Lexical borrowing in Austronesian and Papuan languages: concepts, methodology and findings

Klamer, M.A.F.; Moro, F.R.

CHAPTER 1

Lexical Borrowing in Austronesian and Papuan Languages: Concepts, Methodology and Findings

Marian Klamer and Francesca R. Moro

Introduction

A fundamental idea in linguistics is that similarities between geographically close languages are not accidental, but point to a shared history of their speakers. Either, the speakers descend from a common ancestor, and the similar features were passed down the generations; or they are, or once were, in mutual contact, and adopted features from each other. This volume studies the latter type of contact-induced similarities, focussing on lexical borrowing.

Lexical borrowing involves the transmission of lexical material from one language to another. Lexicon is easily borrowed, and the lexicon of a language can provide important traces of the social and cultural past of its speakers (Ross 2013). For example, loanwords often signal contact in particular socio-semantic domains such as governance, technology, religion or trade at specific moments in time, and the contact may be datable by the spread of loanwords through a group of languages and level of integration into individual languages. As one of the most widespread and extensively documented form of contact-induced language change (Grant 2015), lexical borrowing is probably the most fruitful part of a language to look at in search of traces of a past history of contact.

Island South East Asia and New Guinea are ideal regions in which to study language contact. The region hosts thousands of languages and has a long history of contact through trade and marriage exchanges, or by culturally dominant groups, both colonial and indigenous. Coupled with the sharp lexical and typological contrasts between the Austronesian and non-Austronesian (Papuan or Indo-European) languages spoken in the region, this provides numerous opportunities to study many different types of language contact situations. The present volume studies language contact particularly in the Philippines, Indonesia, Timor-Leste, and New Guinea.

Although linguistic research on language change induced by contact between Austronesian and Papuan languages is increasing, the number of studies is still rather limited, and their scope varies. Most publications on lexical borrowing describe how a single language is influenced by a (regionally

or nationally) dominant language—recent examples from the region include Saad, Klamer & Moro (2019); Klamer & Saad (2020). Studies incorporating a wider set of Austronesian and Papuan languages typically study the borrowing or 'diffusion' of grammatical features (Ross 1996; Dunn et al. 2008; Foley 2010), sometimes in order to define so-called 'linguistic areas' (Klamer, Reesink & van Staden 2008; Ewing & Klamer 2010; Schapper 2015; Holton & Klamer 2017). The two edited volumes published so far on contact-induced change in the Austronesian world, namely *Language contact and change in the Austronesian world* (Dutton & Tryon 1994) and *Language change in Austronesian languages* (Ross & Arka 2015) focus mainly on Austronesian languages and discuss various types of (contact-induced) change not restricted to the lexical domain. The volume by Andersen (2003), *Language contacts in prehistory: studies in stratigraphy*, includes only one example of an Austronesian language, the language Rotuman (Fiji). Articles specifically centred on borrowing in the lexicon of Austronesian or Papuan languages include Reid (1994) on possible non-Austronesian lexical elements in Philippine Negrito languages, Terrill (2003) on lexical stratigraphy in the central Solomon Islands, Edwards (2018a) on lexical stratigraphy in Timor, Robinson (2015) on Austronesian borrowings in Alor-Pantar languages, and Gasser (2019) on borrowed colour and flora/fauna terminology in North-western New Guinea.

The current volume similarly focusses on borrowing of lexicon, including both Austronesian and Papuan languages, while expanding the geographical focus to include both Island SE Asia and New Guinea. Compared to existing studies it is innovative in three respects. First, most contributions study borrowing of lexicon *across* family borders. For example, Papuan lexicon entering Austronesian languages, Austronesian lexicon entering Papuan languages, lexicon transferring from one Papuan language family into another, or lexicon from an Indo-European language entering an Austronesian language. Second, some chapters (e.g., the chapters by **Edwards** and **Fricke**) systematically examine the entire lexicon of a set of Austronesian languages, focussing on the words that can *not* be shown to have an Austronesian origin. Third, most contributions address the question what loanwords can tell us about the social history of the speaker populations. This question is crucial in Island SE Asia and New Guinea where written historical records and archaeological evidence is very much lacking in most regions. The study of loanwords can provide a window to contact events that happened in the past.

This introductory chapter is organized as follows. In section 1, we give an overview of the concept of loanword, how to define it, the different types of loanwords, and the processes leading to lexical borrowings (1.1). We then discuss methods and practical considerations for detecting loanwords (1.2), and

the data types and data sets that can be used in research on loanwords (1.3). In section 2, we review some of the current models of language contact, relating specific contact settings to amounts and types of lexical borrowings. Section 3 introduces the volume by offering an overview of the chapters.

1 Lexical Borrowing: Concepts, Methods and Data Sets

1.1 *Concepts*

A central concept in this volume is the concept of loanword, which can be defined as 'a word that at some point in the history of a language entered its lexicon as a result of borrowing' (Haspelmath 2009: 36). The process of borrowing comprises all kinds of transfer or copying of linguistic elements from a source language (SL) into a recipient language (RL), including lexemes, derivational morphology, (morpho-)syntactic and lexical-semantic structures. Most contributions in this volume (i.e., **Hoogervorst; Klamer; Edwards; Gerstner-Link; Moro, Sulistyono & Kaiping; Fricke; Schapper & Huber**) are concerned with the borrowing of lexemes, two are concerned with the borrowing of derivational morphology (**Baklanova & Bellamy; Gallego**), and one investigates contact-induced semantic changes in the lexicon (**Saad**).

