding a given reduplicant twice to the same base. The best documented example to date is Thao, for which Blust (2001:330) adduces the following example among others: *qa-qa-qucquc* 'to tie or bind tightly or securely' with doubled Ca-reduplication from the base *qucquc* 'tie, bind' (see Müller-Gotama (2001:16) for possible examples from Sundanese).

Very many western Austronesian languages allow Ca-reduplication or one or more variants of CV-reduplication in addition to foot reduplication and full base reduplication. It is rare that a language allows both CV- and Ca-reduplication as productive processes applying to the same set of lexical bases. One example, already mentioned above, is Siraya, where numerals occur both with CV- and Ca-reduplication (see also Chang (1998:285) and Blust (2001:326f) on Thao and Li and Tsuchida (2001:21f) on Pazeh). It is thus not impossible that C-reduplication, Ca-reduplication and (light-syllable) CV-reduplication are actually variants of each other (in the case of Siraya, an extinct Formosan language, it is unclear whether the reduplicant in CV-reduplication was heavy).

While the majority of reduplicants are clearly prefixes (i.e. they are attached to the left edge of the base), there are a few examples which suggest attachment further to the right (sometimes called rightward reduplication). Thus, Chang (1998) lists the following examples from Thao in which material from the end of the base, excepting the final C if there is one, is copied and appears to be attached close to the right edge of the base (similar formations are found in Rukai and Pazeh (Chang 1998:296) and Siraya (Adelaar 2000:40f):

(4) **Thao 'rightward' reduplication** (cf. Chang 1998:284)

	BASE	REDUPLICATED FORM (reduplicant in italics)	
(a)	su.hu.i	p-i-suhui*hui*	'to be put there repeatedly'
(b)	shna.ra	pa-shnara*nara*	'to burn sth. repeatedly'
(c)	ag.qtu	agqtu*qtu*	'think about'
(d)	sig.ki	sigki*gki*	'to kneel'
(e)	ma-ku.tnir	mia-kutni*tnir*	'to harden'
(f)	dut.khun	mia-dutkhu*khun*	'to hunch over'
(g)	m-ar.faz	m-arfa*rfaz*	'to keep flying around'

The reduplicant is either monosyllabic with a heavy onset (as in c-g) or disyllabic with simple onset(s) (a and b). The choice of the reduplicant depends on the shape of the base. Roughly, if the rightmost syllable of the base has a heavy onset, then this syllable is copied (c, e, f). If its onset is light and the preceding syllable is closed, the reduplicant consists of the coda of the penultimate syllable plus the final syllable (d, g). Otherwise the last two syllables are copied (a, b), a possibly complex onset of the penultimate syllable being simplified (b). This reduplication pattern is in complementary distribution to 'regular' foot reduplication which applies to monosyllabic bases and disyllabic bases without complex onsets or a coda in the penultimate syllable (Chang 1998:279).

Since the reduplicant occurs to the left of the final consonant of the base (if there is one) this is not a clear-cut example of suffixing a reduplicant. In fact, as it stands, there is no clear-cut evidence to consider the second string of identical segments as the reduplicant. That is, rather than *agqtu**qtu*** and *marfar**faz*** it is also possible to analyze these forms as *ag**qtu**qtu* and *mar**far**faz*, locating the reduplicant to the left of the main (i.e. stressed) part of the base. In this way, these forms look much more like standard examples of CV- or CVCV-reduplication, the major difference being that the reduplicant is not attached to the left edge of the base but placed directly before the stressed syllable of the base. This is how Van Klinken analyzes somewhat less complex, but still similar forms in **Tetun Fehan**.

In Balinese, there are a few examples in which the case for suffixation is clearer in that the reduplicant in fact includes the base-final consonant as in *pə-cəpol**pol*** 'collapse (plural)' (< *-cepol*) or *ngəbət**bət*** 'throb (plural)' (< *-kəbət*). But Clynes (1995:157) notes that the first syllable(s) in bases allowing final syllable reduplication is always from a severely restricted set (*kə, cə, kati/kali/katu, nərə/nylə*), which has prefix-like characteristics. Consequently, he argues for leftward attachment (i.e. *pə-cə**pol**pol*) in analogy to the standard cases such as *me-laib-laib* 'run (plural)' (< *laib*) where prefixes are also excluded from reduplication. See also Li and Tsuchida (2001:22) on Pazeh.

As the preceding examples from Thao and Balinese show, the structure of the base often determines the choice of reduplication pattern. Consequently, reduplication patterns are often in at least partial complementary distribution within a given language. In addition to the number of syllables, a major parameter here is whether the base is vowel- or consonant-initial. Thus, for example, in Lauje CV-reduplication is only possible for consonant-initial bases (e.g. *me-**ní**-nimpis* 'rather narrow' (< *nimpis*)). Vowel-initial roots only allow foot reduplication as in *me-**ito**-itong* 'rather black' (Himmelmann 2001:81f). Other languages allow the reduplicant to consist simply of a vowel which then is usually separated by a glottal stop from the base (e.g. in Ratahan *Ca*-reduplication for vowel-initial roots consists in prefixing the vowel /a/ to the root as in ***a**-ingka?-en* [ʔaʔiŋkaʔen] 'messenger' < *ingka?* (Himmelmann and Wolff 1999:15)).

A very complex area with much crosslinguistic variation is the interaction of reduplication and 'regular' affixation. The general tendency here is that many affixes, in particular monosyllabic ones, do not take part in reduplication but rather appear to be added to a derived stem consisting of reduplicant+lexical base. Thus, for example, infixes are almost always added to the reduplicant and not to the lexical base (e.g. Tagalog *bini-bilí* 'bought', not **bí-binilí*; see **Kimaragang** for a noteworthy exception to this generalization). When prefixes involve the homorganic nasal, however, it is generally the case that nasal assimilation and replacement modify the base before reduplication takes place (e.g. Tagalog *ma**mú**mútol* 'will cut (a lot)' (< *maN-putol*) not **ma**mú**pútol*).

Further morphonological aspects of reduplication relevant to western Austronesian languages pertain to the simplification of base-initial onset clusters (e.g. Tagalog *mag-**tá**-trabaho* 'will work', not **mag-tra-trabaho*), the insertion of nasals or glottal stops in between reduplicant and base, metathesis and various assimilation processes at the boundary of reduplicant and base. See the language sketches most of which contain data on reduplication.

2.3 Lack of morphonological transparency

Lack of morphonological transparency as understood here means that the relation between basic and derived words is formally not transparent in that the constituent formatives of a morphologically complex word are not easily identifiable. Once the changes brought about by the homorganic nasal are mastered, identifying constituent formatives in a western Austronesian language is generally not very difficult. However, there are two important and often ignored types of exceptions. On the one hand, there are a few languages such as **Leti** which lack morphonological transparency due to pervasive metathesis (and apocope). On the other hand, there are a few languages such as Atayal, Siraya, **Seediq**, **Tsou** or Gorontalo where in particular verbal derivations involve a whole set of largely reductive alternations, the combined application of which tends to obscure the relation between base and derived formation. This second type of exception is briefly illustrated here with data from Atayal.

The most pervasive morphonological alternation in this language is the reduction of all antepenultimate vowels to schwa. Thus, when a disyllabic root with penultimate stress is suffixed, the formerly stressed vowel is reduced to schwa as in *tápih* 'call' → *təpíhun* (*-un* is the patient voice suffix). (Since schwa is predictable, Atayal words are usually written without schwa, i.e. *təpíhun* is represented as *tpíhun* in standard orthography.) If a suffixed formation is stressed on the final syllable, all prefinal vowels are reduced to schwa: *tehúk* 'arrive' → *təhəkán*, *laqúx* 'win' → *ləqəxán* (<q> represents a uvular stop and *-an* is the locative voice suffix). In addition to this regular and pervasive reduction, whole syllables are dropped under various conditions. For example, when the two last syllables of a base are segmentally identical, one of them may be dropped in suffixation, e.g. *kəŋúŋu?* 'fear' → *kəŋún* (with suffix *-un*) and *kəŋúi* (with suffix *-i*).

In addition to these reductions, various consonant alternations occur. As the previous example *kəŋúŋu?* → *kəŋún* shows, root-final glottals may be dropped with subsequent contraction of sequences of identical vowels. But sometimes glottals are retained (e.g. *həŋəʔán* 'dip in water' < *həŋú?*), or they become a glide, which happens in roots ending in /iʔ/ or /uʔ/ (e.g. *siyun* < *siʔ* 'to place').

This does not yet complete the inventory of morphonological alternations in Atayal. But the following examples suffice to show what happens to morphonological transparency when several of these alternations apply together to a given derivation: *gálan* < *ágan* 'take' + *-an* (consonant alternation, stress shift, syllable loss), *kətón* < *kíta?* 'see' + *-un* (loss of glottal stop, vowel contraction, stress shift, vowel reduction), *həbəgán* < *həgúp* 'do magic' + *-an* (consonant alternation, metathesis, vowel reduction).

2.4 A note on morphological type

The large majority of western Austronesian languages shows a moderate inventory of affixes, mostly prefixes and a few suffixes (less than five, not counting pronominal suffixes). Productive infixation (involving usually two infixes) is largely confined to the northwest (Taiwan, Philippines and northern Borneo and Sulawesi). The typical number

of prefixes is somewhat difficult to determine since it depends very much on whether complex prefixes are counted as prefixes in their own right (e.g. Tagalog *maka-* could be analyzed as an unanalyzable prefix of its own or as a combination of the two independently attested formatives *ma-* and *ka-*). The number of clearly monomorphemic prefixes rarely exceeds thirty, many of which have very special functions (e.g. the **Iloko** prefix *agat-* deriving words for smelling as in *agat-layá* 'smell like ginger'). Prefixes with a high functional load and text frequency typically do not exceed a dozen (counting person marking prefix sets as single prefixes). Such prefixes, as well as all suffixes, tend to be highly multifunctional.

Major exceptions to these generalizations are found on the Southeast Asian mainland, in western Borneo, where various Land Dayak languages exhibit very little morphology, and the Flores-Timor region, where a number of (nearly) isolating languages occur (e.g. Keo, Waima'a). Keo, in fact, is said to have no affixes whatsoever, only a single proclitic and no productive native reduplication patterns (Baird 2002:171). **Kambera**, spoken in neighboring Sumba, is also rather poor in affixes (one productive prefix, possibly two suffixes), but very rich in clitics.

In terms of morphological typology, most western Austronesian languages have been considered to be agglutinative. This assessment needs to be qualified, however. First, there is a considerable number of languages which are clearly not agglutinative because they exhibit pervasive metathesis, consonant mutation, ablaut or reductive alternations of one form or another, as was briefly exemplified in the preceding section. Second, the traditional categories *agglutinative* and *fusional* are not simple properties, but sets of logically independent properties, as explained in Plank (1999:282f). The formally more transparent Austronesian languages are agglutinative only with regard to some of these properties.

In terms of the formal concatenation of formatives, in particular transparent segmentability and phonological cohesion, many Austronesian languages indeed show few signs of fusion. Fusion with regard to these properties is very much limited to the ubiquitous homorganic nasal and a few other assimilatory processes. But even these languages, with their easily identifiable morpheme boundaries, are not agglutinative in the same way as Turkish, for example, because they deviate quite significantly from the ideal of 'one form – one meaning'. The major formatives not only tend to be multifunctional, but they also often convey a *bundle* of morphosyntactic features rather than representing a single category (for example, many so-called voice affixes are strictly speaking voice-aspect-mood affixes because they always convey a combination of these categories; see section 4.2.2). Conversely, the same category (e.g. locative voice) may be represented by two different formatives (as with locative voice which is represented by both *-an* and *-i*). Furthermore, most languages allow combinations of affixes (prefix plus suffix, two or more prefixes, etc.), and while some of these affix combinations are semantically transparent, others are quite idiosyncratic and demand an analysis as unit morphemes.

3 BASIC MORPHOSYNTAX

3.1 Lexical and syntactic categories (parts of speech)

In discussing lexical and syntactic categories in western Austronesian languages, it is necessary to make a fundamental distinction between morphological and syntactic (distributional) levels and between lexemes (lexical bases) and morphosyntactic words. It is a prominent feature of these languages that categorial distinctions on these different

levels do not necessarily align in the same way as in other language families. For instance, two morphosyntactic words may differ clearly in that they participate in different paradigms and thus belong to two distinct morphological categories. At the same time, however, their syntactic distribution may be identical, thus belonging to the same syntactic category. Similarly, morphosyntactic words in a given language may clearly belong to different morphological or syntactic categories but at the same time there may not be a corresponding distinction on the level of lexical bases (roots) (see Sasse 1993, Evans 2000 and Himmelmann (to appear a) for a more detailed overview of different alignment possibilities). In the following discussion, morphosyntactic words are taken up first.

A basic distinction with regard to morphosyntactic words which is widely attested in the languages of the world is the one between content words (open word classes) and function words (closed word classes). This distinction is also found in all western Austronesian languages, with most morphosyntactic words clearly belonging to one or the other category. Exceptions pertain to weakly grammaticized items such as **Belait** *saw'* which functions both as a main predicate 'use' and as an instrumental preposition. Many function words are clitics, which are further discussed in section 3.2.

Western Austronesian languages differ somewhat with regard to the further subdivision of content words. In many languages there is a clear-cut syntactic distinction between verbs and nouns in that there are syntactic slots in which only nouns may occur (note that in most languages nouns can occur without further modification in predicate function so that typically there are no syntactic slots that are exclusively filled by verbs). This is the case in most preposed possessor and transitional languages including **Biak**, **Taba**, **Leti**, **Mori Bawah**, and **Nias**, where verbs are marked for person. In **Kambera** there is some distributional overlap between nouns and verbs, but the basic distinction is still clear.

In many symmetrical voice languages, on the other hand, the syntactic distinction between nouns and verbs is often somewhat less clearly delineated in that word-forms which semantically appear to be verbs easily and without further morphological modification occur in nominal functions and vice versa. Compare the following examples from Riau Indonesian (see also **Tagalog** and **Seediq**):

(5) *orang* ***bahasa*** ***Inggeris*** *sama* *David*
person language English with David
'Did people speak English with you (i.e. David)?' (Gil 1994:182)

(6) ***tunggu*** ***taksi*** *susa* *sekali*
wait taksi difficult very
'Waiting for a taxi is very difficult.' (Gil 1994:182)

In (5) the 'noun' *bahasa Inggeris* 'English' is used in predicate function and then conveys the clearly verbal meaning 'speak English'. Conversely, in (6) the 'verb' *tunggu* 'wait' with its complement *taxi* is used in subject function and has to be rendered by a nominalization in English.

Sometimes a distributional distinction between nouns and verbs pertains only to one or two fairly specific syntactic contexts. Thus in Standard Indonesian a major distributional difference between nouns and verbs pertains to the fact that nouns cannot be negated with *tidak* and that when verbs are negated with the nominal negator *bukan* the negation is emphatic or contrastive (see also **Sama**, **Kimaragang** and **Belait**).

The analytical consequences of a pervasive overlap in the syntactic distribution of putatively nominal and verbal word-forms remain controversial. Some authors (e.g. Gil 1994, 2000, Himmelmann 1991, to appear a) argue for a basic lack of a morphosyntactic noun/verb distinction. Most descriptive grammars and formal analyses of symmetrical voice languages assume underlying syntactic differences based on the semantics of the forms and analyze the examples above as involving zero conversion. Apparently nominal uses of 'verbs' are often also considered headless relative clause formations (see **Kimaragang**).

Analysts also differ widely as to whether it is useful and necessary to distinguish a separate morphosyntactic class of adjectives in addition to nouns and verbs. Here, however, one has to be particularly careful to determine the level at which such a distinction is claimed, and to distinguish differences in actual fact from differences in descriptive practice. It would appear that in most western Austronesian languages putative adjectives have the same kind of syntactic distribution as intransitive (particularly stative) verbs. Thus, for example, in languages where negators provide a diagnostic context for distinguishing nouns and verbs, putative adjectives always behave like verbs. Exceptions usually pertain to very small classes of words conveying property concepts which show some minor distributional differences distinguishing them from verbs (e.g. **Biak** *kasun* 'small' and *babo* 'new'; see also **Seediq**).

While there is thus little evidence to set up *adjective* as a distinct syntactic category, in various languages throughout the area there are good reasons to set up *adjective* as a distinct morphological category (e.g. **Iloko**, **Karo Batak**, **Malagasy**, **Leti**, **Tetun Fehan**). In these languages, adjectives are distinguished from other morphological categories by their distinctive morphological potential. Perhaps the most widespread distinguishing feature of adjectives is the fact that they have a specific way for expressing plural in agreement with some nominal constituent (this is generally optional and usually involves some form of reduplication). In some languages (e.g. **Iloko**, **Karo Batak**) adjectives additionally occur with specific comparative or superlative affixation unavailable for other classes.

As with adjectives, it is usually also possible to define a category of verbs in purely morphological terms, i.e. as the class of morphosyntactic words which are marked for either voice (and aspect/mood) or for person (and mood). An important difference exists between these two basic types of morphologically defined verbs in that for person-marked verbs the morphological category corresponds to a syntactic one (a person-marked verb usually has a clearly distinct syntactic distribution). Such a correspondence is generally much less clear in the case of voice-marked verbs. Their distribution, in particular in Philippine-type languages, often very closely resembles the distribution of (morphological) nouns (see **Seediq** and **Tagalog** for examples).

From a morphological point of view, nouns in western Austronesian languages are usually unmarked. That is, they are not overtly marked for case, number or gender but rather occur in their lexical base form in most uses (a major exception is **Nias**, where nouns occur in either mutated or unmutated form). Optional plural marking (usually via reduplication) is fairly widespread.

Turning now to the lexeme level, it is frequently noted in descriptions of western Austronesian languages that lexical bases (roots) are underdetermined in allowing both nominal and verbal derivations or uses. Alternatively, a basic distinction between nouns and verbs (and possibly adjectives) is made for lexical bases but then it is stated elsewhere in the grammar that nominal bases can be used as (morphosyntactic) verbs essentially in the same way as verbal bases (for example, by prefixing person markers). Once again one has to separate here differences in actual fact from differences in descriptive practice. The issue tends to be further confounded by the terms *root* and *precategorial*, which are used in widely differing ways in the literature. The following basic scenarios have to be distinguished:

- precategorial bound roots, i.e. lexical bases which do not occur without further affixation or outside a compound in any syntactic function and from which items belonging to different morphological or syntactic categories (nouns and verbs, for example) can be derived, without there being clear evidence that one of the possible derivations from a given root is more basic than the other one(s). This is the way Verhaar (1984:2) defined the term *precategorial root*. There is no western Austronesian language where this type of root is very common although sporadic examples are attested in a number of languages, including **Nias** and a number of **Malayic varieties** (Adelaar 1992:145f). In any event, it would be useful to restrict the use of the term *precategorial* to precisely this state of affairs (see also Clynes 1995:203–205).
- morphologically or syntactically subcategorized bound roots, i.e. lexical bases which do not occur without further affixation in any syntactic function but which clearly belong to one particular morphological or syntactic category because of the affixations they occur with. This type of root is attested in some preposed possessor and transitional languages, including **Nias** and **Biak**. It usually concerns verbal roots which obligatorily have to be marked with a person marking prefix before they can function as predicates. Note that such bound verbal roots typically may also co-occur with nominal derivational morphology (for example, to form instrumental nouns or action nominals). But this does not mean that the roots per se are precategorial because the same derivational morphology also occurs with free or derived verbal bases. It is obviously the function of this morphology to turn verbs into nouns regardless of whether the derivational base is a bound verbal root or some other kind of verbal base form.
- multifunctional lexical bases, i.e. lexical bases which occur without further affixation in a variety of syntactic functions. This type of lexical base is attested on the one hand in the isolating languages of Flores and East Timor (e.g. Keo, Waima'a). On the other hand, it occurs with some frequency in the languages of western Indonesia (e.g. Acehnese, many **Malayic varieties**, Ngaju Dayak and possibly Balinese) where 'verbs' are not necessarily marked for voice or person and the same base allows for verbal as well as nominal uses (e.g. Acehnese *jeu* which denotes both 'a type of net' and 'to catch with a net' (Durie 1985:44)).
- morphologically or syntactically subcategorized lexical bases.

In the large majority of western Austronesian languages, most lexical bases would appear to be of the morphologically or syntactically subcategorized type. Note that this assessment allows the possibility that lexical bases are only morphologically but not syntactically subcategorized (or vice versa) or that they are both morphologically and syntactically subcategorized but that morphological classes do not match syntactic classes.

In some languages (e.g. **Kambera**, **Tetun Fehan**), the evidence for a subcategorization of lexical bases is clear and uncontroversial because there are clear-cut distributional differences between nouns and verbs and most (underived) lexical bases fit either one or the other slot. In many of these languages (e.g. **Leti**), there are very productive conversion processes which, for example, allow all non-human nouns and many adjectives and numerals to be turned into verbs simply by adding a subject-marking prefix. But there appears to be sufficient evidence to assign each lexical base to one basic category which then serves as input to the conversion process.

It is exactly the lack of such evidence which causes the problem of assigning a given base to a single basic category in languages with a large number of multifunctional lexical bases. Three types of analyses for multifunctional bases are found in the literature: (a) The bases are assumed to be 'precategorial' in the sense of being unspecified for a morphological or syntactic category. (b) Lexical bases are assumed to occur in homonymous doublets, one subcategorized as a noun, the other as a verb. (c) Each base is subcategorized as a noun or as a verb (or an adjective), with productive conversion processes allowing for non-basic uses. This last option is primarily applied in those instances where lexical bases differ in morphological potential.

In another set of languages, notably the Philippine-type languages, evidence for the subcategorization of lexical bases also appears to be lacking, and in more recent literature these bases have repeatedly been classified as 'precategorial'. A major characteristic of these languages is that almost all lexical bases can be affixed with voice and aspect-mood affixes. For example, Tagalog *p<um>utol* 'to cut down' from *putol* 'cut' and *b<um>ató* 'to stone/throw stones at' from *bató* 'stone' both contain the actor voice marking infix *-um-*. Consequently, one cannot claim that *bató* and *putol* are different kinds of lexical bases because only one of them needs extra derivational morphology in order to become available for voice and aspect-mood morphology. On first sight, then, it may indeed appear that there is little evidence for a morphological subcategorization.

