Siraya

Trends in Linguistics
Documentation 30

De Gruyter Mouton

Siraya

Retrieving the Phonology, Grammar and Lexicon of a Dormant Formosan Language

by

Alexander Adelaar

De Gruyter Mouton

ISBN 978-3-11-025295-8
e-ISBN 978-3-11-025296-5
ISSN 0179-8251

Library of Congress Cataloging-in-Publication Data

Adelaar, K. Alexander.
Siraya : retrieving the phonology, grammar and lexicon of a dormant Formosan language / by Alexander Adelaar.
p. cm. − (Trends in linguistics. documentation; 28)
Includes bibliographical references and index.
ISBN 978-3-11-025295-8 (alk. paper)
1. Siraya language. 2. Bible. N.T. Matthew. Siraya.
3. Anthropological linguistics. I. Title.
PL6162.A34 2011
499'.2−dc23
2011027909

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Typesetting: Frank Benno Junghanns
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
∞ Printed on acid-free paper

Printed in Germany

www.degruyter.com

To Ren and Niuniu

Contents

Figures and Tables

Acknowledgements

This publication was made possible with support from the Chiang Ching-kuo Foundation for International Scholarly Exchange (project SS-002-P-02, Taiwan, 2003), the University of Leuven (Fellowship SF/08/009) in 2009, and the VLAC (Flemish Academic Centre for Science and the Arts, Brussels, 2010). I am grateful to the Institute of Linguistics at the Academia Sinica (Taipei) for inviting me for a two-month stay in 2002–2003. I am also grateful to the Department of Linguistics at Leuven University for hosting me in the second half of 2009 and the VLAC in Brussels for its hospitality in the first half of 2010 when I was involved in the final writing up of this book. In January 2003 I received a generous welcome and a research award from the Tainan Ping-pu Siraya Culture Association in Tainan on behalf of my Siraya research project.

While writing this book I received invaluable feedback from Laurent Sagart (École des Hautes Études et Sciences Sociales, Paris), Laurie Reid (Osaka), Elizabeth Zeitoun (Academia Sinica, Taipei) and Marian Klamer (Leiden University). They are in no way responsible for any errors in the present text, which are no doubt many. I am grateful to Frank Benno Junghanns (Berlin) for his very professional editorial work on this book. I also wish to thank Marian Lang for her excellent proofreading of the Introduction, and Laurent Sagart, Robert Blust (Hawai'i), Paul Jen-kuei Li (Academia Sinica), Lewis Mayo (Melbourne), Jean-Christophe Verstraete (Leuven University) and Petra Couvée (Leiden) for their encouragements and their belief in my Siraya project at times it did matter.

Asia Institute, University of Melbourne, October 2010.

Conventions

The following conventions are used in this book:

– Quotations from the original orthography are in semibold. Square brackets, hyphens and diacritics are maintained as found in the original source. The only alteration to this orthography is the addition of an equal sign (=) to indicate a line break in the original text.

– Phonemic and near-phonemic representations of Siraya text, words, parts of words and phonemes, are in italics.

  N.b. A rationalised near-phonemic representation is used to bring out better the phonemics of the language and to avoid unnecessary spelling variation found in the original orthography. This spelling is not totally phonemic: it maintains the letter combination *ng* to indicate a velar nasal, and it still allows spelling variation where the existence of a given underlying phoneme is probable but not attested. In such cases the relevant distinctions made in the original spelling are generally maintained, but a more rigorous phonemic spelling in italics and between slashes may be used in the Grammatical sketch or Lexicon to bring out the diagnostic structure of the form in question. For instance, *ĭmid* 'all, everything' is usually spelled as such in the gospel text but it also has variant forms *ĭm'd-* and (in the Catechism) *ĭmŭd* and even *ĭmŭt*. This variation is maintained, although the Lexicon also gives /*ĭməd*/ to show that the form underlying this spelling has a final voiced stop (*d*) and most probably also a schwa in the last syllable.

– Phonetic representations (in regular lettertype) are between square brackets.

  Square brackets in combination with italic lettertype are also used to indicate extrapolated and corrected segments in phonemic words.

– A hyphen (-) within a word indicates a morpheme boundary; when attached to the beginning and/or end of a base, it indicates that the base in question does not occur underived (e.g. the base *-tamŭd* only occurs in derivations such as *t‹m›amŭd* (xiv:33) 'to worship' and *ta-tam'd-ən* (vii:16) 'religious service').

– An equal sign (=) in the original orthography indicates a line break in the original text (see above); in the phonemic representation, it indicates the boundary between a clitic and the word it is cliticised to.

– In glossed text, if the meaning of a Siraya word or segment is rendered in English by several words, these English words are connected with a dot

directly between them, e.g. 'come.from' is the semantic gloss for Siraya *măka-*. If the meaning of a Siraya word or segment corresponds to several related meanings in English, these meanings are given with a comma directly between them, e.g. 'when,if' is the semantic gloss for Siraya *ru*.

– In near-phonemic writing, a combination of italics and semibold is used for words or phrases that remain unanalyzed and are glossed with their composite meaning only (this is the case with words or phrases of an obscure structure and some frequently occurring biblical terms).
– In glosses and explanations, a question mark indicates that the meaning or explanation is uncertain.
– English glosses are given between quotation marks; 'id.' means 'same meaning as preceding gloss'. Chevrons (small angled brackets) are used to indicate infixes (e.g. *d‹m›arang* 'to go away'). English glosses between inequality signs ("large" angled brackets) concern Siraya roots that do not occur in isolation (e.g., *-v'li* <to reciprocate>).
– Quotations are followed by a code between brackets referring to their place in the sources. The number of places given for a quotation occurring more than once is not necessarily exhaustive. Roman numbers refer to chapters in the gospel text, and subsequent Arabic numbers to verses in these chapters. A letter 'C' followed by an Arabic number refers to original folio page numbers in the Catechism; an additional 'v' refers to the back ('verso') of a folio page. The letters 'UM' refer to the Utrecht manuscript.[1]
– Where quotations in the original spelling are listed in the Grammar Sketch, I give the equivalent form in my own spelling to the right on the line where these quotations occur. The frequency of a particular spelling variant is sometimes given between brackets after a source reference, e.g. **'jngau** (xix:12) (1x), **yngau** (5x)" indicates that the (phonemic) form *ĭngaw* is spelled as **'jngau** in one case, and as **yngau** in five other cases, in the gospel text.
– 'PAn' refers to Proto Austronesian, PF to Proto Formosan, and PSF to Proto South Formosan.[2] Etyma preceded by an asterisk are assigned to PAn unless indicated otherwise. PAn etyma are taken from Blust (online) unless another source is mentioned.

---

[1] No page numbers are given for this source as Murakami's (1933) edition has Siraya glosses in alphabetical order, and the original Dutch version by van der Vlis (1842) is ordered according to subject.