Traditionally, languages in contact are viewed to directly influence each other in two ways: 'borrowing', affecting the lexicon; and 'interference' affecting the grammar (Weinreich 1953). Van Coetsem (1988; 2000) adds a psycholinguistic dimension to these two processes of transfer, which he refers to as 'borrowing' and 'imposition', introducing the notion of agentivity of the speaker, and the relative dominance of languages in contact in the individual. While the direction of the transfer of linguistic material is always from source language SL to RL, the agent involved in the transfer is either the RL speaker or the SL speaker, depending on which language is their dominant language. A speaker is generally dominant in the language in which she is most proficient or fluent, which is usually, but not necessarily, her first language (van Coetsem 1988: 13). In Van Coetsem's terms, 'borrowing' is then by speakers who show 'RL agentitvity' and adopt elements from one or more SL into their dominant RL, while 'imposition' is the result of speakers who show 'SL agentivity' by transferring features of their dominant SL onto the RL. In this volume, examples of both processes are discussed. 'Borrowing' with RL agentivity would be involved when a speaker of a Timor-Alor-Pantar (TAP) language uses words originating from an Austronesian language (**Klamer**), or when a speaker of an Austronesian language uses words from a TAP language (**Schapper &Huber; Moro et al.**). An example of 'imposition' with SL agentivity would be when a speaker of an

Austronesian language uses derivational morphology from another Austronesian language (**Gallego**) or from a non-Austronesian language (**Baklanova & Bellamy**). In terms of contact-induced outcomes, borrowing typically results in transfer of lexicon to the RL, while imposition typically results in phonological or morpho-syntactic changes in the RL (see section 2).[1]

The word from the SL that served as a model for the loanword in the RL may be called the source word, which may be morphologically simplex or complex. If it is complex, typically the internal structure of the word is lost when it enters the RL. This is in fact one of the ways in which the direction of borrowing can be established: if we attest similar lexemes across two or more languages, and the word is morphologically analyzable in language A, but not in language B, then A is likely to be the SL (see section 1.2 for further discussion of ways to establish loanwords and direction of borrowing). However, while it is rarely attested, complex loanwords can also be borrowed along with their structural properties. Such loanwords give rise to words in the RL that show combinations of non-native affixes with native stems, and native affixes with non-native stems; besides the regular native-native and non-native-non-native combinations. An example of this is Ibatan, which combines non-native prefixes and stems borrowed from Ilokano with native Ibatan affixes and stems (**Gallego**).

1.2 *Methods*

A loanword has a form and a meaning that is identical or similar to the form and meaning of a lexeme in a SL with which plausible contact exists, or existed. For example, contact is plausible when the languages are spoken in adjacent geographical regions, or are known to be (or have been) involved in trade or marriage exchange. If similarities between lexemes are explainable by their common descent, they are not loanwords. Sound imitations and nursery forms are known to be crosslinguistically formed in similar ways without having a shared history, so similarities between such forms cannot be taken to point to contact either.

In some cases, it is not known whether a word is a loanword or a native form in a particular language or language group; then, the form-meaning pair(s) are referred to neutrally as 'lexemes', and the investigation of their history considers 'shared lexicon' (**Schapper & Huber**) or 'lexeme sets', sets of formally similar words that appear across languages (**Fricke**; **Moro et al.**). Lexeme sets can be distinguished into two types: cognate sets and similarity sets. Cognate sets trace

1 Van Coetsem's notion of 'imposition' corresponds closely to 'interference through shift' in Thomason & Kaufman (1988) (see Winford 2020).

back to a reconstructible proto form in a proto language (represented with an asterisk ⟨*⟩ preceding it, e.g. Proto Malayo-Polynesian *pitu 'seven'), while similarity sets are not known to be reconstructible to a common proto form. They do however show striking form-meaning similarities that suggest some shared history: either common descent, or contact, or a combination of both. If the assumption is that they may share a common ancestor, the possible/hypothetical proto form is preceded by a hashtag ⟨#⟩ to distinguish it from established proto forms (e.g., #kafo 'eight', **Schapper &Huber** Table 6.3; Lamaholot-Kedang #dahe-k 'near', **Fricke** Table 5.2).

In most studies in this volume, loanwords are diagnosed using the results of earlier historical comparative work. For example, one way to argue that a lexeme (set) has been borrowed into Timor-Alor-Pantar languages is to demonstrate that it has a Proto Austronesian (PAN) or Proto Malayo-Polynesian (PMP) reconstructed form with a similar form and meaning, from which it can be regularly derived. Similarly, to argue that a lexeme attested in an Austronesian language is from a non-Austronesian (Papuan) SL, it is useful to show a similar form that has been reconstructed for a non-Austronesian group of languages.

For the etymology of Austronesian lexemes, the database of Austronesian and its subgroups as listed in Blust & Trussel (2016) is used. In addition, several chapters in this volume make use of recent reconstructions of lower-level subgroups within Malayo-Polynesian that have been proposed in recent years: the Flores-Lembata subgroup, and within it, the Lamaholot subgroup (Fricke 2019); the Central Flores subgroup (Elias 2018); the Timor-Babar subgroup and the Central Timor subgroup (Proto Timor-Babar being a sister to Proto Central Timor and Helong, Edwards 2018b; 2018a); the Rote-Meto cluster (Edwards 2021) and the Alorese cluster (Sulistyono 2022). For the etymology of lexemes from Timor-Alor-Pantar languages, forms from Proto Alor-Pantar (Holton et al. 2012; Holton & Robinson 2017), or Proto Timor-Alor-Pantar (Schapper, Huber & van Engelenhoven 2017) can be compared. With such detailled etymological information available it is possible to establish which forms in a similarity set share an Austronesian or a TAP ancestor, and which forms do not (**Klamer; Moro et al.; Schapper & Huber**). It also allows us to identify which lexemes are of 'unknown origin' or 'non-Austronesian' (**Fricke; Edwards**); forms that can then be hypothesised to have been acquired through language contact.