However, such an assessment is somewhat misleading because lexical bases in these languages typically differ with regard to their morphological potential. Thus, for example, Tagalog lexical bases differ as to whether they take either *-um-* or *mag-* or both of these affixes for actor voice marking (see also **Iloko**, **Tsou**, **Seediq**). Consequently, each Tagalog lexical base has to be subcategorized as belonging to the *-um-*, the *mag-* or the *-um-/mag-* class. Morphological classes of this kind are widely considered to be of purely morphological import, comparable perhaps to the inflection and declension classes well known from Indo-European languages. At least, they are usually ignored in the literature on lexical categories and it is a matter for further research whether there are morpholexical classes which are of interest and relevance beyond the realms of morphology (see Himmelmann (to appear a) for further discussion).

As for a possible syntactic subcategorization of lexical bases in Philippine-type languages, the assessment presented above for morphosyntactic words – i.e. that there is no distributional evidence for distinguishing major syntactic categories – of course also applies to lexical bases. The interesting and often overlooked observation here is that almost all lexical bases in these languages may occur without affixes. This is no surprise in the instance of a putatively nominal basis such as *bató* 'stone'. But putatively verbal bases such as *putol* 'cut' may also be used without affixes (in the same slots as putatively nominal bases), and perhaps even more surprisingly, they convey nominal meanings when used in this way. Thus, *putol* without verbal affixations means 'a cut, a piece' as in:

(7) *ang* ***putol*** *ng* *buhók* *ni* *Huán*
SPEC cut GEN hair PN.POSS John
'John's hair-cut' (Bloomfield 1917:220)

On first sight, these Philippine-type lexical bases may resemble the multifunctional lexical bases mentioned above. However, there are two important differences. In languages with multifunctional bases, there is a genuine morphosyntactic (distributional) difference between nouns and verbs, with verbs being excluded from at least some nominal slots and vice versa. Concomitant with this morphosyntactic difference, multifunctional bases also convey two clearly different meanings depending on which slot they occur in

(Acehnese *jeu* denotes 'a/the net' in a nominal slot and 'to catch with a net' in a verbal slot). In Philippine-type languages, there are no clearly nominal and verbal slots and consequently the meaning of lexical bases does not (in fact, cannot) change in correspondence with a different morphosyntactic function. Instead, it only changes via affixation. In this view, Philippine-type lexical bases are not precategorial, but belong to the last type of bases listed above (i.e. morphologically or syntactically subcategorized): they are syntactically subcategorized as content words and morphologically as belonging to different (derivational) paradigms.

3.2 Clitics

Clitics are probably attested in all western Austronesian languages although there are considerable differences in the number and type of clitics found in a given language (see **Kambera** for a particularly complex example). Formally, a distinction needs to be made between peripheral and second position clitics. Peripheral clitics appear at the beginning or end of the constituent to which they belong (immediately before the verb or at the end of a clause, for example). They are found in practically all western Austronesian languages.

Second position clitics are confined to Philippine-type languages and a few transitional languages (e.g. **Makassar** (only the absolutive pronoun in intransitive clauses), **Mori Bawah** (aspectual clitics only)). They are called second position clitics because they occur after the first constituent of the phrasal unit to which they belong (another term for these clitics is *Wackernagel clitics*). In the following Tagalog examples, the clause core is this unit and the two second position clitics *namán* and *namin* follow its first constituent. In example (a), the first constituent is the predicate *alam*. In example (b), it is the negator *hindí* and the clitics now occur before the predicate *alam*. Non-clitic expressions such as the personal name phrase *ni=Pepito* do not change their position when a negator precedes the predicate (example (c)).

(8) (a) *alam=**namán**=**namin***
knowledge=really=1pe.POSS
'of course we knew'

(b) *hindí=**namán**=**namin*** *alam*
NEG=really=1pe.POSS knowledge
'of course we didn't know'

(c) *nguni't* *hindí* *alam* *ni=Pepito*
but NEG knowledge GEN.PN=Pepito
'but Pepito did not know'

Note that conjunctions such as *nguni't* in (c) do not belong to the clause core and hence do not attract second position clitics. Therefore, 'but of course we knew' is rendered by *nguni't alam=namán=namin* (**nguni't=namán=namin alam* is ungrammatical). See **Tagalog** (section 3.3) and **Kimaragang** (section 3.1.3) for further examples and discussion.

What counts as 'the first constituent of the phrasal unit to which they belong' is not quite as easily determined as the above examples might suggest. Schachter and Otanes (1972:187–193) list a number of complex constructions which appear to function as an uninterruptible unit with regard to clitic placement (see also Kroeger 1993:118–123). Thus, according to Schachter and Otanes, in *isáng taón at apat na buwan siyá* (one LK

year and four LK month 3s) 'he is one year and four months old' the second position clitic *siyá* cannot be placed earlier in the clause (**isá siyá-ng taón at …, *isáng taón siyá at …*, etc.). Unfortunately, such restrictions on clitic placement have been very little studied to date. The Tagalog regularities are not yet completely uncovered, and next to nothing is known about this topic in other Philippine-type languages (see Sneddon (1975: 238–246) for the basic rules in Tondano, which are quite different from those found in Tagalog).

Both peripheral and second position clitics are usually unstressed and form a prosodic unit with either the following word (proclitics) or the preceding word (enclitics). While the lack of stress and the concomitant need to attach to a phonological host are core features of clitics, second position clitics in western Austronesian languages present conflicting evidence in this regard. Most disyllabic clitics such as Tagalog *namán* and *námin* and some monosyllabic ones carry their own stress and thus could be considered independent words. But the fact that with regard to their position they clearly pattern with unstressed second position clitics strongly suggests an analysis as clitics. There is probably also evidence for a clitic analysis in the way the stress of the second position clitics interacts with the stress of the base, but this once again is a topic which has not yet been investigated.

Because of their positional variability, it is generally easy to distinguish second position clitics from affixes. This distinction, however, can be a problem in the case of peripheral clitics, especially proclitics. Thus, for example, the preposed pronouns *ku* and *kau* which are widely attested in **Malayic** constructions such as *buku ini sudah ku=baca* (book PRX already 1s=read) 'I already read this book' are sometimes analyzed as prefixes (*ku-*, *kau-*), and sometimes as proclitics (*ku=*, *kau=*). In many varieties, including Standard Indonesian, there appears to be no clear-cut evidence for preferring one or the other analysis (see also the discussion of cognate forms in **Javanese** and **Karo Batak**). In other instances, there are essentially three types of evidence for distinguishing clitics from affixes: (1) As opposed to affixes, clitics generally do not trigger morphonological alternations of the stem to which they are attached (this includes, for example, the fact that suffixes may cause stress shifts, while enclitics do not). (2) Clitics tend to be less selective than affixes with regard to the category of their hosts (for example, clitics may attach to nouns and verbs, while affixes usually are restricted to (a subclass of) either nouns or verbs). (3) Clitics are sometimes somewhat variable with regard to their position while affix order tends to be rigid.

The phonological attachment of a clitic does not have to match its morphosyntactic function. For example, in Central (or Guinaang) Bontok (northern Philippines) a number of phrase-marking clitics phonologically attach as enclitics to the preceding word while morphosyntactically they indicate the function of the following phrase (cf. Reid 1970, 1992). In example (9) this mismatch of phonological attachment and morphosyntactic function is illustrated by the general locative marker *as* which is usually reduced to just *=s* after vowel-final words (the square brackets indicate syntactic constituency).

(9) *in-manok* *nan* *babái=[$_{PP}$**s*** *nan* *masdem]*
AV-chicken SPEC woman=LOC SPEC night
'the woman performs a chicken sacrifice' [$_{PP}$ in the evening] (Reid 1970:23 and p. c.)

The orthographic representation of clitics differs widely across the area. Thus, for example, in the northern Philippines enclitics are usually not separated from their host (and are thus orthographically indistinguishable from suffixes), while in the central Philippines all

kinds of clitics are represented as independent orthographic words (compare the examples in the **Iloko** chapter with the ones in the **Tagalog** chapter; the indication of clitic boundaries in the Tagalog example (8) above does not conform to the standard orthography, which is otherwise adhered to throughout this chapter).

With regard to their function, the following major types of clitics are attested in western Austronesian languages:

- **Pronominal clitics**, which are widely attested throughout the area both as peripheral and as second position clitics (see section 3.7).
- **Aspectual/modal clitics**, which are also widely attested throughout the area. In Philippine-type languages they are usually second position clitics, while in the other languages they cluster around the verb as in the following example from Buru:

(10) *kami* ***la=ma=te=iko.***
1pe IRR=1p=ABLE=go
'we want to be able to go' (Grimes 1991:217)

- **Clitic particles**, which cover a wide range of interaction-related functions such as question and politeness markers, evidentials, etc. In Philippine-type languages these are often also second position clitics, while in the other languages they tend to be non-clitic particles, i.e. phonologically and morphosyntactically independent words which do not belong to a major word class.
- **Phrase-marking** and **determiner clitics** such as the Bontok locative preposition *as* (example (9)), the **Tagalog** phrase markers *ang*, *ng* and *sa* (see also section 3.6), definiteness marking *=nya* in **Colloquial Indonesian** and other **Malayic varieties** (see also section 3.7) or nominalizing proclitics such as Ratahan *to* in *tapi [**to** napók tee] nangule taa* (but NR AV.PST-CUT DIST AV.PST-return AND.DIR) 'but [those who split them] returned' (Himmelmann and Wolff 1999:34). Most western Austronesian languages have at least one clitic of this type but the distribution across the languages is very uneven and not easily amenable to generalizations.
- **Emphatic clitics** such as Standard Indonesian=*lah* or Buru=*an* which give various kinds of pragmatic prominence to the constituent to which they are attached. Compare the following Buru example:

(11) *tawe,* *yako=**an*** *naa* *te=keha* *moo.*
friend 1s=FOC PRX ABLE=ascend NEG
'Friend, it is me here who can't climb (the tree).' (Grimes 1991:193)

3.3 Clause types I: multiple basic verbal clauses

In most western Austronesian languages, there are a number of verbal clause structures which appear to be equally basic in that they do not seem to be derived from each other or to be clearly rankable with regard to some markedness metric. The best known and most widely discussed example of such multiple basic clause structures is the different voices in symmetrical voice languages to be discussed further below. But there are also other types of variation in basic clause structure for both transitive and intransitive clauses.

Here and in the remainder of this chapter, the terms *transitive* and *intransitive* refer to semantic transitivity, which does not necessarily match morphosyntactic transitivity. A mismatch between the two types of transitivity is found, for example, in English passives

such as *I was hit by the guy behind me* where a morphosyntactically intransitive predicate (*be hit*) denotes a transitive state of affairs (see also section 3.4, Clynes 1995:189f, 297–300 and Van Valin and LaPolla 1997:147). Similarly, 'verbal clause' is to be understood in purely semantic terms, i.e. a clause which denotes an event ('going', 'falling', 'digging', 'throwing', etc.). It is of no concern whether the predicate of such a clause is a verb in morphosyntactic terms.

Furthermore, it will be convenient to make use of the widely used abbreviations S for the single core argument of an intransitive clause, A for the more actor-like core argument of a transitive clause, and O for the more undergoer-like argument of a transitive clause. Note that S, A, and O here are convenient abbreviations for referring to the core arguments of verbal predicates and not syntactic-semantic relations of some kind (as in Dixon 1994:6 passim).

With regard to intransitive clauses, there is a phenomenon that has been called split intransitivity, i.e. the availability of two (or more) basic constructional patterns for intransitive verbal clauses in a given language. The most typical split pertains to (semantic) control and volitionality (cf. Mithun 1991). Thus, for example, in Acehnese the S argument is obligatorily cross-referenced by a proclitic if it is conceived of as being in full control of the action denoted by the predicate, as in (12) (the full pronoun *gopnyan* is optional).

(12) *(gopnyan)* ***geu****=jak* '(s)he goes' (Durie 1987:370)
3s 3s=go

If the S argument is conceived of as undergoing rather than controlling the action (cf. Durie 1985:55–71), it is *optionally* cross-referenced by an *en*clitic, as in:

(13) *(gopnyan)* *rhët (=geuh)* '(s)he falls' (Durie 1987:369)
3s fall (=3s)

These coding properties mirror the properties of the A and O arguments of transitive clauses, where the A argument (*lôn* in (14)) is also obligatorily cross-referenced by a proclitic while the O argument may optionally be cross-referenced by an enclitic.

(14) *(gopnyan)* *ka* ***lôn****=ngieng (=geuh)*
3s CPL 1s=see (=3s)
'I saw him/her.' (Durie 1987:369)

Durie (1987) argues that these similarities between A and S_a arguments (i.e. those S arguments which share their cross-referencing properties with A arguments) on the one hand and O and S_o arguments on the other are not just superficial similarities but rather reflect the fact that Acehnese grammar is basically organized around the macro-roles of Actor and Undergoer. He claims that arguments with the same macro-role exhibit the same set of coding and behavioral properties, regardless of whether they occur in transitive or intransitve constructions (for a further illustration of relevant behavioral properties, see section 3.8.1). On this account, Acehnese is a Split-S language (cf. Dixon 1994:70 passim; Dixon's further distinction between Split- and Fluid-S languages is ignored here). Other possible examples for split-S languages in the area are Dobel (Hughes 2000:147f), Selaru, and **Mori Bawah**, where syntactic factors (rather than control and volitionality) play a major role in determining the choice of a given intransitive construction.

Not all split-intransitive languages are also split-S languages in the strict sense. That is, there may be no single basic intransitive verbal clause type in a given language, but the structural differences between the two (or more) basic intransitive clauses do not have

to align with distinctions in transitive clauses in a straightforward way. **Kambera** has been described in this way. **Taba** is another possible candidate since here clear-cut alignments are complicated by the fact that O arguments also appear to be split into two basic types (see below and the detailed discussion in Bowden 2001:154–166).

Arka (1998:50–65) analyzes Balinese as a split-intransitive language based on the fact that intransitive predicates differ with regard to their morphological marking: some are prefixed with *N-* or *ma-*, and others remain unmarked. This difference in morphological marking resembles the kind of morphological marking found on transitive predicates. In undergoer voice, the predicate is morphologically unmarked; in actor voice it is usually prefixed with *N-* (compare example (15)b below). Note, however, that the fact that transitive predicates are never marked with *ma-*, and that therefore the proposed alignment between the intransitive and transitive system is only partial, remains unexplained. Furthermore, it is unclear whether the differences in morphological marking correlate with differences in clause structure.

In a similar spirit, Donohue (1999:482–484) proposes to analyze Tukang Besi as a split-intransitive language based on the fact that intransitive verbs fall into different classes with regard to their morphological potential: only some intransitive verbs may occur with the comitative applicative suffix *-ngkene*, while others may occur with factitive causative *hoko-*, etc. Grimes (1991:99 passim) makes a similar point for Buru. In both instances, however, there appears to be only a single basic intransitive clause structure in which S precedes V (in Tukang Besi, S is also cross-referenced by a pronominal prefix on V).

In most of the world's languages, dynamic and stative intransitives in all likelihood differ with regard to at least one morphological or syntactic property. It is therefore not quite clear whether it is useful to expand the notion of split-intransitivity in line with the proposals just sketched for Balinese, Tukang Besi and Buru, where differences between intransitive predicates primarily pertain to morphology (and semantics). There is, however, no doubt about the fact that the distinction between dynamic and stative predicates is of fundamental importance to the grammar of most western Austronesian languages, as further discussed in section 4.1. If the notion of split-intransitivity is applied to instances of differing morphological marking or potential, then all these languages are split-intransitive. Otherwise, split-intransitivity which is clearly manifest on the level of clause structure would appear to be found only in a number of transitional and preposed possessor languages, a (probably relatively small) subset of these being Split-S languages in the strict sense.

Turning now to transitive clauses, symmetrical voice systems provide one extremely common example of multiple basic transitive clauses in that for any transitive event there are at least two representations, one in actor voice and one in undergoer voice. This was already illustrated with the Standard Indonesian examples (1) and (2) above. Here is another example pair from Balinese:

(15) (a) *bawi-ne* *punika* *tumbas* *tiang.* UNDERGOER VOICE
pig-DET DIST UV:buy 1
(b) *tiang* *numbas* *bawi-ne* *punika.* ACTOR VOICE
1 AV:buy pig-DET DIST
'I bought the pig.' (Arka 1998:10)

As already mentioned in section 1.2, it is a matter of controversy whether actor and undergoer voice clauses in such pairs are in fact equally basic. For Balinese and some other symmetrical voice languages, it has been argued that the undergoer voice clause is in fact the more basic one. Thus, for example, the undergoer voice verb form *tumbas* in example (a) is morphologically unmarked, while the actor voice form consists of the

prefix *N-* plus the base *tumbas*. Note that such an argument does not hold in the case of the Standard Indonesian examples (1) and (2) because in Standard Indonesian both actor and undergoer voice forms are usually prefixed.

Other arguments that are usually invoked in order to show that undergoer voice clauses are more basic or less marked than actor voice clauses include claims that they are more frequent in discourse, that they occur in a wider range of discourse contexts, often being the required construction in a given context, or that they are acquired earlier in language acquisition. We will have the opportunity to look at some of these arguments in more detail in sections 3.8 and 5. Here it will suffice to note that none of these arguments appears to be without problems and that all contributions on symmetrical voice languages in this volume consider the relevant alternations to instantiate different, but equally basic transitive constructions.

Apart from symmetrical voice alternations, other types of multiple basic transitive clause constructions are attested in western Austronesian languages, usually involving some kind of pronominal marking. One type is found in Standard Indonesian and many other **Malayic varieties** (except **Old Malay**) as well as in most symmetrical voice languages of western Indonesia (prominent exceptions are Balinese and Sundanese). In Standard Indonesian, there is a third basic transitive construction type next to the two constructions exemplified in (1) and (2) above. In it, a pronominal actor occurs immediately before the (unaffixed) verb:

(16) *orang itu* ***ku***=*lihat*.
person DIST 1s.ACT=see
'I saw that person.'

It is widely agreed that in this construction the undergoer (*orang itu*) is the subject and the actor pronominal (*ku*=) is a non-subject core argument and that therefore the overall construction is equally transitive as the actor voice construction *anak saya melihat orang itu* in (1) (see Musgrave 2001 for references and discussion). Once again, however, it is a matter of debate whether the constructions are equally basic. Most important in this regard is the fact that the actor+unaffixed verb construction is restricted to pronominal actors and terms of address, including kin terms (*bapak* 'father, Mr.') and personal names (see also **Javanese** and **Karo Batak**).

A somewhat different type of multiple basic transitive clause construction appears to exist in Tukang Besi (cf. Donohue 1999:51–54, 2002). In one type of transitive construction, the A argument is marked with the so-called nominative marker *na* and cross-referenced by a verbal prefix. The O argument is preceded by the so-called core argument marker *te* and not cross-referenced on the verb (schematically: A-V *te* O *na* A):

(17) *no-kiki'i te* *iko'o* *na* *beka*
3.RLS-bite CORE 2s NOM cat
'The cat bit you.' (Donohue 1999:53)

In a second type of transitive construction, the O argument is cross-referenced by an enclitic on the verb and is also marked with *na*. The A argument is still cross-referenced by a verbal prefix but now it is marked with the so-called core argument marker *te* (schematically: A-V=O **na** O *te* A).

(18) *no-kiki'i=ko* *na* *iko'o* *te* *beka*
3.RLS-bite=2s.OBJ NOM 2s CORE cat
'The cat bit you.' (Donohue 1999:53)

Donohue (1999:53f) suggests that despite its greater morphological complexity, this second construction is the more basic one of the two constructions because it occurs more frequently in texts and also has a somewhat wider distribution. At the same time, he considers this alternation a symmetrical voice alternation, the construction with an O-enclitic (example (18)) corresponding to an undergoer voice construction, and the one without it (example (17)) to an actor voice construction (Donohue 1999:160–164). While there are some conspicuous similarities with Philippine-type voice alternations, the fact that there are no voice-marking affixes involved makes it doubtful whether this is really best analyzed as a *voice* alternation (rather than some other kind of symmetrical alternation between transitive construction types).

Note that the specific interaction of person markers and nominal phrase markers illustrated by examples (17) and (18) appears to be attested only in Tukang Besi. But in neighboring Muna a somewhat similar alternation in person marking occurs, which is called *definiteness shift* by Van den Berg (1989: 59–66). In this alternation, (mostly) transitive verbs change their subject class prefix when occurring with a definite object. In the following examples, the two subject prefix classes are simply glossed as I and II respectively:

(19) (a) ***ne-rabu*** *nuhua*
3sI.RLS-make pitcher
'She is making a pitcher.'

(b) ***no-rabu-e***
3sII.RLS-make-3s.OBJ
'She is making it.' (Van den Berg 1989: 59f)

As this example shows, the different subject marker is triggered inter alia by the presence of an object marking suffix.

Furthermore, other characteristic features of Tukang Besi, including the person markers and the occurrence of a clear passive construction (e.g. *'u-to-kiki'i na iko'o* (2s.RLS-PASS-bite NOM 2s) 'you were bitten' (Donohue 1999:53)), are frequently found in other transitional languages of Sulawesi such as **Mori Bawah** and **Makassar**. For this reason, Tukang Besi is considered a transitional language here rather than a symmetrical voice language. Nevertheless, the example of Tukang Besi shows that the occurrence of multiple transitive constructions is not restricted to symmetrical voice languages.

Finally, it should be mentioned that in a number of transitional and preposed possessor languages, a distinction is made between transitive and semi-transitive clause types which possibly also constitutes an example for multiple transitive constructions. Bowden (this volume) correlates the distinction between transitive and semi-transitive constructions in **Taba** with two types of O arguments, direct and remote undergoers, which differ in that remote undergoers may be optionally marked with an adposition. Inasmuch as this can be usefully analyzed as a split-O system (Bowden 2001:164–166), it would instantiate another type of multiple transitive constructions.

3.4 Clause types II: existential, possessive and equational clauses

It is a common feature of most western Austronesian languages that there are at least one or two non-verbal clause types which play a major role in grammar and discourse. These are, on the one hand, existential clause constructions which often also form the basis for possessive constructions, and on the other hand, equational (equative) clause constructions which play a role in cleft (emphatic focus) and question formation.