[2] PF and PSF are labels used in Tsuchida (1976). I maintain these labels, although many etyma labeled with them would also qualify as PAn etyma.

List of abbreviations

Abbreviations used in glossing

1S/2S/3S	first/second/third person singular
1PE	first person plural exclusive
1PI	first person plural inclusive
2P/3P	second/third person plural
ADD	additive (‘and’, ‘also’)
ADV	adversative (‘but’, ‘however’)
AO	actor-oriented
AO1, AO3, AO4	actor-oriented affixes respectively belonging to class 1,3,4 verbs
AS	anticipating sequence
CAUS	causative
COLL	collective
COM	comitative
DF	default case marker
DST	distal
GEN	genitive
IMP	imperative
INCH	inchoative
INCL	inclusive
INDEP	independent
INV	inversive
LK	linker
LOC	locative
MOT	motion
NEG	negator
NOM	nominative
OBL	oblique
PA	personal article
POSS	possessive
PRF	perfective
PRX	proximal
PST	past

QU	question word
RDP	reduplication
ST	stative
SJ	subjunctive
SJ.UO	portemanteau suffix expressing subjunctive and undergoer voice
UO	undergoer-oriented affix
V1, V4	prefixes forming respectively class 1 and class 4 verbs

Other abbreviations and symbols

a.s.	root used as anticipating sequence
aux.	root used as auxiliary
bibl.	Biblical term
bvb	bound verb
bvb1, bvb3, bvb4	bound verbs respectively belonging to class 1, 3 and 4
C	Catechism
lit.	literally
n.	noun
n.deriv.	nominal derivation
or.	root combined with an orientation prefix
PAn	Proto Austronesian
PF	Proto Formosan
PSF	Proto South Formosan
v	verso
vb	verb
vb1, vb2, vb3, vb4	verbs respectively belonging to class 1, 2, 3 and 4
UM	Utrecht Manuscript
→	in the Grammatical Sketch: 'becomes', 'results into' in the Lexicon: 'see further'
<	'derives from'
>	'became', 'developed into'
<…>	contains English gloss of a root that does not occur in isolation
‹…›	contains an infix
*	indicates that the following segment or word belongs to an older stage and is unattested.

Part A

Introduction[1]

1. General

This study is a linguistic analysis of a 17th century translation of the Gospel of St. Matthew into Siraya (henceforth "Gospel"), a language that is now extinct but used to be spoken in a variety of dialects on the plains of Southwest Taiwan. The Siraya Gospel was translated from the Dutch in 1661 by Daniel Gravius.

For the present analysis, I have used additional data from two other Siraya sources. One is a Dutch-Siraya bilingual shortened version of the (Heidelberg) catechism (henceforth "Catechism"), which was also published by Gravius (1662) and is in the same dialect as the Gospel. The other is the so-called "Utrecht Manuscript,"[2] an anonymous 35-page Dutch-Siraya wordlist representing a different dialect (van der Vlis 1842).

The study has three main parts.

The first part (B) is a grammatical analysis of the Gospel text based on all 28 chapters of the Gospel.

The second part (C) is a glossed text consisting of ten chapters (chapters 2 to 11)[3] of St. Matthew. In it, lines in the original Siraya orthography are followed by corresponding lines representing the original Dutch text, a phonemicised spelling, interlinear lexical glossing and an English translation of the line in question.

The third part (D) is a lexicon including all roots and relevant derivations occurring in the Siraya Gospel, augmented with as many roots and derivations as I was able to find in the Catechism that are not already in the Gospel.

---

1 This chapter includes adapted versions of sections that have already appeared in Adelaar (1997) and (2007).

2 Referring to the rediscovery of this wordlist as a manuscript in the city of Utrecht some 150 years ago.

3 Chapter 1 was not included because it largely consists of genealogical information of a rather repetitive nature.

2. Siraya primary sources and linguistic literature

There are three broad categories of primary sources for Siraya, each of a very different nature.[4]

(1) 17th century Dutch sources. These are the oldest sources and consist of the following items:

(a) Gravius' translation of the gospel of St. Matthew;
(b) Gravius' translation of the Heidelberg catechism;
(c) the 35-page Dutch-Siraya wordlist or Utrecht Manuscript;
(d) four short dialogues between schoolchildren, which appear as an appendix to the Utrecht Manuscript (1842) and have been analyzed in Adelaar (2006).

The original Gospel translation contains 174 pages of Dutch and Siraya parallel text. A later version was edited by Campbell, who added corresponding lines from the King James Bible in footnotes (Campbell 1888). It was reprinted in 1996 (Campbell 1996).

The Catechism has 288 pages of Dutch-Siraya parallel text. The Utrecht Manuscript is by an anonymous author. It dates from the same period as the Gospel and Catechism but represents a different dialect and was published two centuries later.

A number of factors complicate research on these sources. For one, there are clear dialect differences between the gospel and catechism texts on the one hand (representing the "Gospel" dialect), and the dialogues and wordlist in the Utrecht Manuscript on the other (the "Utrecht Manuscript" [or "UM"] dialect, see further below). The Siraya data used in comparative-historical linguistic studies have usually been taken from the wordlist in the Utrecht Manuscript, which also has a Siraya-English version (Murakami 1933). Furthermore, the gospel text is a translation of the *Statenbijbel*, an official Dutch Protestant bible version which had appeared only a few years earlier (in 1648) and would become a major unifying influence on Dutch language and spelling. In comparison to the King James Bible, it is a more literal translation from the Greek and Hebrew originals. This has important implications for the linguistic analysis of the Siraya gospel text: for an

[4] Incidentally, the Dutch missionaries also collected and published materials for another now extinct Formosan language, Favorlang, which was spoken on Taiwan's west coast to the north of the Siraya region. These materials consist of a dictionary (Happart 1650) and religious teachings and sermons (Vertrecht 1888).

adequate linguistic interpretation, it is imperative to match it against the *Statenbijbel* translation, and no other version.

(2) Land contracts or "Sinkang manuscripts". After the Dutch had left Taiwan and until the early nineteenth century, the Siraya continued to use their writing and spelling. This is evidenced in the survival of 170 land contracts that had been drawn up between 1663 and 1818 by Siraya locals in their dealings with members of the in-migrating and expanding Chinese community. The language of the land contracts might be considered more authentic than that of the 17th century liturgical texts as they were composed by Siraya speakers themselves, but they are highly formulaic and lack the grammatical and lexical variation of the gospel and catechism texts. The contracts are also very difficult to interpret and show regional variety: most contracts are in Siraya Proper, but 23 are in the Taivuan dialect, and four others in the Makatau dialect (Li 2009; see below for dialect variety). The contracts have been studied by the historian Weng Chia-yin (1989a, 1989b, 1990a, 1990b, 1999c) and by Paul Jen-kuei Li, who recently published a comprehensive volume covering all contracts (Li 2010).