When loanwords are attested across two or more languages, the next step is to formulate a hypothesis about the SL, or the direction in which the borrowing took place. The chapters of this volume have applied several practical considerations for this, including the following.

1. If similar forms across language family A are demonstrably historically related (e.g., because they are regularly derived from a known proto form,

show regular sound correspondences), while a similar form is only attested in one language of family B, then the direction of borrowing is from A to B.

II. If similar forms in a language or language family A are more similar to each other and/or show a larger geographical spread than those attested in language (family) B, then the direction of borrowing is from A to B.

III. If a word is morphologically analyzable in language A, but not in language B, then A is the SL.

IV. If a word is integrated into the phonological system of language A but not in that of language B, then A is the SL.

V. If a word is attested in language A, language B, and a sister of B, language C, and language C cannot have been under influence of language A, then B is the SL.

If a word in a particular sub-branch of a language family has no similar forms in the rest of the family, this may be seen as evidence for its status as a loanword. However, this individual word may in fact be an inherited word whose cognates happened to be lost elsewhere in the family, so such instances are not considered to be strong evidence for a contact event (Haspelmath 2009: 44). However, the more words a language has without cognates in the family, the less likely the scenario that all of these words got lost in all the other branches. A large amount of words of unknown ancestry in a particular language or language group is therefore suggestive of a contact event, even if no SL is currently attestable (**Fricke; Edwards**).

1.3 *Data Sets*

As pointed out above, in Island South East Asia and New Guinea, where most indigenous communities do not have written traditions, it is often impossible to exactly date when certain linguistic changes and language contact events took place. This is reflected in Part I of the volume where the dating of pre-modern contacts often remains vague, placing it between the time of the expansion of Malayo-Polynesian languages into Island SE Asia 4000 Before Present time (BP) and the arrival of the first western colonial powers about 500 BP. The data used in the chapters of Part I are generally from previously unwritten sources, including primary data collected through recent fieldwork and oral histories.

Only a few languages in the region have old written traditions. The two main ones are Malay and Javanese, whose written traditions can be traced back to respectively the 7th Century CE (1300 BP), and the 9th Century CE (1500 BP) (**Hoogervorst**). It is the written tradition of Javanese in particular that provides insights into the history of this language and the languages it has been in contact with. At the same time, Malay was the language of the powerful Malay

empire that had its centre in Malacca on the west coast of Malaysia (located between today's Kuala Lumpur and Singapore). By the end of the 15th C, Malacca exerted its influence on its immediate region with its literature in Malay, its style of government and culture, thus accelerating the spread of the Malay language. At the height of Malacca's power, the Malay influence even spread to areas beyond their political control, such as the islands of Ternate and Tidore in the Northern Moluccas. Malay thus became the language of literature and the language of court in many parts of the archipelago, and was thoroughly established by the time the European colonizers arrived in the 16th C. It was subsequently taken up by the Portuguese, Dutch and British colonial powers as a tool of centralisation and modernisation (Collins 1997). Malay as the language of trade has retained its role to this day. Malay was (and is) thus the vehicle by which many loanwords from other language families (Dravidian, Indo-Aryan and Indo-European) entered the local languages of Island SE Asia (**Hoogervorst**).

Sometimes, important regional languages were recorded on paper by the colonial powers. This includes for example Tagalog, the current national language of the Philippines, sources of which go back to the time of the Spanish rule in the late 16th C (**Baklanova & Bellamy**). However, in most of the regions discussed in this volume, linguistic documentation only started about fifty years ago, with the bulk of the work taking place during the last twenty years. So, most chapters use synchronic data sets without information on past stages of the languages.

Apart from the fact that they are mostly synchronic in nature, the data sets as used in the studies of this volume are very different in type and size, an overview is given in Table 1.1. Three contributions (**Klamer; Fricke; Moro et al.**) have made use of the data in the online lexical database LexiRumah (Kaiping, Edwards & Klamer 2019). The reader is referred to Lexirumah for the sources of the data.

TABLE 1.1 Data types and data sets used in the chapters of this volume, organised according to size of data set

Chapter	Recipient language(s)	Source language(s)	Data type	Data set size
7	**Alorese** (Moro et al.)	TAP languages	Mainly synchronic lexical data from LexiRumah	Very large: 13 Alorese dialects, 55 Austronesian language varieties, 42 TAP language varieties × ~600 words = more 66,000 lexemes

TABLE 1.1 Data types and data sets used in the chapters of this volume, organised according to size of data set (*cont.*)

Chapter	Recipient language(s)	Source language(s)	Data type	Data set size
8	**Kilmeri** (**Border**) (Gerstner-Link)	Nimboran / Sentani	Synchronic lexical data from (sketch) grammars, wordlists, dictionaries	Relatively large: 14 Papuan languages (Kilmeri, Waris, Imonda, Amanab, Taikat, Auyi, Nimboran, Sentani, Skou, Wutung, Dumo, Dusur, I'saka, Barupu), from each language ~100 items
3	**TAP languages** (Klamer)	Malayo-Polynesian	Syncronic data from wordlists and reconstructed forms in LexiRumah	Large: 54 TAP language varieties and 55 AN language varieties. For each language, 75 concepts were inspected, i.e. 109 lects × 75 lexemes = 8,175 lexemes
4	**Proto Rote-Meto** (Edwards)	extinct non-AN	Synchronic lexical data; reconstructions based on these forms	Large: 1,173 Proto Rote-Meto reconstructions; the presence of cognates in other languages in the region has also been tracked
5	**Lamaholot** (Fricke)	extinct non-AN	Synchronic lexical data from wordlists in LexiRumah and from dictionaries, reconstructed forms	Large: 46 Flores-Lembata language varieties, from which over 400 lexeme sets were extracted
9	**Tagalog** (Baklanova and Bellamy)	Spanish	(a) Historical data from the 19th–early 20th century lexica (b) Contemporary data of the 20th–early 21st century	Large: Older Spanish-Tagalog dictionaries; 34 sample Tagalog texts, 6 pieces of literary texts; modern Tagalog dictionaries, the Tagalog Leipzig Corpus
11	**Abui** (Saad)	(Alor) Malay	Synchronic data set with utterances	Large: 6 videoclips × 66 speakers = 396 utterances
2	**Malay, Javanese and other AN languages** (Hoogervorst)	Indo-Aryan (e.g., Sanskrit) and Dravidian (e.g., Tamil)	Written sources, dictionaries, old texts	Unspecified
6	**KAWAIMINA languages** (Schapper & Huber)	TAP languages	Synchronic data from (sketch) grammars, dictionaries, fieldnotes; reconstructed forms	Unspecified