The most common existential construction in western Austronesian languages consists of an existential particle or verb which is immediately followed by its

complement. Existential particles are unaffixed and thus differ clearly from verbs. Existential verbs usually differ from other verbs by combining only with a small subset of the verbal morphology available in the language (see **Nias**, **Mori Bawa**, **Leti** and **Kambera** for examples). The following Cebuano example illustrates an existential particle:

(20) ***may*** *tulú* *ka* *tawu-ng* *nangita?* *nimu.*
EXIST three LK person-LK AV.RLS:see 2S.DAT
'There were three people looking for you.' (Wolff 1972:679)

This example also illustrates a very typical use of existential constructions, i.e. as presentative constructions introducing a new participant into the universe of discourse. In this use, the complement is often further expanded by a relative construction (as in (20)). The other major function of existentials is to indicate availability (or, in negated form, non-availability), as in Cebuano *may sigarilyu* 'there are cigarettes/we have cigarettes'.

The following three features of existential constructions appear to be restricted to symmetrical voice languages, in particular Philippine-type languages. First, in these languages it is possible to combine an existential particle directly with a voice and aspect-mood-marked form, as in Cebuano:

(21) *may* ***miinom*** *ug* *tubig* *sa* *kusina.*
EXIST AV.PST:drink OBL.NSPEC water OBL kitchen
'Someone drank water in the kitchen.'

This use of voice and aspect-mood-marked forms is just another indication of the lack of distributional differences between nouns and verbs characterizing these languages (see section 3.1 above).

Second, numerals and other quantifiers may be used in the function of an existential operator as in **Tagalog** *ma-rami-ng tao-ng na-matáy doón* (ST-amount-LK person-LK RLS.ST-dead DIST.LOC) 'there were many people who died there' (further examples in the **Seediq** chapter).

Third, it is very common that negative existentials are expressed by a negative existential particle which is not in any obvious way derived from the positive one. In Cebuano this is *walá?*:

(22) ***walá?*** *na=y* *tubig*
NEG.EXIST CPL=LK water
'There is no more water.' (Wolff 1972:1122)

The alternative is illustrated by **Malayic varieties** where existential *ada* is negated by the common verbal negator *tidak* (e.g. *tidak ada air* 'there is no water'). This is also the standard pattern attested in transitional and preposed possessor languages.

In a few languages, the element functioning as existential operator also functions as a locative preposition, as in Tetun Dili *iha foos iha ka'ut* (EXIST rice LOC sack) 'there is rice in the sack' (Hull and Eccles 2001:99; see also Muna *bhe*, Van den Berg 1989:160f). **Taba** is claimed to have no native existential expression. And in **Biak** there is an existential verb (based on a deictic root) which appears to be a full verb.

The existential construction also widely functions as the predicate in a clausal possessive construction, to which then another NP denoting the possessor is added. The NP denoting the possessor may occur in two grammatical functions. It may be the subject of the overall construction as in the following examples from Tetun Dili and Cebuano respectively (in (24) the subject pronoun *ka* is a second position clitic):

(23) ***ami*** *iha* *telemovel* *ida*
1pi EXIST mobile phone one
'We have a mobile phone.' (Hull and Eccles 2001:100)

(24) *may* *sigarilyu* ***ka*** *dihá??*
EXIST cigarettes 2S MED.LOC.PST
'Do you have any cigarettes on you?' (Wolff 1972:679)

Alternatively the possessor NP may be expressed as a genitive attribute to the possessum (so that the literal meaning of the overall construction is something like *X's Y exists*). Compare the following Muna example:

(25) *miina* *bhe* *doi-**ku***
NEG EXIST money-1S.POSS
'I do not have any money.' (Van den Berg 1989:161)

Possessive constructions built from a more basic existential construction are by far the most common strategy for possessive clauses throughout the area. The major alternative is an equational construction in which the possessum occurs in subject position and is ascribed to the possessor which functions as the predicate of the overall construction. In Tetun Dili, the possessor occurs in a special form, marked with the possessive suffix *-nian*:

(26) *ne'e* *sira-nian,* *la'ós* *ami-nian.*
PRX 3P-POSS NEG 1PI-POSS
'This is theirs, not ours.' (Hull and Eccles 2001:34)

In most languages which allow such a construction, however, the possessor predicate is marked as a dative or locative phrase, as in Tagalog *sa nanay ang relos* (LOC mother SPEC watch) 'the watch belongs to mother' (Schachter and Otanes 1972:273).

Full verbs meaning 'have', 'own' or 'belong' are only sporadically attested. Where they exist, they tend to be used infrequently (major exceptions include Balinese where the most common possessive clause type involves the verb *ngelah* 'possess, own'). See **Taba** for a very unusual derived possessive verb based on possessive pronominals.

The Tetun Dili and Tagalog examples just mentioned also illustrate the basic pattern for equational clauses. These consist of a simple juxtaposition of a subject and a predicate phrase. There is no copula in most western Austronesian languages (major exceptions are West Papuan languages such as Ambai and **Biak**). In Philippine-type languages the basic order in equational clauses tends to be PREDICATE–SUBJECT while in most other languages, including many symmetrical voice languages in Indonesia, the order is SUBJECT–PREDICATE. Typical predicates are simple nouns (Madurese *Siti ghuru* 'Siti is a teacher') or prepositional phrases (Madurese *Buku-na nəng meja* (book-DET at table) 'the book is on the table' (Davies 1999a:26)).

The equational clause format is often used in clefts and content questions. In questions, the question word usually becomes the predicate while the remainder of the clause functions as the subject, often in the form of a headless relative construction or some other kind of nominalization. Compare the following example from Tukang Besi:

(27) *te* *emai* *na* *'umelo-'elo-aku* *iso?*
CORE who NOM REL: RDP-call-1S.OBJ DIST
'Who is it that's calling me there?' (Donohue 1999:57)

In Tukang Besi and most other languages, this strategy is either optional or restricted to a subset of questions. In Tukang Besi, for example, it is obligatory only for subject questions (Donohue 1999:451f). In these languages, there is usually also a question-forming strategy where the question word remains in situ. **Biak** is somewhat exceptional in requiring the fronting of question words without imposing an equational structure.

In many Philippine-type languages, the equational clause-strategy is obligatory for clefts and content questions (at least for core roles; there are exceptions, in particular in Taiwan, e.g. Pazeh). In fact, starting with Bloomfield (1917) it has repeatedly been suggested that the basic clause structure in these languages is equational (see, for example, Scheerer 1924, Lopez 1937, Capell 1964, Lemaréchal 1991, Naylor 1995, Egerod 1988, DeWolf 1988, Himmelmann 1991). That is, the structure of verbal (or narrative) clauses is said to be essentially identical to that of equational clauses. Schachter and Otanes (1972:62) hint at this possibility with the following observation:

> It may, in fact, quite reasonably be argued that the distinction made above between equational and narrational sentences in Tagalog is a somewhat arbitrary one, and that all Tagalog basic sentences, including those here treated as narrational, are essentially equational in nature, involving a balancing of two elements – the predicate and the topic [i.e. subject, NPH] – against one another.

Compare the following Tagalog examples (based on Schachter and Otanes 1972:61f), where the (a) examples are generally considered equational, while the (b) examples are generally considered verbal:

(28) (a) *artista ang babae*
artist SPEC woman
'The woman is an artist.'

(a′) *babae ang artista*
woman SPEC artist
'The artist is a woman.'

(b) *y<um>aman ang babae*
<AV>riches SPEC woman
'The woman got rich.'

(b′) *babae ang y<um>aman.*
woman SPEC <AV>riches
'The one who got rich is a woman.'

The only difference between these sentences is that in the (b) examples there is a voice and aspect/mood-marked word (*yumaman*). But it is questionable whether voice and aspect/mood marking has any clause-structural consequences. Thus, for example, voice and aspect/mood-marked words like other content words may appear in predicate as well as in subject position without any concomitant morphosyntactic changes. Furthermore, there are many ways of rearranging the content words in these examples (for example, one could topicalize *babae* as in *ang babae ay artista/yumaman*), but there is not a single alternative pattern which would clearly distinguish the (a) from the (b) examples. Note that essentially the same possibilities hold for semantically transitive expressions:

(29) (a) *asawa ko siyá*
spouse 1s.POSS 3s
(a′) siyá ang asawa ko
(a″) siyá ang aking asawa
'She is my wife/he is my husband.'

(b) *m<in>ura ko siyá*
<RLS(UG)>scolding 1s.POSS 3s
(b′) siyá ang minura ko
(b″) siyá ang aking minura
I scolded him/her.

The equational clause hypothesis rests on the fact that essentially the same set of pronominal forms and phrase makers are used for arguments and adjuncts regardless of whether the predicate is (semantically) nominal or verbal. There are also no particles, negators or other kinds of grammatical markers which would clearly distinguish between a verbal and an equational clause type.

Nevertheless, there are two major empirical problems for this hypothesis. First, it is unclear how to account for control constructions and other kinds of multipredicate constructions under this hypothesis. In particular, the fact that in some languages (not standard Tagalog) verbal subjunctive forms occur in these constructions would appear to be not easily explainable within this framework (see sections 3.9 and 4.2.2). Second, there are in fact some minor differences in the marking of adnominal and verbal arguments, as is further discussed in section 3.6.

To date, these two problems have not yet been explicitly addressed by the proponents of the equational hypothesis. Hence it is unclear whether and in which form this hypothesis can be upheld. But in line with the quote from Schachter and Otanes (1972) given above, it should be obvious that verbal and equational clauses are very similar indeed in many Philippine-type languages and that the degree of their similarity constitutes an important parameter for both the internal and external typology of western Austronesian languages.

Very roughly, western Austronesian languages fall into three major types of languages with regard to this parameter: (1) languages where the two clause types are so similar that their distinction may be questioned (many symmetrical voice languages of the Philippine type); (2) languages where the distinction is not in doubt but where there are still quite a number of important similarities (the remaining symmetrical voice languages and some transitional languages); and (3) languages where the distinction between the two clause types is very clear and little overlap exists (some transitional languages and most preposed possessor languages).

3.5 Word order and constituency

Western Austronesian languages differ with regard to preferred basic word order and the strictness of ordering relations. Symmetrical voice and transitional languages are either predicate-initial, with a tendency to be also subject-final (i.e. 'VXS'), or favor a subject-predicate ('SVO') order. In most of these languages, however, there is some word order flexibility and more often than not it is unclear whether the so-called basic or unmarked order reflects a syntactic constraint or a pragmatic preference (see also Cumming 1991). Preposed possessor languages tend to follow an SVO pattern and to be somewhat more rigorous in adhering to this basic pattern.

The preceding generalizations are easily challengeable on methodological and empirical grounds. There is no need to repeat here the well-known methodological problems of statements on basic word order (see LaPolla and Poa (to appear), for a recent summary). Empirically, we may note that in symmetrical voice languages, for example, statistically manifest preferences vary according to voice type. For Balinese, Pastika (1999, chapter 6) finds that in more than 90% of the actor voice clauses in his corpus, the subject (=actor) NP precedes the predicate. In undergoer voice clauses, on the other hand, there is no such clear-cut preference, undergoer voice subjects being equally likely to precede or follow the predicate (see also Artawa *et al.* 2001).

Other word order generalizations are empirically much more robust. Adpositions are generally prepositions in western Austronesian languages (some of which may not have any adpositions at all, see section 3.6), although the odd postposition is also sporadically attested (e.g. **Karo Batak**, **Taba**). Auxiliaries generally precede main verbs. Negators also generally precede the negated constituent, with the exception of most preposed

possessor languages, where negators usually occur in clause-final position (not in **Tetun Fehan** and only in part in **Leti**). Possessors generally follow the possessum, except of course in preposed possessor languages where non-pronominal possessors precede the possessum (see section 3.11). For (cardinal) numbers and other quantifiers the converse tendencies hold: they generally follow the head noun in preposed possessor languages, but precede it in most symmetrical voice and transitional languages. Otherwise, adnominal modifiers generally follow the head, with demonstratives being placed at the very end of an NP. In Philippine-type languages, however, the order of constituents in noun phrases is highly flexible and there are no clear-cut ordering rules, except that cardinal numbers tend to be placed at the beginning of an NP and genitive-marked possessors have to follow the possessum.

It has occasionally been claimed in the literature that some western Austronesian languages are free word order languages, sometimes even allowing every possible commutation of a verbal predicate and its core arguments (Uhlenbeck 1975 is an example). This is almost certainly an overstatement of the facts. To begin with, no western Austronesian language allows the constituents of noun phrases or prepositional phrases to be distributed discontinuously across the clause as is found in Latin or in Australian languages. Hence, if anything, western Austronesian languages could be free phrase order languages.

But even phrase order is probably not completely free in any of these languages. As shown by Davies (1999b) for Madurese and Javanese and Arka (1998:119–182) for Balinese, apparently free phrase ordering options are structurally constrained and usually intonationally marked in such a way that they clearly instantiate topic, cleft, or rightward expansion constructions. That is, for core constituents the phrase ordering options in a basic unmarked clause are usually restricted, allowing at most for one or two alternations. As in many other languages, the placement of adjuncts (including adverbs) tends to be somewhat less restricted. The alternation between absolute vs. oblique forms in Kerinci (**Malayic varieties**, section 5) provides strong historical evidence for the relevance of phrasal boundaries on various levels.

In many western Austronesian languages there is, in fact, good evidence for a VP constituent which contains the predicate and all non-subject core arguments. Non-subject core arguments generally have to occur in immediate post-predicate position. That is, neither the subject nor adjuncts may intervene in between the predicate and its non-subject arguments (exceptions include postverbal clitics and sometimes a restricted set of adverbs). Compare the following example from Totoli (see also Donohue (1999:151 passim) for Tukang Besi and Artawa *et al.* (2001:15) for Balinese):

(30) *gaukan* *[$_{VP}$no-gutu* *ponguman* *itu]*
king AV.RLS-make story DIST
(Yesterday) the king made this announcement: ...
(a) *nogutu **gaukan** ponguman itu
(b) ***ponguman itu** nogutu gaukan

As the (a) example shows, it is not possible to place the subject (*gaukan*) in between the verb (*nogutu*) and the non-subject argument (*ponguman itu*). Furthermore, it is not possible to exchange the positions of subject and non-subject argument, as seen in (b). It is, in principle, possible to place the non-subject argument in clause-initial position (i.e. *ponguman itu, gaukan nogutu*). But this is clearly a topicalization construction in which the non-subject argument occurs outside the clause core and, among other things, forms a prosodic unit of its own (as indicated by the comma).

In symmetrical voice languages, the same kind of evidence points to the fact that in undergoer voice clauses the verb and the non-subject actor argument form a kind of VP constituent, as seen in the following example from Totoli:

(31) *kopi* *ia* *[$_{VP}$ni-pogutu* *i* *Andris].*
coffee PRX RLS(UG)-make PN Andrew
'Andrew made this coffee.'

Once again, subjects (*kopi ia*) or adjuncts cannot intervene in between verb (*nipogutu*) and non-subject actor argument (*i Andris*). The topicalization of the non-subject actor argument (i.e. placing *i Andris* in clause-initial position) is strongly dispreferred by speakers and not attested in spontaneous data.

It is a matter for further debate whether the 'VP' constituents in (30) and (31) are indeed sufficiently similar to warrant the same label. In fact, the nature of clausal constituents in western Austronesian languages has been very little studied to date and there are almost certainly many details still to be uncovered (useful but far from complete discussion can be found in Kroeger 1993:118–166, Arka 1998, Davies 1999b, and Musgrave 2001). Two problems, however, are already reasonably clear at this point.

First, while there is good evidence that the predicate and non-subject arguments form a constituent in most western Austronesian languages, it is also clear that the evidence varies with the nature of the non-subject argument. Thus, as just mentioned, non-subject undergoer arguments in actor voice constructions (such as *ponguman itu* in (30)) can usually be topicalized without any problems, but non-subject actor arguments in undergoer voice constructions (*i Andris* in (31)) cannot (see also **Karo Batak**). Furthermore, Arka (1998:124f) claims that in Balinese there is a significant difference between indefinite or non-referential and definite non-subject undergoer arguments. Only definite non-subject undergoer arguments (such as *ia* in *cang ng-runguang ia ditu* (1 AV-care.for 3 DIST.ADV) 'I cared for him/her there') can be topicalized and allow the insertion of certain adverbs after the verb.

Second, for some Philippine-type languages it is occasionally claimed that the order of arguments in post-predicate position is essentially free. In particular, the subject is said to be allowed to intervene in between predicate and non-subject arguments, resulting in a VSX order as in the following widely used Tagalog example:

(32) *b<um>ili* ***ang*** ***babae*** *ng* *tinapay* *sa* *tindahan* *para* *sa* *bata'*
<AV>buy SPEC woman GEN bread LOC store for LOC child
'The woman bought some bread at the store for the child.'

While sentences such as these are judged acceptable by some native speakers, they do not occur in natural discourse and are thus of questionable value. There are two problems with these examples. First, in Tagalog natural discourse, as probably in most languages, clauses containing more than one full NP are rare (see DuBois 1987). Second, and more importantly, full (non-pronominal) subject NPs never precede genitive-marked arguments and usually also follow locatives and benefactives (i.e. in natural discourse *ang babae* in (32) would always follow *ng tinapay* and in most instances in fact occur at the very end of the clause).

For Tagalog, there is one type of exception to this claim. Subject expressions consisting of a (short) personal name, which are marked by *si* rather than by *ang*, sometimes

occur in immediate post-predicate position:

(33) *nag-pa-sundó'* ***si Andrés*** *ng* *isa-ng* *pare'*
RLS.AV-CAU-fetch PN GEN one-LK priest
'Andrés sent for a priest' (Bloomfield 1917:92)

Bloomfield calls this ordering *enclitic positioning* (1917:153) since in this example the proper noun occupies the clitic position that is usually occupied by second position clitics (see section 3.2). This analysis implies that there are severe restrictions on the placement of non-pronominal subject expressions before other argument expressions. So far, however, the nature of these restrictions has not yet been investigated in any Philippine-type language.

3.6 Case marking, adpositional phrases and the core vs. peripheral distinction

Western Austronesian languages are generally not case-marking languages. No western Austronesian language has case affixes. However, in a very few instances alternations exist which are similar to affixal case alternations. These include the alternation between mutated and unmutated forms in **Nias** and the absolute vs. oblique alternation in Kerinci (cf. ADELAAR, MALAYIC VARIETIES). Furthermore, many symmetrical voice languages and a few transitional languages have paradigms of phrase-marking clitics which are often called case markers but which usually also convey specificity, definiteness or even deictic distinctions.

Before looking more closely at these phrase-marking clitics, a few remarks on adpositions are in order. Since there are only very few postpositions in western Austronesian languages (see section 3.5 above), the discussion here will be limited to prepositions. Compared to Indo-European languages, the inventory of primary (or simple) prepositions tends to be fairly small. A typical inventory contains a smallish number of local prepositions (e.g. Standard Indonesian static locative *di* 'in, on, at', etc., allative *ke* 'to', ablative *dari* 'from') and a comitative-instrumental preposition (e.g. Standard Indonesian *dengan* 'with'). It is not uncommon that a single primary preposition can be used for stative locative as well as source and goal relations (e.g. **Kambera** *la*, Muna *we*). Local prepositions may usually also be used for temporal relations (**Kambera** *la mbaru* 'in the morning'). Further additions to the basic inventory tend to have fairly specific meanings and to be attested only in a few languages (Standard Indonesian, for example, also has *untuk* 'for', *oleh* (for agents in passives), and *tentang* 'about, concerning').

More specific local relations are usually expressed with the help of relational expressions for body parts ('head', 'back', etc.) or relational object parts ('front', 'centre', etc.). These relational expressions are often combined with a locative primary preposition as in Standard Indonesian *di atas* 'above', *di bawah* 'beneath', or *di muka* 'in front'. Another strategy is the use of deictic or directional particles in prepositional functions. Thus, for example, Buru proximal *na(a)* may not only be used as a pronoun or adnominal modifier but also as a locative preposition. In the following example, it occurs in both functions (the short form *na* is conditioned by the non-final position).

(34) *da* *kaduk* *na* *huma* *naa.*
3s come PRX house PRX
'He came here to this house.' (Grimes 1991:172)

In fact, much of what Lichtenberk (1991) observes for directional and prepositional elements in Oceanic languages also holds for western Austronesian languages, especially preposed possessor languages.

TABLE 5.2: CEBUANO AND TAGALOG NON-PERSONAL PHRASE MARKERS

	Cebuano	Tagalog	
SPECIFIC (ARTICLE)	ang	ang	SPECIFIC (ARTICLE)
OBLIQUE SPECIFIC	sa	ng [naŋ]	GENITIVE
		sa	LOCATIVE
OBLIQUE NON-SPECIFIC	ug	ng/LINKER	

TABLE 5.3: CEBUANO AND TAGALOG PERSONAL NAME MARKERS

	Cebuano	Tagalog
(NOMINATIVE)	si	si
POSSESSIVE	ni	ni
DATIVE	kang	kay

Apart from primary prepositions and complex prepositions consisting of a primary preposition and another relational expression, many western Austronesian languages also show weakly grammaticized prepositions, i.e. words which may be used as content words (usually verbs) and as prepositional function words. Standard Indonesian *sampai*, for example, functions both as a preposition 'until, as far as' (as in *sampai sekarang* 'until now') and as a verb 'arrive, reach' (as in *pukul enam pagi kami sampai* (strike six morning 1pe arrive) 'we arrived at six a.m.'). See also **Belait** and **Tetun Fehan**.

Philippine-type languages (and Tsou and Rukai) are languages with absolutely minimal inventories of primary prepositions. In fact, several of these languages have been analyzed as having no primary preposition at all (e.g. **Tsou**, see also **Kimaragang**). Such an assessment depends very much on the analysis of the phrase-marking clitics which occur in most of these languages. These clitics usually come in two paradigms, one for personal names and one for all other kinds of nominal expressions. Table 5.2 and Table 5.3 give the paradigms for two closely-related Meso-Philippine languages, Cebuano and **Tagalog**.