(3) Wordlists. When the Japanese assumed control in Taiwan in 1895, Siraya was almost extinct.[5] Nonetheless, some Japanese (linguists and others) were still able to collect a number of fragmentary wordlists, which show forms that are not found in other sources and suggest a greater dialect variation than that reflected in the 17th century texts.[6] Ogawa (1917) made a comparative study of Formosan languages on the basis of these and other lists and distinguished three Siraya dialects: Siraya Proper, Makatau and Taivuan. Li and Toyoshima (2006) is an annotated edition of all Ogawa's lists. Tsuchida and Yamada is an annotated edition of his Siraya data augmented with a few other Siraya worldists collected by Chinese and European scholars (Tsuchida and Yamada 1991).

Linguistic literature dealing specifically with the analysis of the Siraya language consist of an overview of Siraya grammar (Adelaar 1997),[7] two

---

5 According to Tsuchida and Yamada, in 1895 only a limited number of old people still remembered the Siraya language and customs. (Tsuchida and Yamada 1991: 1). Li believes that the language already became extinct about two centuries ago. (Li 2000: 52).

6 Gravius and 17th-century Dutch archives also make mention of a much larger dialect variety (cf. Campbell 1888: xiii, Ferrell 1971: 217–226).

7 This is the published version of a paper with the same title presented by the author at the 7th International Conference of Austronesian Linguistics in

papers by Tsuchida about pronouns (1996) and lexical prefixes (2000), and three papers by Adelaar about spelling and phonology (1999), reduplication (2000) and complex verb morphology (2004). Bien-horn Chen (2001) is a glossed version with Chinese and English translation of the entire Gospel text, together with Siraya-English-Chinese wordlists. Chen (2005) is a similar treatment of the Catechism text.

Finally, a recent sociolinguistic thesis by Huang (2010) details about language revitalisation and identity politics involving Siraya.

3. Dialect variation

From the previous section it is clear that the several Siraya sources together represent a considerable dialect variation. The UM dialect and the dialect of the Gospel and Catechism differ predominantly in their reflexes of Proto Austronesian *d, *R and *S, and *-an, as follows:[8]

Proto Austronesian	UM	Gospel
*d	*s*	*d-*/*r-* in initial position, *r* elsewhere
*R and *S	*x* [9]	*x* in root-final position, *h* or ø elsewhere
*-an	*-ang*	*-an*

The title page of the Gospel states that it is written in the "Formosan" language for the inhabitants of the villages Soulang, Mattau, Sinckan, Bacloan, Tavokan, and Tevorang, and (as the author Daniel Gravius specifies a few pages into the Introduction), "possibly also for some of the people in Dorko and Tilocen". Gravius also explains that he and his co-researcher Anthonius Hambrouck conducted their linguistic fieldwork and Bible translation activities in Sulang (Introduction to Gravius 1661, see Campbell 1888: xiii). Apparently these eight villages had the same language.[10] The title page of

---

Noordwijk (Netherlands) in 1994.

8 For examples, see §2.2.1–3 and Adelaar (2007: 26–27).

9 A voiceless velar fricative, cf. Part B §2.1.

10 George Candidius, the first missionary in Taiwan, worked in Sinckan. He mentions "Sinckan, Mattau, Soulang, Backeroan, Tafalan, Tifalukan, Teopang" and "Tefurang" as eight villages sharing the same customs, religion, way of dressing and language, allowing for minor differences. Except for Tefurang (a mountain village), these villages were all close to the coast, and one could easily reach them all on foot "from the fort" (Fort Zeelandia) and be back within two days (Candidius 1628, see also Blussé et al. 1999: 92, 113).

the Catechism explicitly mentions the *Sideis-Formosaansche tale* ('Sideyic-Formosan language'), which, as Gravius indicates again, is spoken in the eight aforementioned villages (Gravius 1662: Introduction).

There are no indications as to where the UM dialect was spoken. The UM wordlist is titled *Woorden-lijst der Formosaansche Taal* ("Word list of the Formosan Language"), without further specifications as to its provenance.

Tsuchida and Yamada (1991: 7) agree with Ogawa (1917) that the latter's Siraya wordlists represent three different linguistic varieties. Siraya Proper was spoken in the coastal area of Tainan (on the plains of Tainan and Kaoshiung Prefectures), Taivuan was used inland of Tainan Plain to the north [of Siraya Proper area], and Makatau was the language of Kaohsiung and Pingtung Prefectures to the South (Li 2009: 399). Ogawa found the differences between these varieties substantial enough to consider them as languages in their own right, rather than dialects, but Tsuchida and Yamada are more cautious. They show that the varieties in question differ particularly in their reflexes of Proto Austronesian *l and *ɬ:

Proto Austronesian	Siraya	Taivuan	Makatau
*l	r	h or ø	r
*ɬ	l	l	n

However, not all lexical items are predictable in the way they reflect *l and *ɬ, nor do lists exemplifying the same linguistic variety in Ogawa always show the same reflex for *l and *ɬ in each lexical item. The way these reflexes are distributed in Siraya Proper, Makatau and Taivuan (in Tsuchida and Yamada 1991: 8–9) is more a matter of degree than a demonstration of clear dialect boundaries. Rather than three distinct dialect areas, one should perhaps think of a variegated dialect continuum throughout the area where Siraya Proper, Makatau and Taivuan were spoken, and consider these varieties as linguistically random reference points within this continuum.

On the basis of 17th century Dutch sources, Ferrell (1971: 217–226) argued that there must have been five separate ethnic groups in the southwestern plain of Taiwan, each with their own dialect: 1. Siraya; 2. Tevorang-Taivuan; 3. Takaraian (Makatau); 4. Pangsoia-Dolatok; 5. Longkiau. This is quite likely but it is also beyond verification.

It is unclear to me whether the Gospel and UM dialects fit in with the dialect divisions proposed by Ogawa, and if so, how exactly.[11]

---

[11] Li (2010: Introduction) tries to identify the UM dialect with Siraya Proper, and the Gospel dialect with Taivuan, but more evidence of a critical nature is required to make the case.

One thing that seems clear, however, is that the Siraya speech area kept changing as a result of demographic and political changes. While early Dutch sources point out that the Sinkan dialect was ill understood south of Sinkan, the Dutch themselves used it as a medium of instruction in South and West Taiwan. After the Chinese took over the island, some Taivuan speakers moved further eastward into the mountains under the pressure of incoming Chinese in the Tainan area. Some Taivuan speakers even crossed the central mountain ridge and built new villages among the Amis people near Taitung on Taiwan's east coast (Tsuchida and Yamada 1991: 2–3).