TABLE 1.1 Data types and data sets used in the chapters of this volume, organised according to size of data set (*cont.*)

Chapter	Recipient language(s)	Source language(s)	Data type	Data set size
10	**Ibatan** (Gallego)	Ilokano	Synchronic data set including an Ibatan dictionary, and recordings of naturalistic speech during fieldwork in 2018	Unspecified

Intuitively, we might expect that the size of a data set would influence the results: the more lexemes of a language are investigated, the higher the chance of detecting new loanwords. This would particularly be the case when the lexeme sets under investigation are not restricted to basic word lists or non-cultural 'core vocabulary' (which are assumed to be more resistant to borrowing than other vocabulary), but also include highly borrowable cultural concepts, such as is the case in the word lists in LexiRumah.

In this respect, it is interesting to note that **Moro et al.** investigated a huge data set of 66,000 forms from LexiRumah, but found that the percentage of Timor Alor Pantar (TAP) loanwords in Alorese is only slightly higher than the (low) percentages found in earlier studies that were conducted on a basic vocabulary Swadesh list. As Moro et al. remark, this suggests that a loanword analysis on the basis of a Swadesh list can give a representative figure of the proportion of loanwords in a language. On the other hand, however, **Edwards** in his contribution shows that in Austronesian Proto Rote-Meto, the basic vocabulary contains fewer non-Austronesian words (31% of 242 items) than the larger lexicon (55% of 1,148 items) (**Edwards**, Table 4.10). Note however, that one third of the basic vocabulary of Proto Rote-Meto was non-Austronesian, a proportion that goes against the generally accepted (but yet unproven) idea that basic vocabulary is immune to borrowing. In general, languages in our region of study appear to be variable in this regard, and core vocabulary items such as body part terms, kinship terms and certain numerals are often borrowed (**Edwards; Schapper & Huber; Moro et al.; Klamer; Gerstner-Link; Hoogervorst**; see also Foley 2010: 799).

2 Contact Settings and Amount of Lexical Borrowing

Generally speaking, when two or more languages are in contact, this means that groups of speakers interact face-to-face to a certain extent. This interaction, as we will see below, can bring about all kind of changes in the structure and the lexicon of the languages involved, usually the more intense the interaction, the more pervasive the changes will be. Linking contact-induced language changes to specific contact settings allows us to make predictions about what will happen in a given scenario, or hypotheses about what has happened in the past. Here is one example (adapted from Aalberse, Backus & Muysken 2019: 13):

> Assume that if a prototypical social setting involving language contact *A* (e.g., contact between North Moluccan Malay and Taba, an indigenous language of Indonesia) has been well studied and produces linguistic properties *p* and *q* (i.e., borrowing of grammatical function words from Malay), then a social setting under study *B* (i.e., contact between the local Malay variety and another indigenous language of Indonesia), resembling *A* in crucial ways, will be likely to also have these properties *p* and *q* (i.e., borrowing of approximately the same grammatical function words from Malay), assuming also roughly the same types of languages involved.

So, we can expect that in other indigenous communities of Indonesia dominated by Malay, the local languages will be influenced approximately in the same way as Taba is. This is exactly what we find, as reported for other Austronesian languages, like West Tarangan, Biak, and Central Lembata, and non-Austronesian Abui (e.g., Nivens 1998; van den Heuvel 2006; Fricke & Saad 2017), all of which have incorporated Malay function words like *kalau* 'if'.

In order to make predictions, like the one above, we need models of language contact, which explain the processes, as well as the psycholinguistic and sociolinguistic mechanisms that underpin outcomes of language contact, and can be used to infer the contact setting that brought about a specific change (Thomason 2001; Kusters 2003; Trudgill 2011; Muysken 2013; Ross 2013).

For example, Thomason (2001: 70–71) proposes the following borrowing scale to predict which types of lexical borrowings can be expected in contact situations.

Intensity of contact correlates with the amount and types of lexical borrowings: under conditions of casual contact only non-basic vocabulary gets

TABLE 1.2 Lexical borrowing in Thomason's borrowing scale

Intensity of contact	Type of speakers	Borrowed elements
1. **Casual**	**Few bilinguals** among borrowing-language speakers, borrowers need not be fluent in the source language.	Only **non-basic vocabulary**. Only **content words**: most often nouns, verbs, adjectives, and adverbs.
2. **Slightly more intense**	**More fluent bilinguals** among borrowing-language speakers, but they are probably still a minority.	Still **non-basic vocabulary. Function words** (e.g. conjunctions and adverbial particles like 'then') as well as content words.
3. **More intense**	**A conspicuous number of bilinguals** among borrowing-language speakers, attitudes and other social factors favoring borrowing.	**Basic** and **non-basic vocabulary**. More **function words**, including **closed-class items** as pronouns and low numerals; derivational affixes.
4. **Intense**	**Very extensive bilingualism** among borrowing-language speakers, social factors strongly favoring borrowing.	**Heavy lexical borrowing in all sections of the lexicon**.