The personal name markers are always obligatory. Considerable variation exists with regard to the obligatoriness of the non-personal phrase marking clitics. In many languages, including Cebuano and Tagalog, the specific article *ang* may be omitted in a number of contexts, for example when the nominal expression occurs in topic position preceding the predicate. The other non-personal phrase markers generally cannot be omitted in Cebuano and Tagalog. In Pazeh, however, most non-personal phrase markers are said to be optional (Li and Tsuchida 2001:31).

Perhaps the most important fact to note about the paradigms given in Table 5.2 is that the distribution of the non-personal markers differs significantly despite the great formal similarities of the markers and the relative closeness of the two languages in geographical as well as typological terms (see **Seediq**, **Tsou**, **Iloko**, **Kimaragang**, and **Buol** for examples of further variation). Their distribution is roughly as follows: the unmarked specific form is used in both languages for subjects, topics and predicates of identificational clauses (e.g. Cebuano *si Ana ang guapa* 'the pretty one is Ana'). The specific oblique form *sa* in Cebuano is used in all other functions, including possessors (e.g. *nanay sa bata?* (mother OBL child) 'the child's mother'), adjuncts (*sa kusina* in (21)), actors in undergoer voice clauses and undergoers in actor voice clauses. The non-specific oblique form is used for obliques the referential identity of which is unknown or irrelevant (e.g. *anák ug hari?* (offspring OBL.NSPEC king) 'a king's offspring'). With regard to

adjuncts, it mostly occurs with expressions corresponding to English manner adverbials, as in *milakáw siyá ug kusúg* (PST.AV:walk.away 3S OBL.NSPEC fast) 'she walked away fast' (Wolff 1972:1077).

In Tagalog, specificity plays a less prominent role in the distribution of the phrase markers. For possessors and actors in undergoer voice constructions only genitive *ng* is used and most kinds of adjuncts are marked by *sa*. The marking of undergoers in actor voice constructions is split among locative *sa* and genitive *ng*, animate and definite undergoers usually being marked by *sa*, and all others by *ng* (see below and **Tagalog** for further details and examples). Thus, there is no single marker corresponding to Cebuano *ug* which in Tagalog is sometimes rendered by *ng*, sometimes by the linker *na*.

The distribution of the personal name markers differs in some important details from those of the non-personal markers and is essentially identical across both languages (and most other Philippine-type languages). Possessive *ni* is used only for possessors and actors in undergoer voice clauses, never for undergoers in actor voice clauses or adjuncts. Dative *kang/kay* is used for recipients, addressees and other core undergoer roles in actor voice clauses. If personal names are used as adjuncts – not a frequent occurrence – the non-personal phrase marker *sa* precedes the *kang/kay*-marked personal name, as in Tagalog:

(35) *t<um>anggáp* *sila* *ng* *isá-ng* *gawáng* *ka-bait-an*
<AV>received 3P GEN one-LK made:LK NR-kindness-NR
sa ***kay*** *Maria*
LOC DAT.PN Maria
'(there were not any who could say that) they had received any kindness from Maria.' (Bloomfield 1917:76)

From a cross-linguistic point of view, the distribution of the non-personal markers in particular is somewhat unusual. To date, there is no standard analysis and terminology in use for these forms in the literature on western Austronesian languages. A recent survey by Reid (2002:286f) shows that more than two dozen labels have been used in reference to them, including *articles/determiners*, *prepositions*, *case markers* and *relation markers*. Note that this lack of terminological agreement also holds for individual markers in individual languages. Thus, for example, Tagalog *ang* has been glossed 'nominative', 'absolutive', 'specific', 'subject', 'topic', 'trigger', etc. Consequently, the labels used in Table 5.2 and Table 5.3 are not standard in any sense.

The analysis of the phrase markers is very closely related to the analysis of other aspects of Philippine-type morphosyntax, in particular basic clause structure, grammatical relations and voice alternations, as will be evident from discussion in the relevant sections (3.3, 3.8, 4.2). Regardless of the choices determined by these aspects, however, it is quite clear that the phrase markers have characteristics of both determiners ('articles') and prepositions. The split between non-personal and personal markers is typical for determiners. Note also that *ang*, Tagalog *ng* and in some uses Cebuano *sa* can be replaced by demonstratives (e.g. Tagalog *sundalo ng sultan* 'the sultan's soldiers' can also be rendered by *sundalo nitóng sultan* (soldier GEN.PRX sultan)). In some languages (e.g. **Tsou**), in fact, all phrase markers also convey deictic distinctions.

On the other hand, Tagalog *sa* (and to a lesser degree Cebuano *sa*) is preposition-like not only because of its uses for adjuncts but also because it is a (usually obligatory) constituent of all complex prepositions such as Tagalog *tungkól sa* 'about, regarding', *hanggáng sa* 'until' or Cebuano/Tagalog *para sa* 'for', Cebuano *human sa* 'after', etc. Perhaps even more importantly, Tagalog *sa* and phrases headed by *sa* allow stative

affixations which clearly bring out its essentially local meaning, for example, *na-sá-sa loób ng katawán* (RLS.ST-RDP-LOC heart GEN body) 'being within the body' (Bloomfield 1917:44).

Historically speaking it is highly likely that some phrase markers were in fact prepositions and others were deictics, and that the present paradigms thus consist of these two types of elements (cf. Himmelmann 1998). This mixed origin also contributes to the problems of their analysis. Note, for example, that the answer to the question of whether there are any primary prepositions in Philippine-type languages depends on whether one considers the oblique marking phrase markers such as Tagalog and Cebuano *sa* prepositions.

Perhaps the most far-reaching problem posed by the phrase marking paradigms is the problem that they rarely, if ever, provide clear-cut evidence for distinguishing core arguments from peripheral arguments (or adjuncts). That is, it is rarely the case that a given marker only applies to core arguments and another only to peripheral arguments. Tagalog *ng*, for example, is not only used for patients, themes and goals – i.e. roles one would expect to occur in core argument positions – but also for instruments:

(36) *p<in>utol* *niyá* ***ng*** ***gulok*** *ang* *kahoy*
<RLS(UG)>cut 3S.POSS GEN bolo SPEC wood
'S/he cut the wood with a bolo.' (Bloomfield 1917:175)

Arguably, *ng* is also used for manner and time (see Schachter and Otanes 1972:437f, 452f and Ross 2002:29) which, however, is somewhat obscured by the fact that in these functions it is orthographically represented as *nang*. Recall from above that Cebuano *ug* is also used for manner expressions.

Similarly, while Tagalog *sa* is used for a broad range of more peripheral roles, including location and time, it is also the regular marker for goals, recipients and addressees, and for definite patients and themes as in:

(37) *itó* *ang* *pusa-ng* *k<um>ain* ***sa*** ***dagá'***
PRX SPEC cat-LK <AV>eat LOC rat
'This is the cat that ate the rat.' (McFarland 1978:157)

Thus, the distribution of the phrase-marking clitics in Philippine-type languages does not reflect in any direct way the distinction between core and peripheral arguments. One may, in fact, doubt that such a distinction exists in these languages, an assumption which forms the core of the equational clause hypothesis mentioned above (section 3.4; cf. also Ross 2002:30).

However, two caveats have to be added to this conclusion. First, there may be other ways to diagnose a core vs. peripheral distinction. Kroeger (1993:40–47), for example, argues that in Tagalog all *ng*-marked arguments, including instruments, are core arguments and that all *sa*-marked arguments are peripheral. The argument invokes control phenomena and the fact that only *sa*-marked arguments allow fronting into a position immediately preceding the predicate. Arka (1998) and Arka and Manning (to appear) use evidence from reflexive binding to argue that only some actors in undergoer voice constructions in Balinese and Standard Indonesian are core arguments, while others are peripheral arguments. Musgrave (2001:67ff) supports the analysis of the Standard Indonesian data with evidence from quantifier floating. (Note that according to this argument, the Standard Indonesian example (2) above (*orang itu dilihat anak saya* 'my child saw that person') is a passive construction, with *anak saya* functioning as an oblique adjunct. However, if the actor is replaced by a pronominal clitic (as in *orang itu*

di-lihat=nya (person DIST PV-see=3.POSS) 's/he saw that person'), then the actor is a core argument and the overall construction a true symmetrical voice construction.)

Second, the distribution of the phrase marking clitics is also not fully accounted for by the equational hypothesis. The problem here is the definiteness alternation associated with *ng* vs. *sa* marking in **Tagalog** actor voice constructions (similar definiteness or specifity-related alternations in most other Philippine-type languages, but the details vary significantly). As shown by (37), undergoers in these constructions may be marked with *sa* and then are clearly definite. If *sa* in this example is replaced by *ng*, the preferred interpretation is specific-indefinite ('a rat') or even non-specific ('the cat that eats rats'). Contrary to what is often claimed in the literature, *ng* marking does not preclude a definite interpretation, as amply illustrated in McFarland (1978). In fact, as McFarland (1978:157) makes clear, there are four options in phrasing this sentence:

(38)	(a)	itó	ang	pusang	kumain	**sa**	dagá'		unambiguously definite=(37)
	(b)	itó	ang	pusang	kumain	**ng**	dagá'		indefinite or non-specific preferred, but definite also possible
	(c)	itó	ang	pusang	kumain	**ng**	**isáng**	dagá'	unambiguously indefinite (*isá*='one')
	(d)	itó	ang	pusang	kumain	**ng**	dagáng	**iyón**	unambiguously definite (*iyón*=DIST)

This definiteness-related *ng/sa* alternation is problematic for the equational hypothesis for the following reason. It seems to occur only in construction with voice-marked forms ('verbs'). In semantically clearly possessive constructions such as *bahay ng lalaki* (house GEN man) 'the/a man's house' there is (a) no preference for an indefinite interpretation of *lalaki* (in fact, a definite interpretation is more likely); and (b) *ng* cannot be replaced with *sa* (such a replacement would result in a barely acceptable expression meaning something like 'house at/on/in the man').

Cross-linguistically, this definiteness alternation is known as *differential object marking* (Comrie 1979, Lazard 2001). As is common in languages with differential object marking, the alternation is most consistently observed for animate beings. Thus, the basic rule is that definite expressions for animate beings receive a more oblique-like marking (in Tagalog *sa*) when occurring in non-subject undergoer functions. Personal pronouns and personal names, which are inherently definite, in fact tend to be restricted to an oblique (dative) form in these functions, as just noted for the Cebuano and Tagalog personal name markers.

The definiteness alternation is also of major relevance to the ergative hypothesis to be discussed in section 3.8.2 below. According to this hypothesis, actor voice constructions are antipassives, i.e. intransitive constructions where undergoers may occur only in oblique positions. Strictly speaking, this hypothesis would predict that in Tagalog all undergoers in actor voice constructions are marked with *sa*, which is evidently not the case. Alternatively, one could argue that *ng*-marked undergoers in these constructions are in some way incorporated objects, which makes sense for non-specific undergoers and perhaps also for indefinite-specific ones (as in (38)c). But the fact that examples such as (38)b allow a definite interpretation, and that (38)d is at all possible, is difficult to account for on the assumption that these are antipassive constructions.

The preceding discussion was mostly concerned with Philippine-type languages. The core-peripheral distinction is usually much clearer in the other western Austronesian languages. Nevertheless, many of these languages allow alternations where a construction with preposition alternates with one without it as in the following Tetun Dili example:

(39) *ha'u hanoin **(kona-ba)** ha'u-nia main.*
1s think about 1s-POSS mother
'I'm thinking about my mother.' (Hull and Eccles 2001:156)

As indicated by the parentheses, the preposition *kona-ba* can be omitted in this example, which in turn raises the question of whether *ha'u-nia main* is a core or a peripheral argument. See Bowden (this volume and 2001:157–166) for a more detailed discussion relating to similar examples in **Taba**.

Finally, it should be noted that in *all* western Austronesian languages all kinds of arguments can be omitted, i.e. there are no clear syntactic constraints on argument omission which could be used for distinguishing core and peripheral arguments (see section 5 for examples).

3.7 Pronouns and person marking

Almost all Austronesian languages make a distinction between first-person inclusive (speaker + addressee) and exclusive forms (speaker(s) only). This distinction is found in all paradigms of personal pronouns, regardless of their form (full, clitic, affixed) or function (subject marker, possessive suffix, etc.). A notable exception to this generalization among the western Austronesian languages is some Malayic varieties where this distinction has been lost (cf. Donohue and Smith 1998 and ADELAAR, MALAYIC VARIETIES).

Western Austronesian pronoun systems do not usually include dual or trial forms (exceptions include Iban and **Biak**). Third person forms are often restricted to human or animate referents, demonstratives being used in pronominal reference to inanimates. Full pronominal forms often include a personal article (or politeness clitic). Second person singular forms are usually used only in addressing friends or inferiors. Otherwise, politeness rules of various complexity are to be observed, which often forbid the use of pronouns in addressing people, for whom special terms of polite address or personal names are used instead (particularly complex systems of this type are found in western Indonesian languages, see FOX, RITUAL LANGUAGES and **Javanese**).

Most western Austronesian languages have a special paradigm of possessive pronouns, usually enclitics or suffixes. More often than not, these possessive pronouns (or forms more or less identical to them) occur also in other functions such as actor pronouns in undergoer voice constructions, etc.

Western Austronesian languages differ significantly with regard to person markers. The term *person marker* is used here as a cover term for both agreement markers (for example, the third singular formative *-s* in English *play-s*) and person markers proper, which are also called *pronominal arguments* (a term avoided here because of its controversial implications). Agreement markers and person markers proper have two characteristics in common: (a) they are affixed or cliticized to the predicate (or occur in a special clitic position); and (b) they may occur with a coreferential full nominal expression within the same nuclear clause. In the following Uma example, the third person singular prefix *na-* co-occurs with the coreferential noun phrase *tobinena* 'his wife'.

(40) ***na-manyu tobine-na*** *pae toe.*
3s-pound woman-3s.POSS rice DIST
'His wife pounded the rice.' (Martens 1988b:248)

The condition 'within the same nuclear clause' is intended to exclude instances of pronominal crossreference to nominal expressions which occur in a clearly clause-external position, for example as topics or rightward expansions ('afterthoughts'). Such constructions are possible in many, if not all, languages (cf. English ***it doesn't bring out the best in people, divorce***) and hence they are of no diagnostic value for person marking.

One way of distinguishing between agreement markers and person markers proper is that agreement markers generally require the co-occurrence of a coreferential nominal expression within the same clause (as in English or German), whereas person markers proper do not. According to this criterion, all person markers in western Austronesian languages are person markers proper.

Otherwise, western Austronesian languages differ widely with regard to the form and function of person markers. In fact, the southern half of Sulawesi and the Lesser Sunda islands are something of a laboratory attesting a bewildering variety of person marking systems. See **Mori Bawah**, **Makassar**, and **Kambera** for a glimpse of this variety, Haaksma (1933) for an early comparative study, and Mead (2002) for more recent discussion and references. Here we will only list some of the more remarkable features.

Person markers occur, on the one hand, in most transitional and preposed possessor languages (major exceptions are the isolating languages in the Flores – Timor region). On the other hand, they also occur in a few of the northernmost languages, i.e. the Philippine languages of northern Luzon (Kapampangan and languages further to the north) and a few Formosan languages, including **Tsou**, Rukai, and Pazeh (which only has one person marker, i.e. the first person inclusive prefix *ta-*; cf. Li and Tsuchida 2001:33,37). Furthermore, they occur in the two Micronesian outliers Chamorro and Palauan.

However, there are important differences between these two main person marking areas. Most importantly, all person marking systems in preposed possessor and transitional languages include at least one series of pronominal prefixes or proclitics while person markers in the northern languages usually only involve enclitics (which are also often second position clitics; Pazeh *ta-* is an obvious exception). Furthermore, in the northern systems third person forms are sometimes missing ('third person is zero'), which does not happen in preposed possessor and transitional languages. With regard to these two features, the two Micronesian outliers clearly pattern with the transitional and preposed possessor languages.

The pronominal prefixes and proclitics in most preposed possessor languages and some transitional languages crossreference S and A arguments and thus are usually called subject prefixes/proclitics. These languages differ as to whether use of the person markers is obligatory (as in **Nias**, Muna, Tukang Besi, Kedang, **Leti**, **Taba**, **Biak**) or optional (as in Buru and Alune). The subject marking prefixes often show very strong fusional tendencies in that they require the setting up of different inflectional (or conjugation) classes (this is particularly widespread in the Moluccas, but it is also attested in Muna and Kedang). In some preposed possessor languages, including **Taba**, Dobel and Selaru, the prefixes only crossreference A and S_a arguments (S arguments of dynamic verbs), which is a typical characteristic of split-intransitive languages (see section 3.3 above).

A conspicuous feature of a number of transitional languages (including **Nias**, Muna, and Tukang Besi) is the presence of two series of pronominal prefixes/proclitics, one for realis mood and one for non-realis (or irrealis) mood, a feature which is also very common in western Oceanic languages (see also the future series in **Mori Bawah**). In these series, it is usually impossible to separate person marking from mood marking formatives. In this regard they are similar to the voice-mood affixes in Philippine-type languages, which also cannot formally be separated into voice and mood marking segments (see section 4.2.2).

With very few exceptions, two person-mood marking prefix series are found neither in symmetrical voice languages nor in preposed possessor languages (Chamorro is a major exception among the symmetrical voice languages). But quite a few symmetrical voice languages in Sulawesi and western Indonesia show one (often incomplete) series of pronominal proclitics marking person and non-realis mood (see section 4.2.2).

Person markers for O arguments are not attested in preposed possessor languages but they do occur in a number of transitional languages (e.g. **Kambera**, Tukang Besi). Note also that there are often special enclitic or suffixal forms for pronominal O arguments in preposed possessor languages (e.g. Alune *=ma* 'first person plural transitive undergoer') but these do not co-occur with coreferential nominal expressions and hence are not person markers according to the definition above.

In a number of transitional languages (including **Makassar**) and northern symmetrical voice languages, in particular Kapampangan, person markers show ergative alignment, i.e. there is one marker for S and O arguments (*=i* in (41) and (42)) and another one for the A argument (*ku=*in (42)), as in the following two examples from Konjo:

(41) *a'-lampa=**i***
INTR-go=3.ABS
'S/he goes.' (Friberg 1996:140)

(42) ***ku**=kanre=**i***
1.ERG=eat=3.ABS
'I am eating it.' (Friberg 1996:153)

However, in most of these languages, there are a number of syntactic contexts where ergative alignment is suspended. In **Makassar** and Konjo, for example, S arguments are regularly crossreferenced by the 'ergative' prefix after negation and a number of adverbials (compare Konjo *anre' **ku**='-lampa* (NEG 1.ERG=INTR-go) 'I am not going' (Friberg 1996:152)). Perhaps even more remarkable from a crosslinguistic point of view is the fact that there are constructions where both core arguments are crossreferenced by 'ergative' prefixes. In Konjo this happens, for example, when the completive maker *=mo* is added to the predicate (see Friberg 1996:168):

(43) ***ku=na**=peppe'=mo* *Ali.*
1.ERG=3.ERG=hit=CPL Ali
'Ali hit me.' (Friberg 1996:168)

In a number of languages, including Uma and **Tsou**, only the A argument is regularly crossreferenced by a person marker, while the S and O arguments are expressed either by pronominal clitics or by full NPs but not both.

In addition to person markers, there are also definiteness or specificity marking clitics or suffixes which are identical to, or derived from, pronouns, in particular in the languages of Indonesia. Thus, for example, the third person possessive pronoun has clearly extended uses in many **Malayic varieties**, including **Colloquial Indonesian**, Balinese, Madurese, **Tetun Fehan**, etc. (cf. Himmelmann 1997:219f on possessive articles). See also the so-called indexer clitic in **Leti** and the pronominal articles in **Biak**.

3.8 Grammatical relations

In most preposed possessor languages the identification of a subject relation in verbal clauses is uncontroversial. The grammatical relation systems in these languages are

usually characterized by nominative-accusative alignments (S and A arguments share the same coding and behavioral properties, while O arguments show a different set of properties; see **Tetun Fehan, Leti** and **Biak** for exemplification). However, there is also often some evidence for split-intransitivity (see section 3.3), and some languages (e.g. **Taba**) are explicitly analyzed as having a mixed nominative-accusative and split-S system.

The nature of grammatical relations in symmetrical voice languages and in many transitional languages continues to be a matter of controversy. The crux of this debate is the question of whether one can identify a subject relation in these languages. A related, though not identical question is whether they are best analyzed as ergative languages. Alternatively, they make up a type of their own or they are split-S languages ('active' languages in the sense of Klimov (1977)). Subjecthood will be discussed first.

3.8.1 Subjecthood

Following a paper by Keenan (1976), it has been widely accepted in the typological literature that the grammatical relation *subject* found in European languages applies to an argument expression which exhibits a specific set of coding (e.g. the subject is morphologically unmarked) and behavioral properties (e.g. the subject can be relativized), distinguishing it from other core arguments of the same predicate. While Keenan lists some thirty properties, the discussion of the distribution of such properties in western Austronesian languages has concentrated on the ones listed in Table 5.4. Following Schachter (1976) and Foley and Van Valin (1984), these properties are grouped in two subsets. Some properties are considered to be *reference-related* in that they reflect the fact that subjects tend to provide topical (or given) information. Other properties are called *role-related* because they appear to be linked to the fact that subjects tend to be actors (in basic underived transitive clauses).

The properties listed in Table 5.4 (and a few others) are discussed and illustrated in detail in Schachter (1976, 1995), Kroeger (1993:19–39), Cena (1995) and Donohue (1999:463–481), among others. Here it will be sufficient briefly to illustrate those properties mentioned in discussions of subjecthood in the grammar sketches assembled in this volume. For *relativization* see section 3.10.

Quantifier floating refers to the phenomenon widely attested in the languages of the world that quantifiers do not have to occur within the nominal expression they quantify. Thus, for example, the Tagalog quantifier *lahát* may occur either within the quantified NP (as in (44)) or immediately after the predicate, as in (45), where it is called a *floated quantifier*.