4. Some observations about authorship and spelling

Daniel Gravius is officially the sole author of the Siraya Gospel and Catechism translations, but his actual role needs further qualification. When he came to Taiwan (where he only stayed from 1647 to 1651), there had already been many attempts at Bible translations, dictionaries and teaching materials in the Sinkan dialect, including an earlier translation of the Gospel of St. Matthew by Joannes Happartius (Ginsel 1931: 99). In the Introduction to his own Gospel translation, Gravius mentions that he had circulated the text among fellow-missionaries in order to obtain the necessary feedback for an improved final edition to which he had been commissioned (Campbell 1888: xiv–xv). Especially his assistant Anthonius Hambrouck had been helpful in providing him with such feedback. As to the Catechism, this was based on a large corpus of questions and answers in Siraya prepared previously by his colleagues Joannes Happartius and Simon van Breen. Gravius' contribution was to extract a shorter version and edit it for publication (Ginsel 1931: 98). This background information clearly shows that Gravius was not alone in his linguistic endeavours. He received the help of others, and at least to some extent he must have built on the foundations laid by his earlier colleagues.

This is also reflected in the spelling in the Gospel, which shows the hand of several editors.

There are of course several factors contributing to spelling inconsistencies in this text. Among others, they reflect the lack of spelling uniformity in 17th-century Dutch itself, where there were two conflicting spelling traditions as well as a more general confusion about spelling rules. Furthermore, the spelling in the Gospel was not based on a rigorous phonemic analysis, and it sometimes tried to express allegro forms and other pronunciation idiosyncrasies along with more diagnostic forms.

However, there is also some method to the madness. While at first it seems that the spelling is inconsistent throughout the gospel text, a careful computer count reveals that some of the spelling principles used in the first 21 chapters differ more or less consistently from those used in the seven chapters at the end. Such organised inconsistency can best be ascribed to interference by different editors in individual sections.

5. Formosan languages: numbers of speakers and vitality

Although most Taiwanese are nowadays of Chinese descent, historically the island belongs to the Austronesian-speaking world. Its indigenous languages are all Austronesian and are linguistically known as the "Formosan" languages. This term is derived from *(Ilha) Formosa* ('beautiful [island]'), originally a Portuguese name for the island, by which it also became known in other European languages. In the seventeenth century there were at least 25 Formosan languages. Today, ten of these have become extinct,[12] and at least five others are on the verge of extinction.[13] Formosan languages that are not under immediate threat of extinction are those spoken by the Amis (177,000), Atayal (81,000), Bunun (50,000), Paiwan (68,000), Truku (24,000), Puyuma (11,000), Rukai (11,600), Saisyiat (5,300), Tsou (6,500) and Yami (3,500).[14]

It is clear that the vitality of these languages cannot be read from the numbers of their speakers alone. For instance, speakers of Truku are all above 20 years of age, and there is no younger generation to continue speaking the language (Tsukida 2005: 291). A similar situation exists with regard to Puyuma, which has hardly any speakers under 40 years of age (Cauquelin 2004: 322), and with regard to many other Formosan languages with large speech communities. Such a generation gap is more alarming for the chances of survival of a language than are low speaker numbers.

---

12 To wit Babuza, Basay, Favorlang, Hoanya, Ketangalan, Kulon, Qauqaut, Papora, Siraya, and Taokas (Zeitoun 2004: 41).

13 To wit Kavalan, Kanakanavu, Pazih, Saaroa, and Thao (Zeitoun 2004: 41); numbers of speakers of these languages run from several hundred (Saaroa, Kanakanavu), to less than fifteen (Thao, Pazih).

14 Council of Indigenous Peoples 2008 (see http://www.apc.gov.tw/main/index_en.jsp). Note that these are population figures and do not indicate how many speakers each ethnic group language has. Note also that Yami is classificatorily a Malayo-Polynesian (extra-Formosan) language (see fn.21).

There are more than 490,000 Aboriginal Taiwanese.[15] They make up slightly more than 2% of the total population of 23,142,460 million,[16] the majority of which consists of Hoklo Chinese (70%); other large groups are Hakka Chinese (12–15%), and Chinese from various parts of the Chinese mainland, who migrated to Taiwan along with the Kuomintang regime in 1949, as well as their offspring (12–15%; Saillard 2004: 362fn3).

6. The linguistic classification of Formosan languages

There are more than 1,200 languages belonging to the Austronesian language family, which has several primary branches. The number of these primary branches remains a matter of debate, with estimates varying between four and ten (see below). However, historical linguists generally agree that the Formosan languages of Taiwan represent several primary branches of Austronesian, whereas Austronesian languages outside Taiwan all belong to one single branch, called 'Malayo-Polynesian'.[17] The latter are the native languages of large parts of insular Southeast Asia and the Pacific as well as of the island of Madagascar.

The fact that Taiwan alone is home to all but one of the primary branches of Austronesian makes the Formosan languages of particular interest to historical linguists: because of their genetic diversity, these 25 languages theoretically encode much more information about the ancestral Proto Austronesian stock language than all other 1200 Austronesian languages together. It correlates with the fact that Taiwan is of great archaeological value as a stepping stone for the spread of Austronesian speakers, who came from the South Chinese mainland and migrated to Southeast Asia and the Pacific some 6,000 years ago.

The number of primary branches in the Austronesian language tree has often been set at four, after Ferrell's (1969) distinction of an Atayalic branch (consisting of Atayal and Seediq), a Tsouic branch (Tsou, Kanakanavu and Saaroa), a Paiwanic branch (containing Paiwan and all other Formosan languages), and a Malayo-Polynesian branch. It has, however, become increas-

---

15 Council of Indigenous Peoples 2008 (see http://www.apc.gov.tw/main/index_en.jsp).

16 July 2010 figures, National Statistics, Republic of China (Taiwan).

17 One apparent exception is Yami, which spoken in the Republic of Taiwan but belongs to the Malayo-Polynesian branch; however, this language is actually not spoken on Taiwan itself but on Lanyu (Botel Tobago), a small island off Taiwan's southeast coast.

ingly clear that this classification is not accurate, primarily because of the heterogeneity of the Paiwanic branch. Several other classifications have been proposed. The historical linguist Robert Blust (2009: 29–30) distinguishes ten branches. Siraya belongs to Blust's East Formosan branch, together with the Basai-Trobiawan, Kavalan and Amis languages.