BASED ON THOMASON 2001: 70–71

borrowed, but as the intensity of contact increases along with the number of fluent bilinguals in the community, then function words, basic vocabulary, and ultimately derivational morphology and all sections of the lexicon can be borrowed as well. Thomason (2001), thus, uses intensity of contact as the main social predictor. The concept of intensity of contact is hard to define, but can be operationalized as a function of the level of fluency of the borrowers, the proportion of borrowing-language speakers who are fully bilingual in the source language, and the speakers' attitudes. Besides intensity of contact, the other major predictor is linguistic: typological similarity between languages enhances the possibility of borrowing, and loose structures are easy to borrow than tightly integrated structures.

Ross (2013) adds a new dimension to the concept of intensity of contact, namely that of age. In his study on shift-induced changes in Melanesia, Ross links life stages of shifting speakers to prototypical linguistic effects: adult second language learning typically leads to the retention of a good amount

of vocabulary from their heritage language into (the version of) the language to which they are shifting (together with phonological transfer, constructional calquing and simplified (morpho-)syntax); while child bilingualism typically leads to lexical calques (together with syntactic copying and complexification).

Taking a cross-linguistic perspective, Tadmor (2009) compares rates of lexical borrowings in the world languages, surveying 41 languages. Tadmor's four levels can be paired with the four types of intensity of contact of Thomason: "low borrowers" (< 10%, casual), "average borrowers" (10–24%, slightly more intense), "high borrowers" (25–50%, more intense), and "very high borrowers" (> 50%, intense). The percentage of lexical borrowing is inevitably linked to specific contact settings, as exemplified by two prototypical cases: Selice Romani (62.7%) and Mandarin Chinese (1.2%). Some of the sociolinguistic circumstances underlying such different borrowing rates are universal multilingualism, minority language status, permissiveness toward borrowings, and donor languages well known in the case of Selice Romani, while we find almost no bilingualism, majority language status, purist attitude and donor languages poorly known in the case of Mandarin Chinese.

We have seen that specific contact settings can predict the amount of lexical borrowing to be found in a given language. However, it is not only the amount of lexical borrowing that varies depending on the sociolinguistic circumstances, but also the meaning of the loanwords, or their semantic fields. Tadmor, Haspelmath & Taylor (2010) investigated the likelihood of borrowing across a list of 22 semantic fields (taken from Buck 1949) in 41 languages. The six fields most likely to be borrowed (> 30%) are: *Religion and belief*, *Clothing and grooming*, *The house*, *Law, Social and political relations*, and *Agriculture and vegetation*. Thus, we can expect that in contact situations that involve casual contact, where few speakers are fluent bilinguals in both languages, the loanwords will come from these semantic fields. One example of casual contact is that of Sanskrit loanwords in Malay and Javanese (and in other languages of the region), as discussed in **Hoogervorst**, that indicate new items or concepts, such as *āgama* 'sacred traditional doctrine or precepts' (*Religion and belief*), or *doṣa* 'transgression', *pañjara* 'prison', *sākṣī* 'witness' (*Law*).

As hinted above, language contact models can be used in two ways (Aalberse et al. 2019: 13):

i. They could predict, given a specific language contact setting and a specific language pair, what the linguistic outcome is most likely to be.
ii. They could help understand, given a specific linguistic outcome, what would be the most likely contact setting leading to that outcome has been.

In Island SE Asia and New Guinea, a region that lacks archaeological data and historical written sources, the study of language contact mostly serves

purpose (ii). In facts, virtually all contributions in this volume try to understand, on the basis of the amount and type of lexical borrowings, what was the most likely contact scenario that gave rise to that type of lexical influence. The languages discussed in this volume can be divided according to the intensity of contact, the level of borrowing, the contact processes and the borrowed elements (see Table 1.3 on the next page).

In this region, we find possibly all types of contact setting and related outcomes, from casual contact to intense contact. Four studies report low levels of borrowings in the recipient languages: Kilmeri (**Gerstner-Link**), Alorese (**Moro et al.**), TAP languages (**Klamer**), and Kawaimina languages (**Schapper & Huber**). The limited lexical influence can be accounted for by lack of long-term contact, and pressure to maintain identity (**Gerstner-Link**), by asymmetric bilingualism patterns and numerous first languages (L1s) interfering with each other (**Moro et al.**), by superficial contacts between speakers (**Klamer**), and by lack of data from the non-AN donor languages of Timor, especially in crucial domains such as plants and animals (**Schapper & Huber**). The study of **Hoogervorst** on lexical influence from South Asia languages (e.g., Sanskrit and Tamil) on Malay, Javanese and other languages of the region does not discuss percentages for the individual languages, nor does it specify the type of speakers who were involved. The transmission of South Asian loanwords was primarily the result of language contact with Malay, both for Austronesian and non-Austronesian languages, and therefore we can hypothesize that the type of contact was casual and involved only few bilinguals among borrowing-language speakers.

Two studies report high level of borrowing in Tagalog (**Baklanova & Bellamy**), and Ibatan (**Gallego**). In Tagalog and in Ibatan, two cases of relatively intense contact, we find borrowing of derivational morphology, as expected according to Thomason's scale (see Table 1.2 above); the contact process is imposition transfer by Ilokano-dominant bilinguals for Ibatan, and by Chinese mestizos for Tagalog. We find only two cases of very high levels of borrowings: **Edwards** who discusses loanwords from an extinct non-AN language into Proto Rote-Meto, and **Fricke** who discusses loanwords from an extinct non-AN language into Lamaholot. Both studies discuss lexical borrowing from a language(s) for which we no longer have direct evidence (also known as 'reconstructio ex silentio', see Ross 2013: 11). The difference is that in the case of Proto Rote-Meto, the contact process was adult language shift, as evidenced by the fact that loanwords come from specific semantic domains, and that we also find traces of phonological transfer (see Ross 2013), while in the case of Lamaholot code-switching was the more likely process, as all domains of the lexicon are involved.