TABLE 5.4: SUBJECT DIAGNOSTICS COMMONLY USED FOR WESTERN AUSTRONESIAN LANGUAGES

Reference-related properties	Role-related properties
Coordinate conjunction reduction	Control of reflexives
Raising	Controllee in control constructions
Relativization	Addressee of imperatives
Agreement	
Quantifier floating	

(44) *sinúsúlat* *ng* *mga* *bata'* *ang* ***lahát*** *na* *mga* *liham*
RDP:RLS(UG):write GEN PL child SPEC all LK PL letter
'The children write all the letters.'

(45) *sinúsúlat* ***lahát*** *ng* *mga* *bata'* *ang* *mga* *liham.*
RDP:RLS(UG): write all GEN PL child SPEC PL letter
'The children write all the letters.' (Kroeger 1993:22)

Floated quantifiers in Tagalog and other western Austronesian languages are of relevance to the problem of subjecthood because they are often unambiguously related to only one of the two core arguments of a transitive predicate, i.e. the one marked with the specific article *ang* (or an equivalent pronominal form). Thus, example (45) cannot mean *All the children are writing letters*. In order to express this with a floated quantifier, one has to convert the clause from undergoer voice to actor voice so that *mga bata'* occurs in the *ang*-phrase:

(46) *sumúsúlat* ***lahát*** *ng* *mga* *liham* *ang* *mga* *bata'.*
RDP:AV:write all GEN PL letter SPEC PL child
'All the children are writing letters.' (Kroeger 1993:22)

If in a given language only one of the two core arguments of a transitive predicate is able to launch a floated quantifier, this is typically the subject argument. Therefore, so the argument goes, the ability of *ang*-phrases to launch quantifiers is one argument in favor of considering *ang*-phrases subjects.

A *raising* construction involves two clauses, a matrix clause and an embedded clause, the latter functioning as a complement of the matrix clause predicate (as in [*I expect* [*that Linda will sing the national anthem*]]). In many languages it is possible to *raise* to the matrix clause an argument which semantically belongs in the subordinate clause, and to make it structurally an argument of the matrix predicate. In (47), a 'normal' Tagalog complement clause construction without raising, *pambansáng awit* 'national anthem' is structurally part of the complement clause, functioning as subject of the complement clause predicate *awitin* 'sing'. In (48) it is raised, now functioning as the subject argument of the matrix clause predicate *inasahan* 'expect, hope'.

(47) *<in>asah-an* *ko* *[na* *awit-in* *ni* *Linda*
<RLS(UG) >hope-LV 1S.POSS COMP song-PV PN.POSS Linda
ang ***pambansáng*** ***awit]***
SPEC GER:nation:LK anthem
'I expected (for) Linda to sing the national anthem.' (Kroeger 1993:28)

(48) *<in>asah-an* *ko* ***ang*** ***pambansáng*** ***awit***
<RLS(UG)>hope-LV 1S.POSS SPEC GER:nation:LK anthem
[na *awit-in* *ni* *Linda]*
COMP song-PV PN.POSS Linda
'I expected the national anthem to be sung by Linda.' (Kroeger 1993:28)

The diagnostic value of the raising construction resides in the fact that typically not all core arguments of the complement clause predicate can be raised in this way. For Tagalog, it is claimed that only phrases marked by *ang* (or an equivalent pronominal form) allow raising (but see Kroeger 1993:28f for some complications). Thus, in (47) it is not possible to raise the other core argument of the complement clause predicate *awitin*, i.e. *Linda*, which is marked by the possessive marker *ni* (**inasahan ko* ***si Linda*** *na awitin ang pambansáng awit* is ungrammatical).

Control (or *Equi-NP deletion*) constructions also consist of a matrix and an embedded clause. In these constructions, the matrix predicate and the embedded predicate share one core argument (thus in *I avoided looking at Linda* the pronoun *I* is an argument of both *avoid* and *looking at*). In many languages, this shared argument can be overtly expressed only once in the overall construction, leaving one of the argument slots of the embedded predicate empty (it is impossible to say in English **I avoided I/me looking at Linda*). But there is never a doubt as to the reference of the omitted argument, the empty argument slot of the embedded predicate (the controllee) being controlled by an argument of the matrix predicate (the controller). Often there are restrictions on the choice of the controllee in that not all argument slots of the embedded predicate can become controllees. In English, the controllee has to be the subject of the embedded predicate. Hence, both I_i *want* e_i *to look at Linda* and *I want Linda*$_i$ e_i *to look at me* are fine, but it is impossible to omit the object *me* in the second clause (*I_i *want Linda to look at* e_i).

In Tagalog and many other western Austronesian languages no such clear-cut grammatical constraint on controllee choice seems to exist. In fact, the relevant facts appear to be quite complex and continue to be a matter of controversy (cf. Kroeger 1993:38f, 71–107 and Schachter 1995:21–27). Here it will suffice to note that at least in some control constructions the grammatical function of the controllee (whether it appears in an *ang*-phrase or in a genitive phrase) is irrelevant as long as the omitted argument is the actor of the embedded predicate:

(49) (a) *<um>iwas* *akó-ng* *t<um>ingín* *kay* *Linda.*
<AV>avoidance 1.SG-COMP <AV>look DAT.PN Linda
(b) *<um>iwas* *akó-ng* *tingn-án* *si* *Linda.*
<AV>avoidance 1.SG-COMP look-LV PN Linda
'I avoided looking at Linda.' (Kroeger 1993:39)

In (49)a, the controllee, if overtly expressed, would have to occur in the *ang*-form (*tumingín* ***akó*** *kay Linda*) because it is the actor in an actor voice construction. In the (b) example, an undergoer voice construction, it would have to occur as a possessive pronoun (*tingnán* ***ko*** *si Linda*). Given such examples, it is clear that *ang*-phrases in Tagalog are not the only possible controllees in control constructions, a point where they differ quite clearly from subjects in European languages.

In *coordinate conjunction reduction* two conjoined main clauses share an argument which remains unexpressed in one of the conjuncts. Thus, in *Peter looked at me and left without another word* the subject of *left* remains unexpressed but, importantly, there is absolutely no ambiguity as to who actually left (i.e. *Peter*). As Kroeger (1993:33f) points out, Tagalog seems to allow basically any core argument to be omitted in such constructions, regardless of its semantic role or grammatical function. In the following example, the actor of the predicate *hinugasan* in the second clause remains unexpressed:

(50) *ni-luto'* *ni* *Josie* *ang* *pagkain*
RLS(UG)-cooked PN.POSS Josie SPEC GER:eating
at *h<in>ugas-an* *ang* *mga* *pinggán*
and <RLS(UG)>washing-LV SPEC PL dish
'The food was cooked by Josie, and the dishes washed (by her).' (Kroeger 1993:34)

Overtly expressed, the actor would have to be a possessive pronoun *niyá*, hence the 'complete' version of the second clause would read *hinugasan* ***niyá*** *ang mga pinggán*.

Kroeger (1993:36) argues that omissions such as the one illustrated in example (50) are actually instances of zero anaphora for which the same constraints hold within a sentence and across sentence boundaries (see section 5 below). Zero anaphora is to be distinguished from 'true' or logical conjunction reduction in which it is possible to omit an argument in the first conjunct, i.e. before its 'antecedent' has actually been mentioned:

(51) *hú-hugas-an* *ko* *at* *pú-punas-an* *mo*
RDP-washing-LV 1S.POSS and RPD-wipe-LV 2S.POSS
ang *mga* *pinggán*
SPEC PL dish
'I will wash and you dry the dishes.' (Kroeger 1993:34)

Here the undergoer of *húhugasan* (i.e. *ang mga pinggán*) remains unexpressed in the first clause. According to Kroeger, logical conjunction reduction is only possible for *ang*-phrases. For example, it would be unacceptable to omit the actor expression in the first conjunct of two undergoer voice clauses (as in ?**húhugasan ang mga pinggán at púpunasan ko* for 'I will wash and dry the dishes').

After this brief review of the subjecthood diagnostics most commonly used in the Austronesian literature, we will now turn to a discussion of their application to western Austronesian languages. It will be useful to begin with a brief review of the discussion of subjecthood in Tagalog (see also McKaughan 1973, Matsuda French 1988, Kroeger 1993:19–22). The Spanish grammarians in the seventeenth and eighteenth centuries as well as authors such as Bloomfield (1917), Blake (1925) and Lopez (1937) use the term *subject* in their descriptions of Tagalog without further comment and consistently apply it to the *ang*-phrase in post-predicate position. In the 1950s, a number of researchers affiliated with the Summer Institute of Linguistics saw the need to highlight the differences between Philippine and European languages with regard to grammatical relations and used *topic* for the *ang*-phrase in post-predicate position (and *focus* instead of *voice*). McKaughan, who seems to be the first to have used *topic* in this sense in print (McKaughan 1958), in a later paper (1973) considers this an unfortunate and possibly misleading move and reverts to calling the *ang*-phrase in post-predicate position a subject.

Schachter (1976) reopens the debate by arguing that in Tagalog subject properties such as the ones listed in Table 5.4 are distributed among the two core arguments of a transitive construction and that therefore neither of them can truly be considered a subject. Schachter's claim continues to be widely quoted and accepted in the typological as well as the formal syntax literature. Within the Austronesianist literature, however, support for this analysis has been on the wane, especially since Kroeger's (1993) detailed and largely convincing critique of Schachter's results (DeWolf's (1979:67–86, 1988:144–150) very similar argument has been largely ignored in the literature). Kroeger argues that there are quite a number of subject properties (including the reference-related ones in Table 5.4) which uniquely apply to the *ang*-phrase. Furthermore, he shows that the subject properties which do not uniquely apply to the *ang*-phrase (most importantly control of reflexives and target in control constructions) also do not uniquely apply to any other core argument of a transitive verb and hence are irrelevant to determining subjecthood in Tagalog. Consequently, the only sensible candidate for subjecthood in Tagalog is the *ang*-phrase in post-predicate position (see Cena 1995 and Schachter 1995 for a critique of this argument and conclusion).

From Kroeger's argument it does not necessarily follow, however, that the *ang*-phrase is indeed the subject of a Tagalog clause. Such a conclusion very much depends on the

additional assumption that it is sufficient for a given argument expression to exhibit a number of subject properties to qualify for subjecthood. That is, it is still possible to argue that *ang*-phrases are not subjects because they do not show enough subject properties or the right subset of subject properties (which in turn presupposes a catalogue of properties that have to hold uniquely of an argument expression in order to qualify for subjecthood).

Although the property list-approach to subjecthood is the one most widely used in the more recent literature, it should be noted that this is not the only possible approach. The original intuition of the Greek and Latin grammarians who coined the terms *subject* and *predicate* pertained to equational (or nominal) clauses such as *She is an actress*. Such clauses are clearly bipartite, consisting of a referential expression denoting the entity about which an assertion is made – the subject – and an expression conveying the assertion – the predicate. It is a matter of cross-linguistic variation to what degree this simple bipartite structure can also be applied to narrative (or verbal) clauses such as *She collects doors*. Transitive narrative clauses consist of a verb (*collects*) and two arguments (*she, doors*). Applying the bipartite structure of equational clauses to narrative clauses presupposes an asymmetry in the relationship of the two arguments to the verb in such a way that one argument forms a closer unit with the verb and thus is interpreted as being part of the predicate (assertion) while the other argument provides the referential anchor or basis for the predication, denoting the entity about which the predication is made.

Languages differ with regard to the degree to which equational and narrative clauses are structurally similar. In Arabic, for example, the two clause types have very little in common (equational clauses have subject-predicate order, narrative clauses have verb-actor-undergoer order, etc.). In Indo-European languages, there are significant overlaps between equational and (many) narrative clauses in that, for example, one argument in a narrative clause usually receives the same case marking as the subject of an equational clause (i.e. nominative case) and the verb of the narrative clause agrees with this argument (and only with this argument) in the same way as the predicate agrees with the subject of an equational clause. This, essentially, is the reason why in grammatical analyses of Indo-European languages the term *subject* has been applied to the one argument in narrative clauses which is most similar to the subject of an equational clause.

The traditional approach to subjecthood outlined in the preceding paragraphs has been repeatedly criticized for being too vague to be of much practical value. The properties approach to subjecthood can be seen as a way to make one insight of the traditional approach more operational by providing tests for diagnosing asymmetries between the core arguments of narrative clauses. And this is in fact the major use that has been made of Keenan's putative subject properties in the Austronesian literature. They are used to test for morphosyntactic asymmetries between core arguments. Since in Austronesian languages these properties rarely cluster in the same way as in European languages, it is a widespread practice to use the term (syntactic) *pivot* instead of *subject* for arguments which are privileged vis-à-vis other arguments in that they uniquely show a number of morphosyntactic properties. (Note that there is a second, overlapping but more restricted use of *pivot* where it refers to the argument which is privileged in clause combining, for example, the omitted argument in coordinate constructions (cf. Dixon 1994:11). On its likely first appearance in Heath (1975:99), *pivot* refers only to the controllee in control constructions.)

However, the asymmetry between the core arguments of a narrative clause is only one aspect of the traditional subject intuition. The other major aspect demands that the privileged argument of a narrative clause displays morphosyntactic properties also displayed

by the subject in an equational clause. These two aspects do not necessarily correlate with each other. To give just one, somewhat superficial example: in most preposed possessor languages, it is possible to mark the A argument in a transitive narrative clause by a prefix or proclitic on the verb (see also section 3.7). In the following Buru example, it is a proclitic:

(52) ***Ya***=*paha* *ringe* 'I hit him.' (Grimes 1991:151)
1s=hit 3s

There is no comparable paradigm of forms to mark the O argument on the verb, hence the existence of the proclitic series could be taken as evidence for an asymmetry between A and O arguments, privileging the A argument. However, the proclitics are restricted to verbal predicates (Grimes 1991:373). They cannot be used with the nominal predicates characteristic of equational clauses. Therefore, the existence of a series of pronominal proclitics does not render A arguments similar to the subjects of equational clauses, and their diagnostic value for determining subjecthood in the traditional sense is thus questionable.

If formal similarity to the subjects of equational clauses is taken to be the major criterion for diagnosing subjecthood in narrative clauses, one could argue that subjects of narrative clauses in Tagalog and many other symmetrical voice languages are in fact *more* subject-like than subjects of narrative clauses in European languages, simply because narrative and equational clause structures in Tagalog are very similar indeed. As discussed in section 3.4 above, proponents of the equational hypothesis in fact hold that there is no narrative clause type in Tagalog and that with the exception of a few minor clause types (e.g. existential clauses) all clauses follow an essentially equational pattern. But even if one does not subscribe to this hypothesis, many analysts would agree that the similarities between the two clause types are considerable in these languages (see, for example, the quote from Schachter and Otanes (1972:62) presented above in section 3.4).

From this point of view it would appear to be ironic that a group of languages where the subject relation is clearly manifest has been at the center of recent discussions concerning subjecthood. In this regard, it will be helpful to take note of the fact that all discussions of subject properties in symmetrical voice languages which are framed within Keenan's property list-approach fail to adhere to a very basic parameter set by Keenan. At the beginning of his 1976 paper, Keenan makes a considerable effort to define the notion of *semantically basic sentences*, and the ensuing subject property list only holds for *basic subjects* which occur in such *basic sentences*. In English, the basic sentence is identified with an active transitive clause. Consequently, the properties of subjects of passive sentences, for example, do not appear on the list.

In discussions of symmetrical voice languages, on the other hand, actor voice as well as undergoer voice clauses are used in the argument (without further discussion!), on the assumption that they are both equally basic. The fact that subject properties appear to be distributed across different arguments in these languages is a direct consequence of this use of multiple basic transitive constructions. It is not clear, however, whether Keenan's methodology is in fact applicable to languages with multiple basic transitive constructions.

3.8.2 Alignment systems: ergative, split-S or none of the above?

As mentioned in the introduction to this section, there is little controversy with regard to the assessment that preposed possessor languages show either nominative-accusative or mixed nominative-accusative and split-S alignments in their systems of grammatical

relations. For symmetrical voice languages and many transitional languages, however, the issue of their basic alignment system continues to be heavily contested. Yet to begin with, we may note that it is widely agreed that most of these languages are *not* nominative-accusative. Furthermore, Acehnese has been forcefully argued to be a split-S language (Durie 1985, 1987). Split-S characteristics are also found in most other symmetrical voice and transitional languages but it is questionable whether these characteristics actually pertain to the clause level and thus are relevant for determining the alignment system (see section 3.3 above).

The major controversy, then, pertains to the question of whether symmetrical voice languages show ergative alignment or rather constitute a type of their own (for a brief illustration of ergative alignment, see examples (41) and (42) above). Ergative analyses of symmetrical voice and transitional languages have been proposed in quite a number of ways (for example, by Payne 1982, DeGuzman 1988, Martens 1988a, Blake 1988, Mithun 1994, Wechsler and Arka 1998). Practically all aspects of these proposals have been criticized in an equally varied number of ways (see, for example, Cumming and Wouk 1987, Shibatani 1988, DeWolf 1988, Kroeger 1993:47f, Schachter 1995:38–51, Foley to appear). Here no attempt is made to review the debate in detail but the following three general observations may be of use in assessing its relevance to typological concerns.

As mentioned in section 3.7, clear cases of ergative alignments in western Austronesian languages usually pertain to the distribution of person markers in languages such as Kapampangan, **Makassar**, Bugis, etc. The interesting point to observe here is that the distribution in western Austronesian languages is the converse of that found in many other ergative languages in, for example, Australia. That is, while in Australian ergative languages it is very common to find a split between a nominal case marking system showing ergative alignment and a pronominal or person marking system showing nominative-accusative alignment, in western Austronesian languages clear-cut cases of ergative alignment are restricted to person marking systems, with little or no evidence of an ergative distribution of noun phrase markers.

A second point that seems to be widely overlooked in the sometimes heated debate regarding ergativity in western Austronesian languages is the fact that very little follows from the assessment that a given language shows evidence of ergative alignment, as stated clearly in Dixon's recent survey of the phenomenon (where western Austronesian languages are excluded from further consideration):

> What then does it mean for a language to be ergative? Exactly what we said in the first paragraph of Chapter 1– that S is treated in the same way as O and differently from A in some part or parts of the grammar. *Nothing else necessarily accompanies this.* (Dixon 1994:219, emphasis added)

That is, showing that a given language is ergative is of comparatively little typological import in that nothing else correlates with it. And it is probably fair to say that most proponents of an ergative analysis would agree that putatively ergative western Austronesian languages have very little in common with ergative languages in other parts of the world, with the possible exception of Mayan languages (Martens 1988a:270f).

Third, it bears emphasizing that diagnosing ergativity or accusativity presupposes a reasonably clear distinction between syntactically transitive and intransitive constructions in a given language. Furthermore, it helps a lot when core arguments are clearly distinguishable from peripheral arguments. It would appear that these two preconditions are not, or at least not very clearly, fulfilled in many symmetrical voice languages, as argued

in sections 3.3 and 3.6. In this view, much of the ergativity debate misses an essential point in that it focuses on providing evidence for ergative alignments rather than for the transitive/intransitive and core/oblique distinctions presupposed by such an alignment.

The last point indicates one possible venue for determining the place of symmetrical voice languages in a comprehensive typology of systems of grammatical relations. That is, in addition to the well-established distinction between nominative-accusative, ergative and split-S ('active') languages, one would need a superordinate parameter which distinguishes languages with clear-cut syntactic transitivity distinctions from those where transitivity is less clearly manifest in the morphosyntax (see also Egerod 1988). Nominative-accusative and ergative-absolutive systems presuppose grammaticized transitivity distinctions while syntactic transitivity distinctions are largely irrelevant for symmetrical voice and split-S systems.

3.9 Multi-predicate constructions: auxiliaries, complex predicates, serial verbs, and the like

Western Austronesian languages are rich in multi-predicate constructions, i.e. constructions which involve two or more phonologically independent predicate expressions within a single clause. However, there are many unresolved descriptive and theoretical issues with regard to these constructions, and it would be premature to attempt any typological generalizations at this point. The purpose of the present section is simply to point out some of the phenomena which eventually would have to be covered by typological generalizations.

Elements termed *auxiliaries* are widely attested in descriptions of western Austronesian languages. Such elements usually convey notions of tense, aspect, mood, negation, or manner. Some of them are clearly clitics and hence do not qualify as phonologically independent predicates in multi-predicate constructions. But others are phonologically independent and also often have some other characteristics of independent predicates. For example, **Tsou** auxiliaries – which are obligatory in all verbal clauses – do not only determine the tense and mood of the clause but are also marked for voice and are suffixed with person markers (in some contexts):

(53) ***mo-'u*** *bonʉ* *to* *tacumʉ*
AV.RLS-1S AV:eat OBL banana
'I ate a banana.'

Some auxiliaries determine the form of the following predicate expression. Thus, the Cebuano past tense negator *walá?* (which also serves as negative existential, see (22) above) requires that the following 'main' predicate occurs in subjunctive form (see also section 3.7 in **Seediq**):

(54) *walá?* *niya* *lutu-**a***
NEG.PST 3S.POSS cook-PV.SUBJ
'He did not cook it.' (Wolff 1972:1121)

Kroeger (1993:139) proposes a complex predicate analysis for a similar construction involving the Tagalog negative imperative auxiliary *huwág*. He also notes (1993:181–201) that so-called 'pseudo-verbs' such as Tagalog *gusto* allow for two alternative constructions, one a biclausal control construction (55), the other a monoclausal complex predicate construction (56).

(55) ***gusto*** *ng* *nanay* *(na)* ***p<um>untá*** *sa* *tindahan*
liking GEN mother COMP <AV>direction LOC store
'Mother wants to go to the store.' (Kroeger 1993: 184)

(56) ***gusto-ng*** ***p<um>untá*** *sa* *tindahan* *ang* *nanay*
liking-LK <AV>direction LOC store SPEC mother
'Mother wants to go to the store.' (Kroeger 1993:184)

Note in particular the different phrase marking for 'mother' in the two examples. In (55) *nanay* is preceded by genitive *ng*, as required by the 'pseudo-verb' *gusto*. In (56) it is marked with *ang*, as required when functioning as the subject of actor voice *pumuntá*.

Tagalog 'pseudo-verbs' (Schachter and Otanes 1972:261–273) are auxiliary-like in that they convey modal meanings ('want, like', 'must', 'ought', 'able', etc.) and are not marked for aspect/mood and voice (or only allow a reduced set of such markers). They differ from typical auxiliaries in that they also allow for main predicate uses (e.g. *gusto ko itó* (liking 1s.POSS PRX) 'I want this').