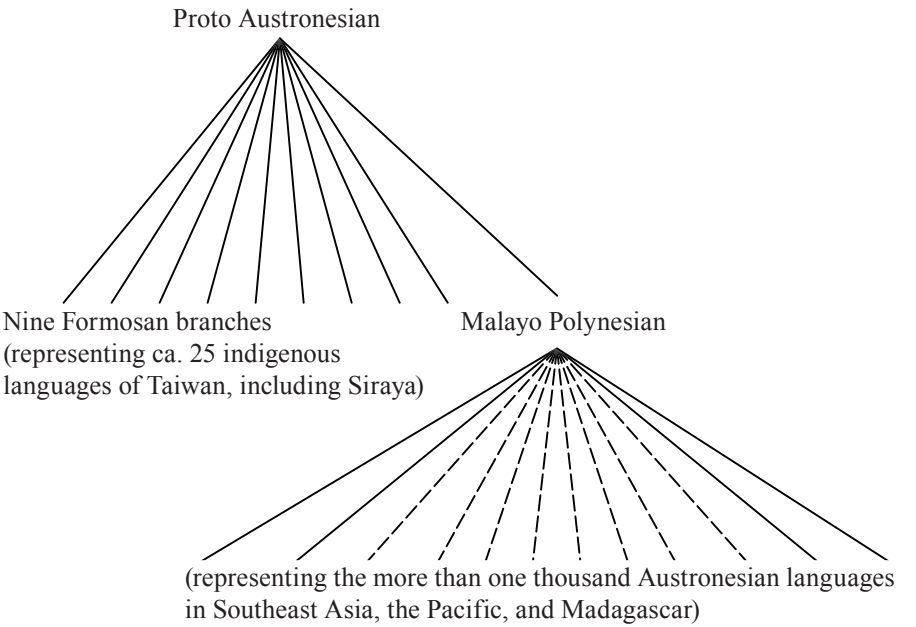

Figure 1. Primary branches of the Austronesian language family tree according to Blust (2009: 29–30).

7. Aboriginal Taiwan and Austronesian prehistory

As indicated above, the linguistic data show that the genetic variety in Taiwan is much greater than anywhere else in the Austronesian language area. They also show that by and large the Formosan languages are phonologically more conservative and complex than the Malayo-Polynesian languages[18]. Both these factors indicate that Austronesian speakers must have migrated from Taiwan to the various regions where Austronesian languages

[18] Phonetic probability and a series of unconditioned mergers of various Austronesian phonemes into one single phoneme in Malayo-Polynesian languages strongly suggest that this is the case.

are spoken today, rather than that they came to Taiwan from any of the other regions where Austronesian speakers currently live.

While the linguistic evidence refutes a 'southern origin' of Formosan speakers, it is not able to trace Austronesian languages back to the Asian mainland.[19] There are currently no Austronesian languages spoken in China.[20] However, archaeology still provides the necessary evidence where the linguistic trail has gone cold. Bellwood (1997: 205–218) demonstrates that there is a geographically and historically continuing trail of neolithic sites beginning in South China and moving via Taiwan[21] into the Philippines[22] and on to Indonesia (Talaud Islands and Halmahera), Sabah and East Timor. These sites contain red-slip ceramics (including decorated and (often) globular vessels), neolithic stone flake tools, and bones of pigs and other animals. The trail also branches off into the Pacific, where it ends in Samoa and is known as the 'Lapita culture' (1,400–800BC).

Although there are no traces of Austronesian languages on the Chinese mainland today, Blust (1999: 73) argues that, given the sinicisation process that has been going on in Taiwan, "it is difficult to imagine that the cultural and linguistic extinction did not occur in coastal regions of southern China and in the Peng-hu Islands (Pescadores) on a much larger scale, leading to the disappearance of any Austronesian or Austronesian-related languages which may have been spoken there prior to European discovery".

8. The Dutch occupation of West Taiwan: historical setting

The Siraya Gospel of St. Matthew (1661) and the Siraya Catechism (1662) were published at the end of the occupation of West Taiwan by the Dutch, which took place between 1627 and 1661. In the early 17th century, the Dutch East India Company had been highly successful in establishing a trade monopoly in large parts of insular South East Asia. It now directed

---

19 Nor is it able to tell whether or not there was a population on Taiwan prior to the Austronesians.

20 Except for Tsat; however, the speakers of this Chamic language migrated to Hainan after the fall of the Cham city of Indrapura in Vietnam in 982 (Thurgood 1999: 225).

21 Represented by the up to 6,300-year-old sites belonging to the Da-Beng-Keng culture.

22 Including the Dimolit site from 2,500 BC in northern Luzon.

its attention to China with a view to monopolise trade between China and Europe, which consisted, among other things, of silk and chinaware. It also hoped to be able to monopolise trade between China and Japan, as the Chinese emperor had prohibited direct trade contacts between the Japanese and the Chinese. But the Chinese, who were not particularly interested in opening up their economy to western countries in the first place, had already granted the Portuguese the rights to establish a small trading colony on China's south coast. The Dutch, not quite realizing whom they were up against, made a display of force to the Chinese and began to attack Portuguese vessels. The Chinese authorities were dismayed by this behaviour and in the end they ordered the Dutch to leave Chinese territorial waters. The latter were allowed to stay in Taiwan, which at the time was foreign territory to the Chinese. They built a stronghold (Fort Zeelandia) on Tayouan, a sandbank off Taiwan's west coast. Initially, their sole objective was to wait for another opportunity to establish trade relations with the Chinese empire, and relations with the local population were to be kept to a minimum. But they would eventually establish their authority in large parts of West Taiwan. They did so mainly at the instigation of Calvinist ministers whom they had invited to Fort Zeelandia. These ministers soon acquired a much better knowledge of conditions on the island and had a better grasp of local intertribal politics than did the local representatives of the East India Company, who became very dependent on them. The ministers' first priority was conversion of the local population, in which they succeeded very well. (Meanwhile, the East India Company obtained a monopoly in the lucrative trade in Taiwanese deer hides). However, Dutch occupation did not last long. From the 1640s onward, China was embroiled in a dynastic crisis. The rule of the Ming had come to an end. In 1644 a new emperor of Manchu origin was established in Beijing and became the first Qing ruler of China. But he still had to contend with dissident armies elsewhere in the country. In 1660 Qing troops defeated Zheng Chenggong (also called Guo Xingye or Koxinga), a trade tycoon-cum-warlord who supported the Ming faction and had maintained a powerful presence on China's southeast coast. Zheng Chenggong was driven from the Chinese mainland and took refuge with his fleet in Taiwan. In so doing he came into direct conflict with the Dutch. He did not accept their claims on West Taiwan and gave them an ultimatum to leave. When this was refused he laid siege to Fort Zeelandia and defeated the Dutch in 1661, chasing them from the island.

Zheng Chenggong's occupation of Taiwan was a decisive factor in the Sinicisation of the island, which until then had had only a few Chinese inhabitants. After Zheng Chenggong's defeat by the Qing armies, Taiwan was

incorporated into the Chinese empire, and the local Austronesian population gradually became a minority.