TABLE 1.3 Contact settings and lexical borrowing in the contributions of this volume

Recipient language(s)	Source language(s)	Intensity of contact	Level of borrowings	Contact process	Borrowed elements
Malay and **Javanese** (Hoogervorst)	South Asian	(not discussed)	(not discussed)	Not specified in the paper. Malay and Javanese were the carriers of loanwords into other local RLs.	Semantic domains of loanwords: precious minerals, and metals, geography, law, plants, numerals, religion, mythology, governance, toponyms, and royal titles.
Kilmeri (**Border**) (Gerstner-Link)	Nimboran / Sentani	Casual	Low (2,3%)	Bilingualism in the family and village contexts due to intermarriage. Language is seen as an emblem of group identity (e.g., for Kilmeri).	Loanwords in the semantic domains of nature, animals, kinship, body parts, and motion. *Wanderwörter* regarding 'water', 'vegetation' and 'arrow' suggestive of trade (bird of paradise).
Alorese (Moro et al.)	TAP languages	Casual	Low (4.7%)	Asymmetric bilingualism, several L1s interfering with each other.	Loanwords especially in the semantic domains of tools, vegetation, and basic actions.
TAP languages (Klamer)	Malayo-Polynesian	Casual	Low (~8%)	No pervasive bilingualism, nor shift; more likely superficial contact.	Loanwords especially in the semantic domains of technology, societal structures, and subsistence and trade.
Kawaimina languages (Schapper & Huber)	TAP languages	Casual	Low (11 items, percentage not given)	(not discussed)	Loanwords especially in the semantic domains of plants and animals, in particular creepy-crawlies.
Tagalog (Baklanova)	Spanish	More intense	High (20–32%)	(not discussed)	Derivational morphology.
Ibatan (Gallego)	Ilokano	More intense	High (40%)	Imposition transfer by Ilokano-dominant bilinguals.	Derivational morphology.
Lamaholot (Fricke)	extinct non-AN	Intense	Very high (50%)	Code-switching.	Basic and non-basic vocabulary, no specific semantic domain(s).
Proto Rote-Meto (Edwards)	extinct non-AN	Intense	Very high (55%)	Adult-language shift.	Basic and non-basic vocabulary, especially in the semantic domains of tools, and vegetation.
Abui (Saad)	(Alor) Malay	Intense	(not discussed)	Transitional bilingualism: (pre)adolescents and young adults dominant in Malay.	Semantic changes in the lexicon: generalization in three verbal domains.

As for the language Abui, **Saad** does not discuss lexical borrowing, but rather the lexical semantic change of 'generalization', whereby some specific words fall into disuse and become replaced by more frequent words. This change is more dramatic in those bilingual speakers who are psycholinguistically dominant in Malay ((pre)adolescents and young adults), thus showing that generalization correlates with intense contact.

Interestingly, in two cases of intense contact out of five, namely Tagalog and Abui, the recipient or donor language is a 'High' variety: a colonial language (Spanish) or a lingua franca or a national language (Malay/Indonesian for Abui). Thus, it seems that when only indigenous local languages are involved in the contact, high or very high levels of borrowing are unlikely. This is possibly connected to the observation that adult language shift (leading to high level of borrowing) is rare in small-scale societies (Ross 2013: 28), such as the ones discussed in this volume.

Finally, an interesting pattern emerges looking at the semantic fields of the loanwords. In the cases of casual contact of Alorese (**Moro et al.**), and Kawaimina languages (**Schapper & Huber**), but also in the case of Proto Rote-Meto (**Edwards**) characterized by intense contact, the semantic fields of *Tools/Technology*, *Agriculture and Vegetation*, *Animals* and *Social and political relations* (including societal structures) are favored. Interestingly, these three case studies discuss possible non-Austronesian lexical influence on Austronesian languages, thus they indicate that non-AN languages of the region mostly contributed with words related to the environment and technology. The case study of **Klamer** on Austronesian influence on TAP languages presents a complementary view, showing that the Austronesian languages contributed with words related to textile technology, societal structures ('slave', 'king/ruler'), subsistence and trade ('salt', 'seed', 'maize', 'skin'), and marriage ('bride price').

3 Introducing the Volume

The volume consists of two parts covering different periods of time. Part I contains five studies of contact that took place in ancient and pre-modern times, and whose contact settings do not exist anymore, or their dynamics have changed dramatically. This is the time between the expansion of Malayo-Polynesian languages into Island SE Asia, which started some 4000 years BP, and the advent of the first western colonial powers about 500 years BP. The contact events in this period cannot be dated with any precision, but must have taken place before the time when western colonial powers produced their written historical records of parts of the region.

The first chapter in Part I is by **Hoogervorst**, who takes the whole of Island SE Asia as region of investigation. His contribution shows traces of ancient East Asian loanwords in the Austronesian and Papuan languages of Island SE Asia, whose dispersal was either direct, or mediated through Malay and Javanese, with Sanskrit mostly a source for cultural borrowings (prestigious concepts), and Tamil for replacive borrowings (every-day items).

The contribution of **Klamer** analyzes Austronesian loanwords attested in TAP languages and shows that the Austronesian influence in pre-modern times involved animals ('pig', 'deer'), textile technology ('needle', 'to weave', 'to sew'); societal structures ('slave', 'king/ruler'), body parts ('breast', 'navel'), subsistence and trade ('salt', 'seed', 'maize', 'skin'), and marriage ('bride price'). She also argues that, while TAP communities have been in contact with Malayo-Polynesian speaking groups since the stage of proto TAP, thousands of years ago, their mutual contacts generally must have remained superficial, being limited to circumscribed domains and individual people.