Many other western Austronesian languages also have predicative expressions which occur both as independent predicates and in some kind of complex predicate construction with other predicates. These are not restricted to modal expressions. Another typical and widely attested class of such predicates are directionals. In the following two examples from Ratahan, the first illustrates the use of a directional as an independent predicate (with further directional proclitics). The second example shows the same directional in construction with another predicate:

(57) *ku=ta=****sá*** *e* *Kinaepesan*
MOTION=AND=DIR CPL Kinaepesan
'(We) **went on down** to Kinaepesan.' (Himmelmann and Wolff 1999:79)

(58) *te* *tintúr* ***sá*** *ngkami*
te in -ntur sa ni=kami
CON PST-deliver DIR GEN=1pe
'So we carried it **down there**.' (Himmelmann and Wolff 1999:79)

Compare also the discussion of **Buol** *maa* 'go' and *magi* 'come' and sections 3.3 in **Mori Bawah**, 3.4 in **Nias**, 3.8 in **Seediq**, and 3.7 in **Belait**.

In a number of western Austronesian languages a second predicate within a clause is used to introduce an additional argument, as in the following Tukang Besi example:

(59) *no-helo?a* *te* *roukau* ***ako*** *te* *ana-no*
3.RLS-cook CORE vegetable do.for CORE child-3.POSS
'He cooked the vegetables for his children.' (Donohue 1999:182)

Such constructions are considered prototypical examples of *serial verb constructions* in the literature. They have been sporadically reported for western Austronesian languages, including some **Malayic Varieties**, **Taba** and **Tetun Fehan** (and of course **Moken/Moklen** and **Cham**, where their occurrence is part of the typological profile of Southeast Asian mainland languages).

Finally, as noted by Kroeger (1993:196), in Tagalog and possibly other western Austronesian languages it appears that some controlled complement constructions show features of complex predications (i.e. they are mono- rather than biclausal) quite similar to the 'pseudo-verb' example (56) given above. In the following Tagalog example, the standard linker *na* is missing and only a few clitic elements would be allowed to appear in between the two predicates:

(60) *ang* *tuláy* *ay* ***ipinagbawal*** ***gamitin***
ang tuláy ay i-in-pag-bawal gamit-in
SPEC bridge PM CV-RLS(UG)-GER-forbidden use-PV
'The bridge was condemned (forbidden to be used).' (English 1986)

Appropriate ways to analyze the preceding constructions are still very much a matter of debate. Apart from considering them auxiliary or complex predicate constructions (with serial verb constructions being one type of complex predicates), some of these constructions are perhaps best analyzed as verbal compounds, as advocated by Klamer for **Kambera**.

3.10 Relative clauses

Western Austronesian languages are famous for their restrictions on the formation of relative clauses and have been of major import for the noun phrase accessibility hierarchy proposed by Keenan and Comrie (1977). The basic and central observation is that among the core arguments of a predicate, only subjects may be relativized. That is, the head to which the relative clause is attributed has to be the subject of the relative clause. Compare the following examples from Tagalog widely cited in the literature (in Tagalog, relative clauses are attached to their heads with a linker):

(61) (a) *isdá-ng* *i-b<in>igáy* *niyá* *sa* *bata'*
fish-LK CV-<RLS(UG)>gift 3s.POSS LOC child
'the fish which she gave to the child'

(b) *bata-ng* *b<in>igy-án* *niyá* *ng* *isdá'*
child-LK <RLS(UG)>gift-LV 3s.POSS GEN fish
'the child to whom she gave the fish'

(c) **isdá-ng* *nag-bigáy* *siyá* *sa* *bata'*
fish-LK RLS.AV-gift 3s LOC child

(d) **isdá-ng* *b<in>igy-án* *niyá* *ang* *bata'*
fish-LK <RLS(UG)>gift-LV 3s.POSS SPEC child

In the (a) and (b) examples, the head of the relative clause (*isdá'* in (a) and *bata'* in (b)) is also the subject of the relative clause predicate which shows the appropriate voice affixation (conveyance voice in (a) and locative voice in (b)). There is also no nominal expression within the relative clause which formally could function as a subject (in which case it would have to appear in the *ang*-form). In the (c) and (d) examples, this is not the case, and these examples are therefore ungrammatical. In (c), the head of the relative clause (*isdá'*) is the theme argument of the predicate *bigáy*, but the predicate is marked for actor voice, not conveyance voice, which indicates that the actor (*siyá*) is its subject. Similarly, in (d) the relative clause predicate is marked for locative voice, indicating that the recipient (*ang bata'*) is the subject.

Note that the grammatical function of the head of the relative clause in the main clause is of no relevance to the grammaticality of the overall construction. Thus, the Tagalog equivalents of *The fish which she gave to the child smelled bad* (*fish* is subject), *I could smell the fish which she gave to the child* (*fish* is object), or *They returned with the fish which she gave to the child* (*fish* is oblique) are all grammatical.

There are two important qualifications with regard to the subjects-only constraint on relativization. First, this constraint is widely attested only in symmetrical voice languages. Transitional languages and preposed possessor languages usually allow all core arguments to head relative clauses. Occasionally, essentially the same relativizing strategy may be used for all core arguments (e.g. **Tetun Fehan**, **Taba**). More commonly, however, there are a number of different relative clause structures depending on the grammatical role of the head noun within the relative clause. This is briefly illustrated here with examples from Tukang Besi (see also **Kambera**, **Leti**, **Nias**, **Biak**). Donohue (1999:367) distinguishes four major types of relative clauses, only two of which are mentioned here. In subject relative clauses the head noun functions as A or S argument of the embedded predicate which lacks subject prefixes but instead is infixed with *-um-*. Otherwise, subject relative clauses are structurally identical to main clauses.

(62) *no-lagu-mo* *na* *La Judi* *[b<um>alu* *te* *loka*
3.RLS-song-PRF NOM PN <REL>buy CORE banana
ako *te* *ina-no]*
for CORE mother-3.POSS
'La Judi, who bought some bananas for his mother, is singing.' (Donohue 1999:372)

In object relative clauses the head noun is the O argument of the embedded predicate, which also lacks subject prefixes and is prefixed with *i-* (also *di-* or *ni-*). The other arguments of the embedded predicate do not receive their main clause marking (with nominative *na* or core *te*) but are coded as possessives or genitives, thus giving object relative clauses a distinctly nominalized appearance.

(63) *te* *po'o* *[i-tompa-api-su* *u* *La Mar]*
CORE mango REL-throw-APP-1S.POSS GEN Mark
no-sangka-mo *ki'iki'i*
3.RLS-exceed-PRF little
'The mango that I threw over to Mark is a bit overripe.' (Donohue 1999:383)

Note that Tukang Besi relative clauses are marked by formatives (*-um-*, *i-*) which are widespread as voice-marking formatives in symmetrical voice languages (see also Donohue 2002:92f). Such specialized and sometimes also fossilized uses of originally voice-marking morphology are very common in transitional languages and are occasionally also found in preposed possessor languages.

The second qualification with regard to the subjects-only constraint on relativization pertains to the fact that in many symmetrical voice languages it appears to be possible to relativize on some non-core arguments or adjuncts, in particular possessors. It is not unusual that in these instances the head noun is overtly represented in the relative clause by a resumptive pronoun such as the enclitic *=nya* in the following Standard Indonesian example:

(64) *sopir* *yang* *nama=nya* *Ali*
driver REL name=3.POSS Ali
'the driver whose name is Ali' (Sneddon 1996:288)

Sneddon (1996:288f) suggests that this type of relative clause, which in Standard Indonesian and some other Malayic varieties can also be used for other non-subject core arguments and adjuncts, basically instantiates an embedded topic-comment structure (see Musgrave 2001, chapter 5 for extensive discussion). The independent (non-embedded)

version of this topic-comment structure is *sopir itu nama=nya Ali* (driver DIST name=3.POSS Ali) 'that driver's name is Ali (lit. as for that driver, his name is Ali)'.

Related to these topic-comment structures are relative clause-like constructions where a word meaning 'place', 'time' or 'reason' is followed by a main clause structure, the overall construction serving as an adverbial clause ('(at the time) when', '(at the place) where', 'the reason why'). A Standard Indonesian example is given in (65) (for other examples, see **Belait**).

(65) *waktu* *saya* *masih* *sekolah*
time 1s still school
'At the time when I was still going to school' (the following incident happened).

Such constructions are widely attested in symmetrical voice languages. While very general words meaning 'place' or 'time' are the most common 'heads', some more specific items are also possible as seen in the following Tagalog example:

(66) ***isá-ng*** ***hapon*** *na* *silá-ng* *dalawá* *'y* *nag-lá-laró'*
one-LK afternoon LK 3p-LK two PM RLS.AV-RDP-play
sa *halamanan*
LOC garden
'One afternoon when the two of them were playing in the garden' (...a butterfly came flying past the two children). (Bloomfield 1917:88)

Other noteworthy facts about relative clauses in western Austronesian languages are as follows. In most languages, relative clauses (like other modifiers) follow their head. In Philippine and Formosan languages, however, they may precede or follow the head (see **Seediq**, **Tsou**, and **Tagalog** for examples). In a few instances, the (semantic) head may also appear within the relative clause, as in the following Tagalog example:

(67) *Wala* *pa* *rín* *yung* *[pinangako*ng
walá' pa rin iyón-ng in-paN-ako'-ng
NEG.EXIST still also DIST-LK RLS(UG)-GER-promise-LK
lamsyed *sa* *akin]*
lamsyed sa akin
lamp LOC 1S.DAT
'The lamp that was promised to me still hasn't come.'

Gil (1994) and Donohue (1999:386f) report internally headed relative clauses for Riau Indonesian and Tukang Besi respectively.

Finally, it may be noted that the functional load of (usually headless) relative clauses in some western Austronesian languages is much higher than in many other languages because they form a regular part of information question and cleft (contrastive focus) constructions as in the following Standard Indonesian example (see also **Buol**, **Javanese**, **Kambera**, **Nias**).

(68) *siapa* *yang* *kau=undang?*
who REL 2s=invite
'Who did you invite?' (Sneddon 1996:316)

3.11 Adnominal possession

As will be recalled from section 1.2, adnominal possessive constructions provide one major criterion for distinguishing symmetrical voice languages from preposed possessor

languages. This section provides further details on the structure and distribution of the relevant constructions.

In symmetrical voice languages, adnominal possessive constructions are straightforward: the possessor generally follows the possessum, often without any intervening grammatical marker (e.g. Sundanese *imah$_1$ paman$_2$ kuring$_3$* 'my$_3$ uncle's$_2$ house$_1$' (Müller-Gotama 2001:36)). If there is a grammatical marker, it is usually a suffixed or enclitic third person possessor pronoun as in Sundanese *imah-**na** paman kuring* 'my uncle's house' (similar to many cognate forms in other Indonesian languages, Sundanese *-na* is no longer strictly a third person possessive pronoun but is also used in definiteness marking and nominalizations (Müller-Gotama 2001:25)). Alternatively, the possessor is preceded by a genitive or oblique proclitic as in Cebuano *maestro sa=bata?* (teacher OBL=child) 'the child's teacher'.

There is one type of exception to the otherwise strict ordering of possessum-possessor in symmetrical voice languages. Quite a number of these languages allow an alternative construction in which pronominal possessors precede the possessum. This is quite regular in Philippine languages where the construction of the form POSSESSUM + ENCLITIC POSSESSIVE PRONOUN (e.g. Tagalog *bahay mo* 'your (sg) house') regularly alternates with a construction of the form DATIVE/OBLIQUE PRONOUN + LINKER + POSSESSUM (e.g. Tagalog *iyong bahay* (2s.DAT-LK house) 'your (sg) house'). In other symmetrical voice languages, this alternative order is more restricted. In Sundanese, for example, it is restricted to the polite first and second person possessive pronouns *pun* and *tuang* (e.g. *pun bojo* 'my wife' (Müller-Gotama 2001:37)). In the Formosan language Paiwan, a special proclitic series exists only for first and second pronouns (e.g. *su=umaq* 'your house' (Egli 1990: 155)). In addition, Paiwan allows all kinds of pronominal possessors to be preposed with a linker intervening (e.g. both *umaq ni=maju* (house GEN=3s) and *ni=maju a umaq* (GEN=3s LK house) are possible for 'his/her house'). The variant with a preposed genitive pronoun appears to be the regular position in Pazeh (Li and Tsuchida 2001:35f) where nominal possessors also generally precede their possessum (e.g. *ni taruat a babizu* (GEN Taruat LK book) 'Taruat's book' (Li and Tsuchida 2001:32)). Li and Tsuchida (2001:35f) suggest that this rather remarkable and exceptional preposing of possessors in a symmetrical voice language could be due to contact with Southern Min.

In preposed possessor languages there are typically two types of possessive constructions, an alienable and an inalienable one. In the alienable construction, the possessor precedes a possessive marker or ligature inflected for person which in turn precedes the possessum as in Buru *ya nango todo* (1s POSS.1s machete) 'my machete' or *ka namo huma* (2s POSS.2s house) 'your house'. The possessor expression is optional, and the inflected possessive ligature, which essentially looks like a preposed possessive pronoun, may appear in reduced form (e.g. *nang todo* instead of *ya nango todo*). As Grimes (1991:279–282) shows for Buru, the possessive ligature sometimes has verb-like characteristics, as it can also occur with typical verbal affixes.

In the inalienable construction, the possessor also precedes the head but there is no ligature intervening in between possessor and possessum. Instead, the possessum is marked with a possessive enclitic (e.g. Buru *fafu olo=n* (pig head=3s.POSS) 'pig's head'). The possessor may be omitted and then the construction looks very much like the typical pronominal possessive construction in symmetrical voice languages (e.g. Buru *olo=m* (head=2s.POSS) 'your head'). This construction is used in Buru primarily for part-whole relationships, including body parts, but not for kinship terms (Grimes 1991:282).

In Buru the distinction between alienable and inalienable is on the constructional level, not on the lexical level (Grimes 1991:287–289). That is, nouns are not subcategorized as

alienable or inalienable. In principle, all nouns can occur in both types of constructions. In fact, the two constructions can occasionally even be merged, as in *nak olo=n* (POSS.3s head=3s.POSS) '(she went away with) her (pig's) head' where the two possessive markers refer to different possessive relationships: the pronominal enclitic *=n* refers to the pig, the 'inalienable' possessor of its own head, and the inflected possessive ligature *nak* refers to the (alienable) owner who in some way acquired a pig's head (Grimes 1991:288).

There is considerable variation regarding the details of the alienable vs. inalienable distinction in preposed possessor languages. In **Leti**, there is actually only a single basic construction (POSSESSOR POSSESSUM + POSSESSIVE SUFFIX) but there is an alienable/inalienable distinction on the lexical level in that inalienable nouns obligatorily take a possessive suffix. In **Tetun Fehan**, the paradigm of the possessive enclitics has been merged into a single genitive marking enclitic *=n* (this has also happened in at least one Buru dialect (Grimes 1991:283)). Furthermore, the pronominal markers precede rather than follow the possessive ligature *=kan*. In Ambai both prefixes and suffixes occur with the ligature, hence *ne-ku* (POSS-1s) 'my' but *u-ne* (3d-POSS) 'their (dual)' (Silzer 1983:124). In **Biak**, which shows perhaps the most complex possessive marking attested in the area, the inflected possessive marker in the alienable construction typically follows rather than precedes the possessum.

Finally, it may be noted that in addition to the position of the possessor and the presence of an alienable vs. inalienable distinction, symmetrical voice languages and preposed possessor languages differ also with regard to another important detail of possessive constructions. Preposed possessor languages typically have pronominal forms (possessive suffixes and inflected possessive markers) which occur exclusively in possessive constructions. In symmetrical voice languages, possessive pronouns typically have at least one additional function: they are also used for actors in undergoer voice constructions (compare Tagalog *tingnán **mo*** (look:LV 2s.POSS) 'look!').

With regard to the position of the possessor, the transitional languages tend to align with the symmetrical voice languages (possessors are usually postposed; a major exception is Banggai in eastern central Sulawesi). With regard to the uses of the possessive pronouns, they often show a closer affinity to the preposed possessor languages in that possessive suffixes or enclitics generally occur only with nouns or nominalized constituents (e.g. Muna, **Nias**, Tukang Besi).

4 MAJOR VERBAL ALTERNATIONS

Western Austronesian languages usually have very little nominal morphology, if any, but they tend to provide for a fairly rich inventory of verbal affixations (in particular Philippine-type languages). The present survey briefly looks at the most widely attested verbal affixations: aktionsart (dynamic vs. non-dynamic), voice, causative and applicative. Zobel (to appear) provides a much more comprehensive survey on these alternations. See **Iloko** for a fuller presentation of a typical Philippine inventory, which usually also includes affixes for requestives, reciprocals, plural actors, etc.

4.1 Dynamic vs. non-dynamic (stative and potentive)

In most western Austronesian languages expressions for non-dynamic events are morphologically overtly marked while those for dynamic events remain unmarked. Expressions for dynamic events typically refer to actions which involve a volitional agent who is in control

of the action (e.g. 'kiss', 'give', 'run', 'throw'). In addition, the dynamic class also includes expressions for meteorological events ('rain', etc.). The two major types of non-dynamic expressions are statives, i.e. states of affairs which in principle do not involve an agent (e.g. 'fall', 'be alive', 'be broken'), and potentives, i.e. events which involve an agent but one who is not in full control of the action. Expressions for processes ('grow', 'turn red', 'freeze', etc.) and spontaneous events ('explode', 'burst', etc.) are fairly variable in their class membership. In some languages they are generally coded as non-dynamic and in others as dynamic, while a third group will allow them to appear in either class. In the remainder of this section we will briefly review some of the forms and meanings typical of non-dynamic expressions.

One major reason to treat statives and potentives together is the fact that in many languages they are marked with the same prefix. One good example is the Acehnese prefix *teu-* which occurs in statives as in (69).

(69) ***teu-****hah* *babah=kah* *sabê*
ST-open mouth=2s always
'Your mouth is always open.' (Durie 1985:73)

The same affix occurs in potentives, which usually convey at least the following three kinds of meanings. First, potentives may refer to *accidental* actions, i.e. the action is done intentionally but the outcome is not the intended one:

(70) *ka=****teu-****koh* *bak=kayee=nyan* *lê=kamoe*
??=POT-cut tree=wood=DIST by=1pe
'We accidentally cut down that tree.' (Durie 1985:76)

Second, potentives are used for *involuntary* actions, i.e. someone happens to do something without having any specific intentions to do so:

(71) *jih* ***teu-****batôk-batôk*
3s POT-RDP-cough
'S/he is coughing a lot.' (Durie 1985:74)

Finally, and somewhat remarkably, potentives may convey an *abilitative* meaning, referring to actions someone is able to do or (just) manages to do:

(72) *batee=nyan* *h'an=****teu-****grak*
stone=DIST NEG=POT-move
'That stone cannot be moved.' (Durie 1985:75)

Not all languages mark all of these event types as non-dynamic, and only some languages follow Acehnese in using a single prefix for these uses. Note also that the most common and widespread non-dynamic prefix is *ma-*.

Also part of the non-dynamic system are so-called adversative passives, which are usually marked by the affix combination *ka--an* (or its cognates) as in Pazeh *ka-udan-an* 'be rained on, be caught in the rain' (<*udan* 'rain') or *ka-lamik-an* 'to catch a cold' (<*lamik* 'cold') (Blust 1999:352). In many Philippine languages, *ka--an* is part of a productive voice alternation paradigm for statives, as seen in **Tagalog**.

4.2 Voice (and aspect/mood)

Two types of voice alternations are attested in western Austronesian languages: passives and symmetrical voice alternations. The latter are also known as *Philippine-type focus* alternations, but *focus* is widely agreed to be a misleading term in this context (cf. Himmelmann 2002b).

4.2.1 Passives and antipassives

A number of western Austronesian languages have voice alternations which are quite similar to the active/passive alternation familiar from European languages. Such passives are characterized by three features: (a) the verb is overtly marked with a passive formative; (b) the undergoer of a transitive verb is the subject of the construction; and (c) the actor may be left unexpressed or if expressed it is clearly marked as oblique, usually by a preposition. Compare the following Standard Indonesian example:

(73) *Orang* *itu* *di-lihat* ***oleh*** *anak* *saya.*
person DIST PV-see by child 1s
'That person was seen by my child.'

This example is almost identical to example (2) above, with the exception of the agent-marking preposition *oleh*. Standard Indonesian is thus a language with both 'standard' passive (as in (73)) and symmetrical voice alternations (as in (2)). This probably also holds true for many symmetrical voice languages other than Philippine-type ones (see Arka (1998) for an analysis of Balinese along these lines).

Passive constructions are also fairly widely attested among transitional languages (e.g. **Makassar**). In **Mori Bawah** the agent argument is obligatorily deleted. In preposed possessor languages, passives are rare (but see **Biak**).

Antipassives are only rarely attested in the region. See **Mori Bawah** for one of the very few good examples. Of course, proponents of ergative analyses for western Austronesian languages (see 3.8.2 above) consider all actor voice constructions to be antipassives.

4.2.2 Symmetrical voice alternations

As mentioned in section 1.2, more than half of the western Austronesian languages are characterized by the occurrence of symmetrical voice alternations. There is a substantial body of literature concerning these alternations, beginning as early as Humboldt (1838: 347 passim). See Constantino (1971) and Matsuda French (1988) for surveys of the older literature, and Sells (1997) and the contributions in Klamer (1996) and Austin and Musgrave (to appear) for more recent discussion.

Basic morphosyntactic properties of symmetrical voice alternations are discussed in sections 3.3 and 3.8. In the current section we briefly review three major parameters of morphological variation: (a) the number of alternations; (b) the interaction with aspect/mood; and (c) the integration of person markers into the paradigm of voice alternations.