9. The Siraya people: some historical and ethnographic data

The heavily sinicised Siraya (Pingbu) people have lost much of their original culture, which includes the ability to speak their ancestral language. However, thanks to the detailed observations of Candidius (1628), Campbell (1903) and other missionaries as well as the meticulous records of the Dutch East India Company published by Blussé et al. (1999, 2000), and the ethnographic studies by Shepherd (1995a, 1995b), we have a fairly accurate picture of who the Siraya people were, how they lived, and what they believed.

When the Dutch established a trading post on Taiwan's west coast in 1627, the Siraya speakers they came in contact with were organised in villages, which were in permanent warfare with each other. In the village communities, the women took care of agriculture and religion, while the men were occupied with hunting, warfare and decision-making. Married couples did not live together until late into their marriage: the wife continued living with their parents, and the husband remained in the men's house; the husband would visit his wife by stealth, and if the wife became pregnant, she would undergo abortion, which was performed by *inib*s, female shamans who were in control of Siraya religious matters. Women underwent these abortions until they were in their late thirties. John Shepherd (1995b) examined the reasons behind this peculiar practice. Not satisfied with earlier explanations based on limitations of food supplies (by Montesquieu) or overpopulation due to sexual promiscuity (by Malthus), he has shown that the abortions were based on cosmological beliefs and on the husband's life cycle. The Siraya, like many other traditional Austronesian societies, must have believed that childbirth and childrearing had an adverse influence on success in warfare. This explains why childbirth was postponed until the husband stopped being a warrior and became a community elder. This would happen when the husband was forty; his wife, who usually was several years younger, would still have a few years left to bear children.

The Dutch missionaries were initially not able to impress the Siraya with their religion and their good works (such as healing and improved agricultural techniques).

A turning point came when the missionaries managed to persuade the reluctant East India Company administration to give military assistance to the Sinkan people in battles against their neighbours. This rather unorthodox

missionary strategy had an instant effect on the Sinkan people, who became much more inclined to accept Christianity and to let the Dutch interfere in the organisation of their society. The latter continued their military alliance with Sinkan and managed to pacify and control a large part of West Taiwan. In the Siraya communities they succeeded, among other things, to abolish abortion, ban the inibs, and merge villages into larger units. They also encouraged cohabitation of newlyweds. They combined their missionary activities with a broad program including medical help and education. By the time the Dutch were ousted from Taiwan by Zheng Chenggong in 1661, they had managed to baptise a large number of the Formosans under their control. However, many converts were only nominal Christians, and after the defeat of the Dutch, Zheng Chenggong succeeded in eradicating the new religion. As has often been pointed out in the literature, literacy outlived religion as far as Dutch heritage in Taiwan was concerned: in the early 19th century some Siraya were still able to write their language in Roman script.

10. The Austronesian ethnic groups in the Taiwanese nationalist debate

For most of last century, the Austronesian ethnic groups played a very subordinate role in the political life of their island, and they were often exposed to severe economic, social and cultural oppression. However, this has begun to change in the last two decades or so. The following account is based on Stainton (1999). Somewhat simultaneously with the political liberalisation of Taiwan and the rise of Taiwanese nationalism in the 1980s, the Austronesian ethnic groups underwent an awareness process and strove for recognition of their cultures and their ethnic rights as the nation's earliest inhabitants. In 1984, they formed the Alliance of Taiwanese Aborigines. The Taiwanese nationalists, seeking to differentiate themselves from the Chinese nationalism of the Kuomintang and, later, the PRC mainland, soon began to capitalise on the unique position of the Austronesian groups. Their historical arguments for an independent Taiwan were based, among other considerations, on the fact that Taiwan originally did not belong to what was traditionally considered the Chinese polity (i.e. during the Ming dynasty and before), and that the annexation of Taiwan to China had been a relatively short one (it had lasted for hardly more than two centuries). The presence of an older non-Chinese population clearly underscored that Taiwan was different. Furthermore, some Austronesian groups had traditional beliefs claiming that their ancestors came from the south (and not from the Asian mainland), which in the view of some nationalists added

weight to the original otherness of Taiwan. The most recent evidence adduced in support of the Taiwan nationalist cause is based on gene tests, which show that the Austronesian groups share part of their genes with the Hoklo majority. This line of argument would mean that the majority of pre-Kuomintang Chinese in Taiwan are of Austronesian ancestry, a point in favour of the otherness of Taiwan as a whole vis-à-vis mainland China.

However, supporters of the annexation of Taiwan by the People's Republic of China have also managed to use the case of the Austronesian inhabitants of Taiwan for their own cause. They consider Taiwan a province of China; it came under Chinese administration in the late 17th century – gaining provincial status in the 1880s – and the majority of the population is culturally and linguistically Chinese. They find historical justification for their case in the fact that, in the past, Taiwan was geologically still part of the mainland. Furthermore, according to recent archaeological evidence, the Austronesian inhabitants originally came from the South Chinese mainland. Some supporters of the annexation are keen to point out cultural similarities between the Austronesians in Taiwan and some of the minority groups in mainland China.

Meanwhile, the Austronesian ethnic rights activists in Taiwan are emphasising the fact that they were the first inhabitants of the island. Some of them take pride in the fact that Taiwan is the prehistoric homeland of the Austronesian languages, which nowadays are spoken almost everywhere Southeast Asia, the Pacific and Madagascar. They also seek some differentiation from Hoklo-dominated Taiwanese nationalism.

There is no need to point out the ad hoc nature of most of the historic arguments used in the above debate, and their irrelevance to linguistic analysis. As some of these arguments are based on linguistics and archaeology, however, it is pertinent to reiterate briefly the current position of linguists and archaeologists on the prehistory of Taiwan. Their evidence suggests that this island was the homeland of Proto Austronesian or, at least, the place from where its speakers 6,000 years ago began to spread over Southeast Asia, the Pacific and Madagascar. Before they came to Taiwan, these early Austronesians must have come from the South Chinese mainland, where some 8,000-year-old Austronesian archaeological sites have been found. The linguistic and archaeological evidence clearly refutes a 'southern origin'. From this it may seem as if these disciplines favour the pro-annexation position, but this is not really the case. While Bellwood (1997: 205) and Blust (1999: 70–73) believe that the ancestors of the Austronesians some 8,000 years ago lived on what is currently the South Chinese mainland,

both work on the obvious assumption that at that stage Chinese cultural and political domination had not yet extended that far South.