The chapters by **Edwards** and **Fricke** present a stratigraphic analysis of the lexicon of Rote-Meto and of Lamaholot, respectively. These two languages have undergone a process of relexification, whereby a good amount of pre-existing words have been replaced with words from an (unattested) language. In such cases, lexical borrowings are the only evidence of the existence of an unattested language or scenario of contact (Grant 2015: 13).

Schapper & Huber investigate the lexical entwinement of the (Austronesian) Kawaimina languages and the (TAP) Maka languages in East Timor, and argue for bidirectionality in lexical borrowing between Papuan-Austronesian languages in the Timor area. They show that Papuan etyma found in the Kawaimina languages have not necessarily been borrowed from the Maka languages. At the same time, Makasae, the largest Maka language, is the immediate source for Austronesian etmya in the Kawaimina languages; and some lexicon that is shared between Kawaimina and Maka languages has no clear origin outside of those groups or appears to have been borrowed in parallel into both group's languages from one or more unknown languages.

Part II of this volume covers studies of contact in modern and contemporary times (from 500 BP to the present), in contact settings that are to some extent still present today.

The contribution of **Moro, Sulistyono & Kaiping** on Alorese, an Austronesian language surrounded by Papuan TAP languages, display a clear example of a language in which, despite a long history of contact, lexical borrowing is not very significant in quantitative terms, but it can be revealing to understand pattern of interactions and dialect dispersal.

Gerstner-Link investigates lexical borrowing in a complex exchange scenario involving the Papuan families of Border, Nimboran, Sentani, and Skou

FIGURE 1.1 Locations of languages or language areas discussed in the chapters of this volume, by their chapter number

Legend to map

2. *Hoogervorst: Lexical influence from South Asia* (*map indicates locations of Malay and Old Javanese*)
3. *Klamer: Traces of pre-modern contact between Timor-Alor-Pantar and Austronesian speakers*
4. *Edwards: Phonological innovation and lexical retention in the history of Rote-Meto*
5. *Fricke: The mixed lexicon of Lamaholot* (*Austronesian*)*: A language with a large lexical component of unknown origin*
6. *Schapper & Huber: Entwined histories: the lexicons of Kawaimina and Maka languages*
7. *Moro, Sulistyono & Kaiping: Detecting Papuan loanwords in Alorese: Combining quantitative and qualitative methods*
8. *Gerstner-Link: Multilateral lexical transfer among four Papuan language families: Border, Nimboran, Sentani, and Sko*
9. *Baklanova & Bellamy: Spanish suffixes in Tagalog nominal derivation: The case of common nouns*
10. *Gallego: The structural consequences of lexical transfer in Ibatan*
11. *Saad: The effects of language contact on lexical semantics: The case of Abui*

located in the island of New Guinea. On the basis of the high number of mutual loans between Border and Nimboran languages, new hypotheses are formulated about the migration routes of the Border people, as well as about the genetic unity of the Border and Nimboran families.

The paper by **Baklanova & Bellamy**, as well as the one by Gallego, both show that loanwords can lead to the transmission and integration of derivational morphemes in the recipient languages. For instance, as shown by Baklanova and Bellamy, Tagalog has absorbed many Spanish words which acted as a conduit for the borrowing of agentive and adjectival suffixes. Similarly, **Gallego** analyses the history and development of the verbal prefix *mag-* in Ibatan, which has been copied from Ilokano as part of complex loanwords.

Saad's is the only contribution that focuses on the outcome of contact-induced change in the semantics of language, by demonstrating that the meaning of certain verbs in Abui (a TAP language) has changed due to the influence of semantically similar verbs in the dominant language Malay (Austronesian).

The linguistic region covered by each of the chapters is indicated on the map in Figure 1.1 on page 17. More detailed maps of these respective areas are provided in the individual chapters.

Bibliography

Aalberse, Suzanne, Ad Backus & Pieter Muysken. 2019. *Heritage Languages: A language contact approach*. Amsterdam: John Benjamins.

Andersen, Henning (ed.). 2003. *Language Contacts in Prehistory: Studies in Stratigraphy*. Amsterdam: Benjamins.

Blust, Robert & Stephen Trussel (eds.). 2016. The Austronesian Comparative Dictionary (web edition). Retrieved from http://www.trussel2.com/ACD

Buck, Carl Darling. 1949. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages*. Chicago: University of Chicago Press.

Collins, James T. 1997. *Malay, world language of the ages. A sketch of its history*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink & Angela Terrill. 2008. Structural Phylogeny in Historical Linguistics: Methodological Explorations Applied in Island Melanesia. *Language* 84(4). 710–759.

Dutton, Tom & Darrell T. Tryon (eds.). 1994. *Language contact and change in the Austronesian world*. Berlin / New York: Mouton de Gruyter.

Edwards, Owen. 2018a. Parallel Histories in Rote-Meto. *Oceanic Linguistics* 57(2). 359–409.

Edwards, Owen. 2018b. Top-down historical phonology of Rote-Meto. *Journal of the*

Southeast Asian Linguistics Society (*JSEALS*) 11(1). 63–90. DOI: http://hdl.handle.net/10524/52421

Edwards, Owen. 2021. *Rote-Meto comparative dictionary*. Canberra: ANU Press. Retrieved from http://doi.org/10.22459/RMCD.2021

Elias, Alexander. 2018. Lio and the Central Flores languages. Research Master thesis Leiden University. https://hdl.handle.net/1887/69452

Ewing, Michael & Marian Klamer (eds.). 2010. *East Nusantara: Typological and areal analyses*. Canberra: Pacific Linguistics.