The typical Philippine and Formosan system, which is also attested in northern Sulawesi and Sabah (northern Borneo), has four alternations, as seen in the following Cebuano examples:

Actor voice

(74) *akú=y* ***mu-****palít* *ug* *isda?*
1s=TOP AV-buy OBL.NSPEC fish
'I will buy some fish.' (Wolff 1972:xv)

Patient voice

(75) *palit-**ún*** *ku* *ang* *isda?*
buy-PV 1S.POSS SPEC fish
'I will buy the fish.' (Wolff 1972: xv)

Locative voice

(76) *bantay-**án*** *ninyú* *ang* *prisu*
watch-LV 2P.POSS SPEC prisoner
'You will watch the prisoner.' (Wolff 1972:38)

(77) *palit-**án*** *ku* *siyá=g* *kík*
buy-LV 1S.POSS 3S=OBL.NSPEC cake
'I will buy some cake for/from her.' (Wolff 1972:38)

Conveyance voice

(78) ***i**-butáng* *niya* *ang* *kwarta*
CV-put.down 3S.POSS SPEC money
'S/he will put the money down.' (Wolff 1972:361)

(79) *wa?* *ku=y* *kwarta-ng* ***i**-palít* *ug* *bugás*
NEG.EXIST 1S=LK money-LK CV-buy OBL.NSPEC rice
'I have no money to buy rice with.' (Wolff 1972:361)

The use of these voices is discussed in the chapters on **Seediq**, **Iloko**, **Tagalog**, **Kimaragang**, **Sama**, **Buol** and **Malagasy**. Note that there is a broad range of terms in use for the different voices. But the forms are usually easily recognizable since there is only very little cross-linguistic variation.

It is not uncommon to find references to more than four voice alternations in these languages in the literature. To give just one example, up to eleven voice alternations have been proposed for Tagalog (Schachter and Otanes 1972:344). However, in most instances the additional voice alternations are morphologically complex, consisting of a stem-forming affix plus a basic voice affix (e.g. the Tagalog 'instrumental voice' prefix *ipaN-* consists of the stem-forming prefix *paN-* and the conveyance voice prefix *i-*). More often than not, such formally complex voice affixes are also semantically compositional and there is little reason to consider them unit morphemes. In a few instances, however, there may be good morphological reasons for proposing a fifth basic voice (see **Iloko** (section 5.1), **Sama** (section 4.1) and **Kimaragang** (sections 3.1 and 4)).

The basic voices usually occur in a number of different aspects and moods. One type of system widely attested in the central Philippines consists of three moods: non-realis, realis, and subjunctive, as seen for Cebuano in Table 5.5. This table also shows that voice

TABLE 5.5: CEBUANO VOICE-MOOD PARADIGM FOR DYNAMIC VERBS (cf. Wolff 1972:xvi, 2001:123)

	Non-realis	Realis	Subjunctive
Actor voice	mu-	mi-/ni-	mu-
Patient voice	-un	gi-	-a
Locative voice	-an	gi--an	-i
Conveyance voice	i-	gi-	i-

and mood marking are closely linked to each other, forming a paradigm. In most instances, it is impossible to formally separate a voice formative from a mood formative. Instead, most formatives represent a combination of these categories.

Voice-aspect-mood paradigms with four or five voice distinctions are formally heterogeneous in that they involve a mixture of prefixes, suffixes and very often also infixes. There are usually a number of different formatives for actor voice (e.g. Cebuano also has *mag-* and *maN-*) while the undergoer voices are represented by just one set of formatives. Patient voice is usually suffixless in realis mood but marked by a suffix in other moods. The voice affixes occur with both semantically transitive and intransitive lexical bases, with only a few lexical bases occurring unaffixed when referring to events or states.

The paradigmatic organization is also evident from the fact that in some instances one formative represents two voice-mood categories. Thus, Cebuano *gi-* represents both realis patient and realis conveyance voice. Only the alternations make it clear which voice-mood category it represents in a given instance (in patient voice, realis *gi-* alternates with non-realis *-un*, in conveyance voice, it alternates with non-realis *i-*). Such multifunctional formatives are found in practically all symmetrical voice-mood paradigms, as amply illustrated by the sketches for **Seediq**, **Iloko**, **Tagalog**, **Buol**, **Kimaragang**, and **Sama**. See also Himmelmann (to appear b) for further discussion.

The subjunctive forms are commonly used in imperatives, in some subordinate clause types, after certain adverbials, and (more sporadically) in narrative sequences (see section 5). The details vary from language to language. In Cebuano, only the undergoer voice forms are used in imperatives as in:

(80) *Bantay-**í*** *uná?* *siyá* *ha?*
watch.over-LV.SUBJ then 3S okay?
'Take good care of her/him, will you?' (Wolff 1966:440)

The adverbials which induce subjunctive mood in Cebuano mark all reference to time and include expressions such as *kagahapun* 'yesterday', *anus?a* 'when (with future time reference)', and *sa miaging Duminggu* 'last Sunday' as well as tense-marked negators and deictics:

(81) *wa?* *niya* *saky-**í*** *ang* *taksi*
NEG.PST 3S.POSS ride.on-LV.SUBJ SPEC taxi
'He did not ride in the taxi.' (Zorc 1977:151)

In a number of languages, especially in Taiwan (e.g. **Seediq**) but also in **Buol**, there are two non-indicative moods rather than just one (the subjunctive). Ross (2002 and elsewhere), who reconstructs these two moods for Proto-Austronesian, calls them *atemporal* and *projective* respectively (see also ADELAAR, A HISTORICAL PERSPECTIVE).

In symmetrical voice languages which do not belong to the Philippine-type, the most common system of symmetrical voice alternations is a simple two-way alternation between actor voice and undergoer voice, as in the Standard Indonesian examples (1) and (2) above (see also **Belait**, **Karo Batak**, **Javanese**). There are also a few languages with three voice alternations (e.g. **Malagasy**). Zobel (to appear) provides a comprehensive survey of the attested systems.

A major characteristic of the symmetrical voice languages found in Indonesia (with the exception of northern Sulawesi) pertains to the fact that pronominal proclitics or prefixes have been integrated into the system of voice markers. In Da'a, a central Sulawesi language, for example, the non-realis undergoer voice prefix is *ra-* as seen in

the following example:

(82) *Loka* *etu* *ma-tasa* *kana* ***ra-koni.***
banana DIST ST-ripe must UV-eat
'(When) that banana is ripe it must be eaten.' (Barr 1988:21)

However, if the actor of an undergoer voice verb is first or second person singular, then *ra-* is replaced by the proclitic person markers *ku=* or *mu=*, as in:

(83) *Da'a* *ma-mala* *aku* ***mu=raga.***
not ST-able 1s 2s.ACT-chase
'You can't chase me.' (Barr 1988:40)

As in the case of Da'a, these pronominal markers are often restricted to the first or second person. If a mood difference is marked on the predicate, the pronominal markers always occur in the non-realis forms. See also **Karo Batak**, **Colloquial Indonesian**, **Javanese**, and Zobel (to appear).

4.3 Causative and applicative

A causative formation involving the prefix *pa-* (or a cognate form) is probably the most widely attested productive morphological derivation in western Austronesian languages. Notable exceptions include Keo and **Biak** which lack productive morphological causatives altogether, and **Malayic varieties** and Modern **Javanese** which allow morphological causative derivations but primarily with suffixes rather than with a prefix (see below). Causatives are mentioned and illustrated in practically all language sketch chapters, but see in particular **Seediq**, **Kimaragang**, **Belait**, **Malagasy**, **Buol**, **Nias** and **Kambera** for examples and discussion.

Most western Austronesian languages, with the exception of Philippine-type languages, also have applicative morphology and constructions (**Tetun** and **Leti** are exceptions among the preposed possessor languages). There are typically two applicative suffixes, one (*-i* or a cognate form) for locative applicatives and one (often *-akan* or a cognate form) covering a broader range of semantic roles, usually including instruments and beneficiaries. Both suffixes also regularly occur in causative derivations (e.g. **Old Malay**, **Colloquial Indonesian**, **Javanese**, **Nias**). See **Mori Bawah** and **Kambera** for more extensive exemplification of applicatives and Sirk (1996) for historical discussion.

Although the morphosyntax is quite different, there is no doubt that the conveyance and locative voice alternations in Philippine-type languages have much in common semantically with applicative alternations in the other languages. There is furthermore a formal similarity in that the most widely attested applicative formative *-i* is also widely attested as (subjunctive) locative voice suffix (in the recent comparative literature (e.g. Wolff 1996, Ross 2002) it is sometimes assumed that the two suffixes are historically speaking identical but this is a conjecture rather than a proven fact). Another related, but not identical phenomenon appears to be the *affectedness alternation* described by Kroeger for **Kimaragang**.

One crosslinguistically remarkable feature of both causative and applicative derivations in western Austronesian languages is that these derivations are not always valency-increasing but convey a broader range of sometimes quite elusive meanings, including intensity and iterativity (cf. **Mori Bawah** and **Kambera**). Once again, Philippine-type languages would appear to be somewhat exceptional in that their causative *pa-* prefix

generally only conveys causation. Since intensity and iterativity are often conveyed by the derivational prefixes *pag-* and *paN-* in Philippine-type languages, it may well be the case that what is described as polysemy of causative *pa-* for a number of transitional and preposed possessor languages is actually a case of homophony, due to a merger of **pa-*, **paR-* and **paN-*.

5 CLAUSE LINKAGE PATTERNS AND ANAPHORA

In all western Austronesian languages, there are few (if any) morphosyntactic constraints on the omission of coreferential arguments in clause sequences. That is, the possibility to omit a coreferential argument is not restricted to subject arguments, as in the following Balinese example where both *nuut* and *rauh* occur without overt subject expressions (see also example (50) above):

(84) *lantas* *ida* *malayar* *nuut* *pasisi.*
then 3.HON INTR:sail AV:follow beach
Rauh *di* *pasisi* *Pajarakan-ne*
arrive LOC beach Pajarakan-DEF
'...then she sails along the beach. When she arrives on the beach of Pajarakan, ...'
(Pastika 1999, chapter 7)

Instead, it is also possible to omit all kinds of non-subject arguments. In the following Tetun Dili example, the subject and the non-subject patient of *hemu* 'drink' are omitted:

(85) *nia* *halo* *kafé,* *hemu.*
3s make coffee drink
'He makes coffee and drinks it.' (Hull 1999:5)

In the following Tagalog example (originally from Martin 1981:313), the recipient of the predicate *ibíbigay* 'give to' in the second clause remains unexpressed:

(86) *kung* *ma-ki-kita* *ko* *siyá,* *i-bí-bigáy* *ko*
if POT-RDP-seen 1s.POSS 3s CV-RDP-gif 1s.POSS
ang *sulat* *mo*
SPEC letter 2s.POSS
'If I see him, I will give [him] your letter.' (Kroeger 1993:34)

Overtly expressed, the recipient would have to be marked with the general locative preposition *sa*, hence the 'complete' version of the second clause would read *ibíbigáy ko* ***sa kaniyá*** *ang sulat mo*.

This (syntactically) free omissibility of argument expressions accounts for the fact that the coordinate conjunction reduction tests widely used in the typological literature on grammatical relations are generally inconclusive in western Austronesian languages (cf. Kroeger 1993:36, Cena 1995:15–18, and section 3.8.1 above).

Given that all kinds of argument expressions are freely omissible, it is somewhat remarkable that actor expressions in undergoer voice constructions are rarely omitted in symmetrical voice languages (cf. Shibatani 1988:93, Himmelmann 1999). Compare the

following Balinese example:

(87) *raris* *kenten* *pepetan* *padi-n-e* *nika* *kaat=**a***
then like.that husks rice-LK-DEF DIST (UV:)cut.off=3.ACT
*jang=**a*** *samping* *jineng-e*
(UV:)put=3.ACT beside rice store-DEF
nika *tunjel=**a*** *pepet=ne* *nika*
DIST (UV:)burn=3.ACT husks=3.POSS DIST
'Then she cut off those rice husks, put them beside the rice store, and burned them, those husks.' (Pastika 1999, chapter 7)

In this action sequence, all three verbs (*kaat*, *jang*, *tunjel*) are in undergoer voice and have the same actor ('she') and undergoer ('rice husks'). But while the undergoer is not mentioned in the second clause, the third person actor clitic=*a* occurs on all predicates. In transitional and preposed possessor languages, transitive actors are also generally overtly expressed in the sense that in many of these languages the use of subject marking proclitics or prefixes is obligatory.

Related to the preceding phenomenon, there is a remarkable pattern of clause linkage which is highly characteristic for symmetrical voice languages and somewhat unusual from a cross-linguistic point of view. Compare the following example from Balinese (and examples (40) and (41) in the **Tagalog** chapter):

(88) *lantas* ***ia*** ***nyemak*** *punggalan-e* *tur*
then 3 AV:take head-DEF and
entung-ang=**a *ke* *tukad-e*
(UV:)throw-APP=3.ACT to river-DEF
'then he took the (severed) head and threw it into the river.' (Clynes 1995:296)

This example consists of two transitive clauses which share the same set of core arguments. The first clause is in actor voice, the actor (*ia*) being the subject and the undergoer (*punggalane*) the non-subject argument. In the second, undergoer voice clause, the linking relations are switched around: the actor (=*a*) is the non-subject argument and the undergoer the subject argument (which is usually no longer overtly expressed, as noted above).

This pattern is found in sequences of main clauses (which do not have to be overtly coordinated) as well as in sequences consisting of a subordinate and a main clause (in either order), as seen in the following Balinese example:

(89) *Mara* *ia* *ningeh* *pamunyin* *panak=ne* *buka* *keto,*
when 3_i AV:hear voice $child_j$=3.POSS as like.that
dadi *ampak-in=**a*** *dogen* *ia* *jelanan.*
thus (UV:)open-APP=3.ACT_i just 3_j door
'When she heard the voice of her child like that, she opened the door for him.' (Pastika 1999, chapter 7)

These examples are remarkable from a cross-linguistic point of view since a similar switch in construction is not attested in other language types. A typical nominative-accusative language, for example, would use two active voice clauses as in the English translations of the two preceding examples.

The examples are also noteworthy with regard to the definiteness constraint on actor voice constructions already briefly discussed in section 3.6 above. It is often claimed that actor voice constructions generally disallow definite undergoers. Examples (88) and (89) show that this is not true. In both examples, the first clause is an actor voice construction but it still involves a definite undergoer (*punggalane* in (88) and *pamunyin panakne* in (89)). Unlike a number of other 'exceptions' to the definiteness constraint, there are no grammatical reasons for the choice of an actor voice construction (such as the occurrence of an actor voice construction in a relative clause, see **Tagalog** and **Karo Batak** for examples and further discussion). Consequently, the definiteness constraint in the strong form it is usually given is certainly wrong in relation to symmetrical voice languages.

Nevertheless, examples (88) and (89) are also good illustrations of the extent to which the definiteness constraint actually holds true. In both examples, it is practically impossible to use actor voice in the second clause. It is not yet clear whether this is a rather strong pragmatic or stylistic preference which allows the odd exception or whether it in fact constitutes a grammatical constraint on clause linkage. Note that the constraint holds between clauses regardless of their syntactic relationship (coordination, subordination, or simple juxtaposition).

A clearly more grammaticized pattern requiring a constructional switch in clause sequences pertains to certain uses of the subjunctive in a number of Philippine-type languages, mostly in Sabah (e.g. **Kimaragang**) and Taiwan (e.g. Paiwan, possibly also **Seediq**). In these languages, use of the subjunctive is required in 'narrative sequences', i.e. action sequences involving the same set of participants. In some languages, including Paiwan, use of the subjunctive is triggered by certain coordinators such as *sa* in the following example:

(90) *pacun-an* *a* *zu'* *a* *gang,* *qucə-quc-ən*
see-LV SPEC DIST LK crab RDP-crush-PV
sa *kan-**i*** *aya*
and.then eat-PV.SUBJ thus
'... he saw the crabs, and crushed and ate them.' (Ross 2002:23, originally from Egli 1990:330)

In this example, the subjunctive only occurs in the last predicate, which is preceded by *sa*. In other languages, subjunctive forms occur throughout the narrative sequence (see example (52) in **Kimaragang**).

6 A FINAL LOOK AT TYPOLOGICAL DIVERSITY WITHIN WESTERN AUSTRONESIAN LANGUAGES

The preceding sections have surveyed some of the major typological characteristics of western Austronesian languages, both from an internal and an external point of view. The survey is obviously far from complete. Among the topics which have been omitted for lack of space are the following ones.

(i) Numeral classifiers

Many western Austronesian languages have small to medium-sized inventories of numeral classifiers, usually involving a special classifier construction. See **Sama**, **Belait**, **Cham**, **Moken/Moklen**, **Mori Bawah**, **Nias**, **Kambera**, **Tetun Fehan**, and **Taba** for examples and discussion. This list already indicates that numeral classifiers are found in both symmetrical voice and preposed possessor languages as well as transitional languages. They appear to be lacking in the Philippine-type languages and Taiwan.

(ii) Plural marking

There is a wide variety of usually optional plural markers for nominal expressions. Next to reduplication (see section 2.2), there are plural words (cf. Dryer 1989) such as **Javanese** *padha* or Tagalog *manga* (orthographically *mga*) and third person plural pronouns, which are used to indicate plural in **Tetun Fehan** (and Tetun Dili) and Buru (Grimes 1991:159), for example. Associative plurals, which refer to a group of people associated with a given person (e.g. Tagalog *sina Maria* 'Maria and company/the others'), are widely attested in Philippine-type languages (more sporadically also elsewhere, for example in Ambon Malay and **Nias**). There is also the possibility to indicate plurality in adjectives (usually by reduplication) and verbs (by prefixes). Apart from a number of languages in western Indonesia, where elaborate politeness rules have led to considerable changes in the pronoun systems (e.g. **Javanese**, Balinese), most western Austronesian languages also clearly distinguish plural from singular pronouns.

(iii) Deictics and reference to space

Western Austronesian languages differ significantly with regard to the basic structure of their (local) deictic systems. Parameters of variation include the number of degrees of distance which are distinguished in a given system (often two (proximal/distal) or three (proximal/medial/distal), but four and five are also attested; **Malagasy** is in fact famous for distinguishing seven different degrees). Visibility plays a role in, for example, **Tsou**, **Seediq**, **Iloko**, **Malagasy**, and **Leti**. In a number of languages, including **Malagasy** and **Iloko**, local deictics are obligatorily tense-marked. In **Leti** the deictics also convey speaker's attitude. Furthermore, in many languages, the basic deictics are closely linked to a system of directionals which specify locations and movements as being UP, DOWN, ACROSS or the like. See in particular **Buol**, **Mori Bawah**, **Taba** and **Biak** for examples. The wider systems of spatial orientation in Austronesian languages have only recently been explored (cf. the pioneering studies edited by Senft 1997 and Bennardo 2002).

(iv) Nominalization/subordination

Western Austronesian languages usually have a few derivational affixes to nominalize action expressions, regardless of whether or not nouns and verbs are clearly distinguished morphosyntactically. The most widespread derivations include instrumental nominalizations such as Pazeh *saa-kudung* 'hammer' (cf. *mu-kudung* 'hit with a hammer', Li and Tsuchida 2001:51), abstract qualities such as Muna *ka-ghosa* 'strength' from *ghosa* 'strong' (Van den Berg 1989:294) and gerunds/action nominalizations such as Ratahan *pangangaaq* 'action of taking/upon taking' from *a(l)aq* 'take' (Himmelmann and Wolff 1999:71). The voice affixes found in Philippine-type languages also have nominalizing functions, and in these functions they are found in many other languages as well. Nominalizing morphology is also widely found in subordinate constructions such as relative, complement and adverbial clauses. Almost every language sketch chapter in this volume provides further exemplification for these observations.

Finally, the proposed typological grouping for western Austronesian languages is also in need of further testing and refinement. Table 5.6 summarizes some of the characteristics of the two major types, symmetrical voice languages and preposed possessor languages.

TABLE 5.6: CHARACTERISTIC FEATURES OF SYMMETRICAL VOICE AND PREPOSED POSSESSOR LANGUAGES

Symmetrical voice languages	Preposed possessor languages	See section
Symmetrical voice alternations	No or asymmetrical voice alternations	3.3, 3.8, 4.2
Postposed possessor	Preposed possessor	3.11
No alienable/inalienable distinction	Alienable/inalienable distinction	3.11
Few or no differences between narrative and equational clauses	Clear-cut differences between narrative and equational clauses	3.4, 3.6
Person marking only sporadically attested	Person marking prefixes or proclitics for S/A arguments	3.7
Numerals/quantifiers precede head	Numerals/quantifiers follow head	3.5
Negators in pre-predicate position	Clause-final negators	3.5
V-initial or SVX	V-second or -final	3.5

With regard to refinements, it is certainly possible to distinguish further subgroups within these larger groupings. One well established and widely used typological subgroup within the symmetrical voice languages is the *Philippine-type languages*. In addition to the defining features already mentioned in section 1.2 (i.e. multiple undergoer voices, paradigms of phrase marking proclitics, pronominal second position clitics), these languages share a number of other features, including productive infixation, honorific articles, a realis/non-realis distinction, and the lack of (numeral) classifiers (these features are of course also found outside the Philippine-type languages). Nevertheless, despite the fact that these languages are very similar indeed in many respects one should be very careful when generalizing from one language over the whole group (practically every article on a Philippine-type language at some point starts to make claims for all Philippine-type languages). Standard Tagalog, for example, is in a number of regards not at all representative of Philippine-type languages (most importantly, perhaps, it lacks subjunctive affixation; see also section 3.6).

A second subtype of symmetrical voice language is the *Indonesian-type languages* (see Wolff 1996, Ross 2002, Zobel 2002 and to appear). Defining features for this subtype are the combination of a symmetrical voice system (with two or three symmetrical voice alternations), applicative morphology and the integration of proclitic actor markers into the paradigm of voice markers. This definition, which is narrower but also more precise than the ones employed in the literature to date, covers roughly the symmetrical voice languages of western Indonesia (including **Javanese**, **Colloquial Indonesian**, and **Karo Batak**), and central Sulawesi (Tomini-Tolitoli and Kaili-Pamona languages) as well as Chamorro and Palauan. It remains to be seen whether this is indeed a useful typological grouping in the sense that there are further features correlating with the defining features.