11. A probable cause of the extinction of Siraya

One may wonder why an apparently important community like the Siraya lost its language, whereas many other, much smaller and politically less prominent groups have been able to maintain theirs. According to Professor Paul Jen-kuei Li, the reason must be the extent to which the Siraya were exposed to Chinese (Hokkien) language and culture. Southwest Taiwan is not mountainous, and it is much more accessible to overseas settlers than most central and eastern parts of Taiwan, where life was no doubt harder, but where Austronesian communities were also much more out of reach of Chinese influence.[23] The fact that the Siraya belong to the southwestern plains, which are much more populated and urbanised than many other parts of Taiwan, and the fact that they interacted more with the Chinese than many other Aboriginal groups did, are the probable causes of their far-going sinicisation and the loss of their language.[24]

12. Attempts at reviving Siraya

The Taiwanese government acknowledges the existence of distinct ethnic groups among the Aboriginal population. Some twelve groups have a separate ethnic status, and various other groups are applying for this. An important condition is to have one's own language. In practice, this is not as straightforward as it sounds, because in some cases the speech of an aspiring group is considered a dialect of the language of another group that has already obtained separate status. In other cases, the language of a group may have lost its importance as a cultural emblem because it has become extinct or is on the verge of extinction (Saillard 2004: 362–363).

---

[23] This incidentally demonstrates a claim often made in language endangerment studies. A language is not only endangered by the small size of its speech community, but also (and possibly even more so) by the regular and intensive exposure of its speech community to another more prestigious language.

[24] Prof. Paul Jen-kuei Li, personal communication.

The Siraya recently applied for separate ethnic status but were unsuccessful. Of course, there is no denial that they are now all speakers of Hokkien Chinese and have not used their original language for more than a century. Nevertheless, *Siraya* still exists as a relatively well documented dormant language.

An organisation called the Tainan Ping-pu Siraya Culture Association has for some time been striving for the maintenance of the *Siraya* cultural heritage, including the revival of the Siraya language. Edgar Macapili, one of the members of the association, has been trying for some years to instil enthusiasm for this language into the Siraya community. In 2002, he wrote a trilingual (Hokkien, English and Siraya) biblical play (Macapili 2002). It was performed in Tainan just before Christmas. The actors were *Siraya* children of all ages. The event obtained much publicity and media coverage. The association was evidently spurred on by the success of another ethnic group, the Kavalan, in obtaining separate ethnic status earlier in the year. (The Kavalan language is on the verge of extinction).

Some who attended the event expressed their doubts at the possibility of reviving a dormant language, especially one that has been almost forgotten for such a long time. Similar doubts were shared by some linguists and other scholars. The author of the play had constructed Siraya sentences from individual words and fairly transparent grammatical elements in the gospel text. The result does not always follow the actual grammatical rules of the Gospel, which is to be expected given the lack of descriptive data available to the author. However, in terms of language revival the project was remarkably successful. Macapili and his Association have recently issued a trilingual Siraya textbook (Macapili et al. 2010).

Another outstanding attempt to give the Siraya community some of its linguistic heritage back were studies mentioned above by Bien-horn Chen, a native-born Siraya.

Part B

A grammatical sketch of Siraya

1. A near-phonemic orthography

1.1. Symbols in the Siraya 17th century materials that were maintained

Some symbols or symbol combinations in these materials do not cause difficulties of interpretation. I maintained the following ones in my near-phonemic spelling, and they do not need further discussion:

a; b; d; f; k; l; m; n; ng; p; r; s; t; v

I also maintained **nḡ** and **z**, although they are problematic and will be discussed further in respectively §1.13 and §1.14.

1.2. Overview of changes made to the 17th century orthography

I made orthographic changes involving the following symbols and symbol combinations, which I explain in the subsections below. I will implement these changes in the near-phonemic spelling throughout this book:

Gospel ⟶	My spelling
æ	*ä*
æu	*äw*
e	either e, schwa or *ä*
i, (syllabic) **j**	*i*
j, (non-syllabic) **i**	*y*
y, (syllabic) **j**	*ĭ*
-**ija**-	-*ya*-
-**ei**	-*ay*
o	-*o* / (*u*)
ou	u / (-o)
-**ow**	-*aw* (i.e. ‘kow’ → =kaw)

uu	*u*
u	*ŭ*
u in alternation with **i**	*ə* (schwa)
-au	*-aw*
auwa-	*-awa-*
ouwa-	*-ua-*
w (elsewhere)	*w*
k, **q**; (only in the UM:) **ch**	*k*
c (in **coh** '1S . NOM')	*k*
c (elsewhere)	*c*
g, **gh**	*x*
h	*x* (if in free alternation with 'g' or 'gh') *ø* (if immediately preceding another consonant) *h* (elsewhere)
double consonants	single consonants (with *breve* on preceding vowel)
vowels preceding double consonants are not systematically marked	vowels preceding double consonants are marked with superscript ˘ (*breve*)
' indicates vowel deletion, loss of historical consonant, or syllabicity of consecutive vowels	' indicates vowel deletion
' in **t'e** (C142v) 'misery' and its derivations	*ʔ* (indicating a possible glottal, see §1.18.1)
¨ indicates syllabicity of a vowel	¨ indicates palatalisation in *ä*
ˆ indicates vowel contraction	ˆ is not maintained (vowels are represented in their uncontracted form)
- indicates morpheme boundaries	- also indicates morpheme boundaries but in a more rationalised fashion

1.3. **æ** is a palatal *ä*

æ represents a palatalised counterpart of **a** and often alternates with it in the dialect of the gospel text and the catechism.

I write **æ** as *ä*, which brings out more clearly the relation between **æ** and **a**. Examples:

tmætææm (xx: 25)	'to call'	*t‹m›ätääm*
mæhæp (xix: 12)	'to castrate'	*m-ähäp*

Conditions of palatalisation from *a* to *ä* and from *u* to *äw* are outlined in §2.2.3.

In 17th century Dutch **æ** stands for a long fronted palatal [ä]. The fact that **a** frequently occurs before double consonants, whereas **æ** usually does not, suggests that the latter is long. Nevertheless, there are three words in which **æ** is always followed by doubled consonants, which I indicate as ă:

pærænnæh (iii: 10)	'tree'	*pärănäx*
ænnæi (xxiii: 13)	'woe unto…'	*ănäy*
mæænnigh (xiii: 45)	'beautiful'	*mä-ănix*

1.4. Siraya must have had a schwa

The assumption that Siraya must have had a schwa is based on the following observations.

(1) **u** and **i** both denote short vowels. They are sometimes in free variation with one another, and when they are, they often have a corresponding form with a schwa (or a vowel reflecting a historical schwa) in other Formosan languages. In such cases, it seems that the translators of the gospel text had difficulties in diagnosing a schwa and used **u** or **i** instead, hence their alternating occurence.

Compare:

Gospel (or UM)	Comparative evidence	Diagnostic structure
'addim (xiii: 7), **addum** (xxvii: 29) 'thorn'		*ădəm*
ninnim (xiii: 8), **nnum** (xvi: title) 'six'	PAn **nəm(nəm)*	*nə-nəm, nəm*
'tdarim (xvii: 9), **'td-darum** (viii:1) 'to go down'	PAn **daləm*	*t-darəm*
voukugh v: 36, x: 30), **voukig** (UM) 'hair'	PF **bukəS* 'head hair'[1]	*vukəx*

[1] Labeled as Proto South Formosan by Tsuchida (1976: 219) and as Proto Formosan by Blust (n.d.).