Foley, William. 2010. Language Contact in the New Guinea Region. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 795–813. West Sussex, UK: Wiley-Blackwell.

Fricke, Hanna. 2019. *Traces of language contact: the Flores-Lembata languages in eastern Indonesia*. Amsterdam: LOT Publications.

Fricke, Hanna & George Saad. 2017. 'If' it exists in one language, why not use it in another? The insertion of kalau 'if' in Abui and Central Lembata, two languages of Eastern Indonesia. Presented at the Ninth Austronesian and Papuan Languages and Linguistics conference (APLL 9), Paris.

Gasser, Emily. 2019. Borrowed Color and Flora/Fauna Terminology in Northwest New Guinea. *Journal of Language Contact* 12. 609–659.

Grant, Anthony. 2015. Lexical borrowing. In *The Oxford Handbook of the Word*, 1–23. Oxford: Oxford University Press. Retrieved from DOI 10.1093/oxfordhb/9780199641604.013.029

Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In Uri Tadmor & Martin Haspelmath (eds.), *Loanwords in the World's Languages: A Comparative Handbook*, 35–54. Walter de Gruyter.

Holton, Gary & Marian Klamer. 2017. The Papuan languages of East Nusantara and the Bird's Head. In Bill Palmer (ed.), *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide*, 569–640. Berlin: Mouton de Gruyter.

Holton, Gary, Marian Klamer, František Kratochvíl, Laura C. Robinson & Antoinette Schapper. 2012. The historical relations of the Papuan languages of Alor and Pantar. *Oceanic Linguistics* 51(1). 86–122.

Holton, Gary & Laura C. Robinson. 2017. The linguistic position of the Timor-Alor-Pantar languages. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology* (*2nd ed.*), Second edition. 155–198. Berlin: Language Science Press.

Kaiping, Gereon, Owen Edwards & Marian Klamer. 2019. LexiRumah 3.0.0. Leiden University Centre for Linguistics. Retrieved from https://lexirumah.model-ling.eu/lexirumah/

Klamer, Marian, Ger Reesink & Miriam van Staden. 2008. East Nusantara as a linguistic area. In Pieter Muysken (ed.), *From linguistic areas to areal linguistics*, 95–149. Amsterdam: Benjamins.

Klamer, Marian & George Saad. 2020. Reduplication in Abui: A case of pattern extension. *Morphology* 30(4). 311–346. DOI: https://doi.org/10.1007/s11525-020-09369-z.

Kusters, Wouter. 2003. *Linguistic complexity*. Utrecht: LOT Publications.

Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(04). 709–730. DOI: https://doi.org/10.1017/S1366728912000727

Nivens, Richard. 1998. *Borrowing vs. code-switching: Malay insertions in the conversations of West Tarangan speakers of the Aru islands of Maluku, eastern Indonesia*. PhD thesis University of Hawaii.

Reid, Lawrence A. 1994. Possible Non-Austronesian Lexical Elements in Philippine Negrito Languages. *Oceanic Linguistics* 33(1). 37–72.

Robinson, Laura C. 2015. The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara. In Malcolm Ross & I Wayan Arka (eds.), *Language change in Austronesian languages*, 19–33. Canberra: Asia-Pacific Linguistics.

Ross, Malcolm. 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In Mark Durie & Malcolm Ross (eds.), *The Comparative Method Revisited: Irregularity and Regularity in Language Change*, 180–217. Oxford: Oxford University Press.

Ross, Malcolm. 2013. Diagnosing Contact Processes from their Outcomes: The Importance of Life Stages. *Journal of Language Contact* 6(1). 5–47. DOI: https://doi.org/10.1163/19552629-006001002

Ross, Malcolm & I Wayan Arka (eds.). 2015. *Language change in Austronesian languages: papers from 12-ICAL, Volume 3*. Canberra: Asia-Pacific Linguistics.

Saad, George, Marian Klamer & Francesca R. Moro. 2019. Identifying agents of change: Simplification of possessive marking in Abui-Malay bilinguals. 4(1), 57. *Glossa: A Journal of General Linguistics* 4(1). 57.

Schapper, Antoinette. 2015. Wallacea, a linguistic area. *Archipel. Études Interdisciplinaires Sur Le Monde Insulindien* (90). 99–151.

Schapper, Antoinette, Juliette Huber & Aone van Engelenhoven. 2017. The relatedness of Timor-Kisar and Alor-Pantar languages: A preliminary demonstration. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology* (*2nd ed.*), 99–154. Berlin: Language Science Press.

Sulistyono, Yunus. 2022. *A history of Alorese*. PhD thesis Leiden University. Utrecht: LOT Publications.

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*., 55–75. Berlin / New York: De Gruyter Mouton.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246. DOI: https://doi.org/10.1075/dia.27.2.04tad

Terrill, Angela. 2003. Linguistic stratigraphy in the central Solomon Islands: Lexical evidence of early Papuan / Austronesian interaction. *The Journal of the Polynesian Society* 112(4). 369–401.

Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah G. & Terence S. Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

van Coetsem, Frans. 1988. *Loan Phonology and the Two Transfer Types in Language Contact*. Dordrecht: Foris.

van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Monographien zur Sprachwissenschaft Vol. 19. Heidelberg: Winter.

van den Heuvel, Wilco. 2006. *Biak: description of an Austronesian language of Papua*. PhD dissertation Vrije Universiteit Amsterdam. Utrecht: LOT Publications.

Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York. Reprinted 1986, The Hague: Mouton.

Winford, Donald. 2020. Theories of Language Contact. In Anthony Grant (ed.), *The Oxford Handbook of Language Contact*, 51–75. Oxford: Oxford University Press.