It is questionable whether the *Formosan languages* form a useful typological grouping, as it is sometimes implied in the literature. All Formosan languages are symmetrical voice languages but not all of them are also Philippine-type languages (see section 1.2 above). Furthermore, even for the Formosan languages which clearly are Philippine-type languages (e.g. Paiwan, Atayal, **Seediq**, Puyuma, Pazeh) it is not at all clear whether they share features which set them as a group clearly apart from other Philippine-type languages. In fact, it would appear that the Formosan languages are morphosyntactically much more heterogeneous than, for example, the northern or central Philippine languages.

Among the preposed possessor languages, the Austronesian languages of Timor form a special subtype due to the rather strong isolating tendencies found in many of them. These

are shared with some languages of Flores (and the close-by islands Solor, Adonara, and Lembata), which are transitional languages according to the definitions employed here. However, once again more detailed investigations are required in order to see whether these languages share enough characteristics to warrant a (sub-)type of their own.

ACKNOWLEDGMENTS

I would like to express my gratitude to Sander Adelaar for his valuable comments on an earlier version of this chapter and to John Hajek for comments on the phonology section. Thanks also to Markus Greif, Emina Kurtić and Jan Strunk for feedback from a student's point of view.

REFERENCES

Adelaar, K.A. (1992) *Proto-Malayic: a reconstruction of its phonology and part of its morphology and lexicon*, Canberra: Pacific Linguistics.
——(2000) 'Siraya Reduplication', *Oceanic Linguistics* 39: 33–52.
Ameka, F., Dench, A. and Evans, N. (eds) (to appear) *Catching Language*, Berlin: Mouton de Gruyter.
Arka, I Wayan (1998) *From morphosyntax to pragmatics in Balinese*, PhD Thesis, University of Sydney.
Arka, I. Wayan and Manning, C.D. (to appear) 'Voice and grammatical relations in Indonesian: A new perspective', in P. Austin and S. Musgrave (eds).
Artawa, Ketut, Artini, Putu and Blake, B.J. (2001) 'Balinese Grammar and Discourse', in Austin et al (eds), 11–46.
Austin, P. and Musgrave, S. (eds) (to appear) *Voice and grammatical functions in Austronesian languages*, Stanford: CSLI.
Austin, P., Blake, B. and Florey, M. (eds) (2001) *Explorations in valency in Austronesian languages*, La Trobe Papers in Linguistics 11, Bundoora: La Trobe University.
Baird, L. (2002) *A Grammar of Keo: An Austronesian language of East Nusantara*, PhD Thesis, Australian National University.
Barr, D.F. (1988) 'Da'a Verbal Affixes and Clitics', in H. Steinhauer (ed.), 11–49.
Bennardo, G. (2002) *Representing Space in Oceania: Culture in Language and Mind*, Canberra: Pacific Linguistics.
Blake, B.J. (1988) 'Tagalog and the Manila-Mt Isa Axis', *La Trobe Working Papers in Linguistics* 1:77–90.
Blake, F.R. (1925) *A Grammar of the Tagalog Language*, New Haven: American Oriental Society.
Bloomfield, L. (1917) *Tagalog Texts with Grammatical Analysis*, 3 vols., Urbana, Ill.: The University of Illinois.
Blust, R.A. (1988) *Austronesian root theory. An essay on the limits of morphology*, Amsterdam: Benjamins.
——(1995) 'Notes on Berawan consonant gemination', *Oceanic Linguistics* 34:123–138.
——(1997a) 'Nasals and nasalization in Borneo', *Oceanic Linguistics* 36:149–179.
——(1997b) 'Ablaut in Western Borneo', *Diachronica* XIV:1–30.
——(1998) 'Ca-Reduplication and Proto-Austronesian Grammar', *Oceanic Linguistics* 37:29–64.
——(1999) 'Notes on Pazeh phonology and morphology', *Oceanic Linguistics* 38:321–364.
——(2001) 'Thao Triplication', *Oceanic Linguistics* 40:324–335.

Bowden, J. (2001) *Taba: Description of a South Halmahera language*, Canberra: Pacific Linguistics.
Capell, A. (1964) 'Verbal systems in Philippine languages', *Philippine Journal of Science* 93:231–249.
Carpenter, K.L. (1996) 'Subordination by reduplication in Wetan', *Studies in Language* 20:37–51.
Cena, R.M. (1995) 'Surviving without relations', *Philippine Journal of Linguistics* 26:1–32.
Chang, M.L. (1998) 'Thao reduplication', *Oceanic Linguistics* 37:277–297.
Clynes, A. (1995) *Topics in the Phonology and Morphosyntax of Balinese*, PhD Thesis, Australian National University.
——(1997) 'On the Proto-Austronesian 'Diphthongs', *Oceanic Linguistics* 36:347–361.
Cohn, A.C. (1990) *Phonetic and phonological rules of nasalization*, PhD Thesis, Univ. of California Los Angeles.
Collins, J.T. (1983) *The Historical Relationships of the Languages of Central Maluku, Indonesia*, Canberra: Pacific Linguistics.
Comrie, B. (1979) 'Definite and animate direct objects: a natural class', *Linguistica Silesiana* 3:13–21.
Constantino, E. (1971) 'Tagalog and other major languages of the Philippines', in T.A. Sebeok (ed.), *Current Trends in Linguistics*, vol. 8/1: 112–154) The Hague: Mouton.
Coward, D. and Coward, N. (2000) 'A phonological sketch of the Selaru language', in C.E. Grimes (ed.), 9–54.
Cumming, S. (1991) *Functional Change. The Case of Malay Constituent Order*, Berlin: de Gruyter.
Cumming, S. and Wouk, F. (1987) 'Is there "Discourse Ergativity" in Austronesian Languages?', *Lingua* 71:271–296.
Davies, W. (1999a) *Madurese*, München: Lincom (Languages of the World/Materials 184).
——(1999b) 'Madurese and Javanese as strict word order languages', *Oceanic Linguistics* 38:152–167.
DeGuzman, V.P. (1978) 'A case for nonphonological constraints on nasal substitution', *Oceanic Linguistics* 17:87–106.
——(1988), 'Ergative Analysis for Philippine Languages: An Analysis', in R. McGinn (ed.), *Studies in Austronesian linguistics*, 323–345, Athens, Ohio: Ohio University Monographs in International Studies.
DeWolf, C.M. (1979) *Sentential Predicates: A Cross-Linguistic Analysis*, PhD Thesis, University of Hawaii.
——(1988) 'Voice in Austronesian languages of Philippine type: passive, ergative, or neither?', in Shibatani (ed.), 143–193.
Dixon, R.M.W. (1994) *Ergativity*, Cambridge: Cambridge University Press.
Donohue, M. (1999) *A grammar of Tukang Besi*, Berlin/New York: Mouton De Gruyter.
——(2002) 'Voice in Tukang Besi and the Austronesian focus system', in Wouk and Ross (eds), 81–99.
Donohue, M. and Smith, J.C. (1998) 'What's happened to us? Some developments in the Malay pronoun system', *Oceanic Linguistics* 37:65–84.
Dryer, M.S. (1989) 'Plural words', *Linguistics* 27:865–895.
DuBois, J.W. (1987) 'The discourse basis of ergativity', *Language* 63:805–955.
Durie, M. (1985) *A grammar of Acehnese on the basis of a dialect of North Aceh*, Dordrecht: Foris.
——(1987) 'Grammatical Relations in Acehnese', *Studies in Language* 11:365–399.
Egerod, S. (1988) 'Thoughts on Transitivity', *Bulletin AS* 59:369–384.
Egli, H. (1990) *Paiwangrammatik*, Wiesbaden: Harrassowitz.
English, L.J. (1986) *Tagalog–English dictionary*, Manila: National Book Store.
Evans, N. (2000) 'Word classes in the world's languages', in G. Booij, C. Lehmann and J. Mugdan (eds), *Morphologie – Morphology. An international handbook on inflection and word-formation*, Berlin: de Gruyter, 708–732.

Florey, M. (2001) 'Verb and valence in Alune', in Austin *et al.* (eds), 73–120.
Foley, W.A. (to appear) 'The place of Philippine languages in a typology of voice systems, in P. Austin and S. Musgrave (eds).
Foley, W.A. and Van Valin, R. (1984) *Functional Syntax and Universal Grammar*, Cambridge: Cambridge University Press.
Friberg, B. (1996) 'Konjo's peripatetic person markers', in H. Steinhauer (ed.) 137–171.
Gil, D. (1994) 'The structure of Riau Indonesian', *Nordic Journal of Linguistics* 17: 179–200.
——(2000) 'Syntactic categories, cross-linguistic variation and universal grammar', in P. Vogel and B. Comrie (eds), *Approaches to the typology of word classes*, 173–216, Berlin: de Gruyter.
Grimes, C.E. (1991) *The Buru language of Eastern Indonesia*, PhD Thesis, Australian National University.
——(2000) 'Introduction: new information filling old gaps in eastern Indonesia', in Grimes (ed.), 1–8.
——(ed.) (2000) *Spices from the East. Papers in languages of eastern Indonesia*, Canberra: Pacific Linguistics.
Haaksma, R. (1933) *Inleiding tot de Studie der Vervoegde Vormen in de Indonesische Talen*, Leiden: Brill.
Hajek, J. and Bowden, J. (2002) 'A phonological oddity in the Austronesian area: ejectives in Waimoa', *Oceanic Linguistics* 41:222–224.
Halle, M. (2001) 'Infixation vs. onset metathesis in Tagalog, Chamorro, and Toba Batak', in M.J. Kenstowicz (ed.), *Ken Hale: a life in language*, 153–168, Cambridge, Mass.: MIT Press.
Heath, J. (1975) 'Some functional relationships in grammar', *Language* 51:89–104.
Himmelmann, N.P. (1991) *The Philippine Challenge to Universal Grammar*, Arbeitspapier 15, Köln: Institut für Sprachwissenschaft.
——(1997) *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*, Tübingen: Niemeyer.
——(1998) 'Regularity in irregularity: Article use in adpositional phrases', *Linguistic Typology* 2:315–353.
——(1999) 'The lack of zero anaphora and incipient person marking in Tagalog', *Oceanic Linguistics* 38:231–269.
——(2001) *Sourcebook on Tomini-Tolitoli Languages. General Information and Word Lists*, Canberra: Pacific Linguistics.
——(2002a) 'Voice in two northern Sulawesi languages', in Wouk and Ross (eds), 123–142.
——(2002b) 'Voice in Western Austronesian: An update', in Wouk and Ross (eds), 7–16.
——(to appear a) 'Lexical categories and voice in Tagalog', in P. Austin and S. Musgrave (eds).
——(to appear b) 'How to miss a paradigm or two: Multifunctional *ma-* in Tagalog', in F. Ameka *et al.* (eds).
Himmelmann, N.P. and Wolff, J. (1999) *Toratán (Ratahan)*, München: Lincom.
Hogan, D. (1999) *Urak Lawoi'*, München: Lincom.
Hughes, J. (2000) 'The morphology of Dobel, Aru, with special reference to reduplication', in C.E. Grimes (ed.), 131–180.
Hull, G. (1999) *Mai Kolia Tetun: a course in Tetum-Praça, the lingua franca of East Timor*, 3rd edition, Sydney: Caritas Australia.
——(2001) 'A morphological overview of the Timoric Sprachbund', *Studies in Languages and Cultures of East Timor* 4:98–205.
——(2002) *Waimaha (East Timor Languages Profiles No. 2)*, Dili: Instituto Nacional de Linguística.
Hull, G. and Eccles, L. (2001) *Tetum Reference Grammar*, Winston Hills: Sebastião Aparício da Silva Project.

Humboldt, W. von (1838) *Über die Kawi Sprache*, Bd. II, Berlin: Dümmler.
Jonker, J.C.G. (1914) 'Kan men bij de talen van den Indischen archipel eene westelijke en eene oostelijke afdeeling onderscheiden?', *Verslagen en mededeelingen der Koninklijke Akademie van Wetenschappen, Afdeeling Letterkunde* (4) XII:314–417.
Keenan, E.L. (1976) 'Towards a Universal Definition of 'Subject'', in Li (ed.), 305–333.
Keenan, E.L. and Comrie, B. (1977) 'Noun phrase accessibility and universal grammar', *Linguistic Inquiry* 8:63–100.
Klamer, M. (ed.) (1996) *Voice in Austronesian*, NUSA 39, Jakarta: Universitas Atma Jaya.
Klimov, G.A. (1977) *Tipologija jazykov aktivnogo stroja*, Moskau: Nauka.
Kroeger, P.R. (1993) *Phrase Structure and Grammatical Relations in Tagalog*, Stanford: Stanford University Press.
LaPolla, R.J. and Poa, D. (to appear) 'On describing word order', in F. Ameka *et al.* (eds).
Lazard, G. (2001) 'Le marquage différentiel de l'objet', in M. Haspelmath, E. König, W. Oesterreicher and W. Raible (eds), *Language Typology and Language Universals*, 873–885, Berlin: de Gruyter.
Lemaréchal, A. (1991) 'Dérivation et orientation dans les langues des Philippines', *Bulletin de la Société de Linguistique de Paris* 86:317–358.
Li, C.N. (ed.) (1976) *Subject and Topic*, New York: Academic Press.
Li, P.J. (1973) *Rukai Structure*, Taipei: Academia Sinica.
Li, P.J. and Tsuchida, S. (2001) *Pazeh Dictionary*, Language and Linguistics Monograph Series No A2, Taipei: Academia Sinica.
Lichtenberk, F. (1991) 'Semantic change and heterosemy in grammaticalization', *Language* 67:475–509.
Lopez, C. (1937) 'Preliminary Study of the Affixes in Tagalog', in E. Constantino (ed.) (1977) *Selected writings of Cecilio Lopez in Philippine linguistics*, 28–104, Diliman: University of the Philippines Press.
Martens, M. (1988a) 'Focus or Ergativity? Pronoun sets in Uma', in H. Steinhauer (ed.), 263–277.
——(1988b) 'Focus and Discourse in Uma', in H. Steinhauer (ed.), 239–256.
Martin, J.R. (1981), 'Conjunction and continuity in Tagalog', in M.A.K. Halliday and J.R. Martin (eds), 1981, *Readings in Systemic Linguistics*, 310–336, London: Batsford.
Matsuda French, K. (1988) 'The focus system in Philippine languages: A historical overview', *Philippine Journal of Linguistics* 18/19:1–29.
McFarland, C.D. (1978) 'Definite objects and subject selection in Philippine languages', in C. Edrial-Luzares and A. Hale (eds), *Studies in Philippine Linguistics*, vol. 2, 139–182, Manila: Linguistic Society of the Philippines.
McKaughan, H.P. (1958) *The Inflection and Syntax of Maranao Verbs*, Manila: Bureau of Printing.
——(1973) 'Subject versus Topic', in A.B. Gonzales (ed.), *Parangal kay Cecilio Lopez*, 206–213, Quezon City: Linguistic Society of the Philippines.
Mead, D. (2002) 'Proto Celebic focus revisited', in Wouk and Ross (eds), 143–177.
Mithun, M. (1991) 'Active/agentive case marking and its motivations', *Language* 67:510–546.
——(1994) 'The Implications of Ergativity for a Philippine Voice System', in B.A. Fox and P.J. Hopper (eds), *Voice: Form and Function*, 247–277, Amsterdam: Benjamins.
Müller-Gotama, F. (2001) *Sundanese*, München: Lincom.
Musgrave, S. (2001) *Non-subject Arguments in Indonesian*, PhD Thesis, The University of Melbourne.
Naylor, P.B. (1995) 'Subject, Topic, and Tagalog syntax', in D. Benett, T. Bynon and G.B. Hewitt (eds), *Subject, Voice and Ergativity*, 161–201, London: SOAS.
Newman, J. (1984) 'Nasal replacement in western Austronesian: an overview', *Philippine Journal of Linguistics* 15:1–17.
Pastika, I Wayan (1999) *Voice selection in Balinese narrative discourse*, PhD Thesis, Australian National University, Canberra.

Pater, J. (2001) 'Austronesian nasal substitution revisited', in L. Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and Representations*, 159–182, Cambridge University Press.
Payne, T.E. (1982) 'Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog', *Studies in Language* 6:75–106.
Plank, F. (1999) 'Split morphology: How agglutination and flexion mix', *Linguistic Typology* 3:279–340.
Rau, D.V. (1997) *A Grammar of Atayal*, Taipei: The Crane Publishing Company.
Reid, L.A. (1970) *Central Bontoc: Sentence, Paragraph and Discourse*, Norman/ University of Oklahoma: Summer Institute of Linguistics.
——(1992) *Guinaang Bontoc Texts*, Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
——(2002) 'Determiners, nouns or what? Problems in the analysis of some commonly occurring forms in Philippine languages', *Oceanic Linguistics* 41:295–309.
Remijsen, B. (2002) 'Lexically contrastive stress accent and lexical tone in Ma'ya', in C. Gussenhoven and N. Warner (eds) *Laboratory Phonology VII*, 585–614, Berlin: Mouton de Gruyter.
Robins, R.H. (1957) 'Vowel Nasality in Sundanese', in *Studies in Linguistic Analysis* (Special volume of the Philological Society), 87–103, Oxford: Blackwell.
Ross, M.D. (2002) 'The history and transitivity of western Austronesian voice and voice-marking', in Wouk and Ross (eds), 17–62.
Samely, U. (1991) *Kedang (Eastern Indonesia). Some Aspects of its grammar*, Hamburg: Buske.
Sasse, H.-J. (1993) 'Das Nomen – eine universale Kategorie?', *Sprachtypologie und Universalienforschung* 46:187–221.
Schachter, P. (1976) 'The Subject in Philippine Languages, Topic, Actor, Actor-Topic or None of the Above', in Li (ed.), 491–518.
——(1995) *The Subject in Tagalog: Still none of the above*, UCLA Occasional Papers in Linguistics vol. 15, Los Angeles: UCLA, Department of Linguistics.
Schachter, P. and Otanes, F.T. (1972) *Tagalog reference grammar*, Berkeley: University of California Press.
Scheerer, O. (1924) 'On the Essential Difference Between the Verbs of the European and the Philippine Languages', *Philippine Journal of Education* 7:1–10.
Sells, P. (1997) 'The functions of voice markers in the Philippine Languages', in S.G. Lapointe, D.K. Brentari and P.M. Farell (eds), *Morphology and its relation to phonology and syntax*, 111–137, Stanford: CSLI.
Senft, G. (ed.) (1997) *Referring to space: studies in Austronesian and Papuan languages*, Oxford: Clarendon Press.
Shibatani, M. (1988) 'Voice in Philippine languages', in Shibatani (ed.), 85–142.
——(ed.) (1988) *Passive and Voice*, Amsterdam: Benjamins.
Silzer, P.J. (1983) *Ambai: an Austronesian language of Irian Jaya, Indonesia*, PhD Thesis, Australian National University.
Sirk, Ü. (1996) 'On the history of transitive verb suffixes in the languages of western Indonesia', in H. Steinhauer (ed.), 191–205.
Sneddon, J.N. (1975) *Tondano Phonology and Grammar*, Canberra: Pacific Linguistics.
——(1993) 'The drift towards final open syllables in Sulawesi languages', *Oceanic Linguistics* 32:1–44.
——(1996) *Indonesian. A comprehensive grammar*, London: Routledge.
Steinhauer, H. (ed.) (1988) *Papers in Western Austronesian Linguistics No. 4*, Canberra: Pacific Linguistics.
——(ed.) (1996) *Papers in Austronesian Linguistics No. 3*, Canberra: Pacific Linguistics.
Stokhof, W.A.L. (1988) 'A modern grammar of Acehnese: Some critical observations', *Bijdragen tot de Taal-, Land- en Volkenkunde* 144:323–350.

Thurgood, E. (1997) 'Bontok reduplication and prosodic templates', *Oceanic Linguistics* 36:135–148.
Thurgood, G. (1999) *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change*, Honolulu: University of Hawai'i Press.
Uhlenbeck, E.M. (1975) 'Sentence segment and word group. Basic concepts of Javanese syntax', *Nusa* 1:1–10.
Van den Berg, R. (1989) *A Grammar of the Muna Language*, Dordrecht-Holland: Foris.
Van der Leeden, A.C. (1997) 'A tonal morpheme in Ma'ya', in C. Odé and W. Stokhof (eds), *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 327–350, Amsterdam: Rodopi.
Van Hoëvell, G.W.W.C. (1877) 'Iets over de vijf voornaamste dialecten der Ambonsche landtaal (bahasa tanah)', *Bijdragen tot de Taal-, Land- en Volkenkunde* 4:1–136.
Van Valin, R.D. and LaPolla, R.J. (1997) *Syntax. Structure, Meaning and Function*, Cambridge: Cambridge University Press.
Verhaar, J.W.M. (1984) 'Affixation in contemporary Indonesian', in B. Kaswanti Purwo (ed.), *Towards a description of contemporary Indonesian: Preliminary Studies*, Part I (=NUSA 18), 1–26, Jakarta: Universitas Atma Jaya.
Wechsler, S. and Arka, I Wayan (1998) 'Syntactic Ergativity in Balinese: an Argument Structure Based Theory', *Natural Language and Linguistic Theory* 16:387–442.
Williams-van Klinken, C., Hajek, J. and Nordlinger, R. (2002) *A short grammar of Tetun Dili*, München: Lincom.
Wolff, J.U. (1966) *Beginning Cebuano*, part I, New Haven: Yale University Press.
——(1972) *A Dictionary of Cebuano Visayan*, 2 vols, Ithaca: Cornell University Southeast Asia Program.
——(1996) 'The development of the passive verb with pronominal prefix in Western Austronesian languages', in B. Nothofer (ed.), *Reconstruction, classification, description — festschrift in honor of Isidore Dyen*, 15–40, Hamburg: Abera-Verlag.
—— (2001), 'Cebuano', in J. Garry and C. Rubino (eds), *Facts about the world's languages*, 121–126, New York: New England Publishing Associates.
Wouk, F. and Ross, M.D. (eds) (2002) *The history and typology of western Austronesian voice systems*, Canberra: Pacific Linguistics.
Zobel, E. (2002) 'The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax', in Wouk and Ross (eds), 405–434.
——(to appear) *A diachronic study of Austronesian verb morphosyntax*, PhD Thesis, Universität Frankfurt.
Zorc, R.D. (1977), *The Bisayan dialects of the Philippines: subgrouping and reconstruction*, Canberra: Pacific Linguistics.