(2) A related case of free variation is observed in the undergoer suffixes. Orthographically, there is quite a variety of endings involving final n. However, these reflect only two morphemic suffixes. The endings **-en**, **-in**, **-un** and **-'n** all reflect the undergoer suffix *-ən*.

Gospel (or UM)	(my spelling)
æillingigh-'noumi (RDP-hear-UO=2P.GEN) 'heard by you'	*ä-ilingix-*[*ə*]*n=umi*
ni-tnamsing-enhou (PST-believe-UO=2S.GEN) 'was believed by you'	*ni-tna-m'sing-ən=hu*
kannin (eat-UO) 'to be eaten'	*kan-ən*
pa-kan-nun (CAUS-eat-UO) 'to be fed'	*pa-kan-ən*
ni-lpogh-un (xxi: 35) (PST-kill-UO) 'was killed'	*ni-l'pux-ən*

-an often reflects the UO suffix *-an*:

ni-'æuloug han (iii: 6) (PST-baptise-UO) '(they) were baptised'	*ni-äwlux-an*

However, after a root ending in *-a*, it can also be a contraction of this *-a* with the undergoer suffix *-ən*:

ka-kytt'an-oumi (RDP-look-UO=2P.GEN) 'seen by you'	*ka-kĭta-*[*ə*]*n=umi*

(3) Some written heterorganic consonant clusters seem phonotactically odd and imply the presence of a schwa. This schwa is often corroborated by comparative evidence. Examples:

Gospel (or UM)	Comparative evidence	Diagnostic structure
tbung (xxiii: 14) 'spouse'	PSF *Cəbung 'to meet' Tsuchida (1976: 215)	*təbŭng*
matmoei (xv: 37) 'full'	PSF *təmuy 'many; full' Tsuchida (1976: 241)	*ma-təmuy*
pchag (UM) 'pounded rice'	PAn *bəRas[2]	*pəxax*

[2] *b > *p* is irregular.

On account of these three factors I assume that Siraya had a phonemic schwa. I write schwa in the undergoer suffix *-ən* throughout this study. If a root contains a schwa, I indicate this in the lexicon in a phonemic representation of this root between slashes, which appears directly after its headword.

Note that there must be cases with **i** or **u** representing a schwa that do not have a corresponding form with respectively **u** or **i**. In such cases it will sometimes be possible to identify a schwa with the help of comparative-historical evidence. For instance, **vullum** (vi: 26) 'sky' has no variant form containing **i**, although comparative evidence suggests that both its vowels are schwa (cf. Paiwan *və-ɬələm* 'overcast, shadows of clouds' Ferrell 1982).

1.5. **e** stands for *e*, *ə*, *a*, *ä* or *i*

In a small number of cases, **e** seems to refer to a mid-vowel *e* in its own right (although its phonemicity must be a recent development).[3] This mid-vowel is represented in the following examples:

Gospel	(my spelling)
veiagh (xxv: 29) 'each, every'	*veyax*
akoumea (xxv: 29) 'to have'	*akumea*
-kame (ii: 2) '1PE.NOM '	*=kame*
vare (vii: 25) 'wind'	*vare*
matt'e (v: 39) 'evil', **paka-t'e** (xiii: 57) 'to offend'	*ma-tʔe, pa-ka-tʔe*
mapoule (xxviii: 3) 'white'	*ma-pule*

In most other cases, however, **e** stands for another phoneme, either as an orthographic variant (of *ə* or ä) or as an allophone (of a or i).

(1) As already discussed in §1.4, **e** stands for a schwa in the suffix *-ən*.

(2) **e** sometimes represents *a* when it precedes *i* on morpheme boundaries:

Gospel	(my spelling)
meirang 'lord' ← *ma-* 'AO1' + *irang* 'great'	*ma-irang*
meireirang (vi: 24) 'lords' ← *ma-* + *ira-irang* (= *irang* + Reduplication)	*ma-ira-irang*

[3] In the examples to follow, the *-eya-* and *-ea* vowel clusters must have developed from *aya (see §1.5(2)), and *vare* and *-kame* reflect PAn *bali and *kami respectively. This *i > *e* development is in line with the history of another PAn midvowel *-u, which usually became *-o* (see §1.7).

Gospel	(my spelling)
mamei-mang (vi: 28) 'how' ← *mama* 'like' + *(i)mang* 'what'	*mama-imang*
peila (xiii: 44) 'to buy' ← *pa-* 'CAUS' + *i-la* 'to do again; to exchange'	*pa-i-la*
peileil'ig (viii:15) 'to serve' ← *pa-* 'CAUS' + *ili-ilix* 'to honour' (= *ilix* + Reduplication and dissimilation of *i* to *e*, see below).[4]	*pa-ile-ilix*

Furthermore, in the final diphthong -**ei**, **e** also represents *a* and has undergone hightening through vowel colouring by final *y*. This hightening does not take place in the first sequence of a reduplication of a root ending in -**ei**, because here the final *y* is deleted, e.g., *ma-pata-patay* 'to brawl, scuffle' etc. (see further §1.6.2).

(3) Before **u**, **e** is sometimes in free variation with **æ**. In this position, there is a correlation between the presence of **e** and **æ** in the Gospel and that of **g** in the UM. Examples:

Gospel	Variant form in gospel	UM dialect	(my spelling)
kaeuloung (v:11) 'person'	**kaæuloung** (xii:12)	**cagoulong**	*kaäwlung*
euyng (v:15) 'candle'	**'æuyng** (vi:22)	**ouging**	*äwĭng*
Euma (v:35) 'village, town'	**æuma** (ix:1)	**gouma**	*äwma*

Cases where **e** is in free variation with **æ** (and/or precedes **u**) are the result of vowel colouring caused by an earlier *x which was subsequently lost in the gospel dialect. This *x had a markedly palatalising effect on neighbouring vowels, changing *a to *ä* and *u to *äw*. It is maintained in the UM dialect, where it is written as **g** and does not have that palatalising effect (cf. §2.2.3). I write a palatal *ä* in cases where **e** is in free variation with **æ** (and/or precedes an **u**).

(4) Two consecutive *i*'s often dissimilate to *ie* or *äi*, a process which usually happens on morpheme boundaries. Progressive dissimilation from *ii to *ie* does not always take place, and in some words there is free variation between *i* and *e*. Examples:

[4] Apparently with regressive rather than progressive dissimilation (possibly through assimilation to the preceding *-e-i-* vowel sequence?)

rii (ix:17), **rïé** (ix:16) ‘old’ (*rii* →)	*rii, rie*
ni-jerroua (xxvii:45) ‘came’ (PST-*irua* →)	*ni-erua*
ierppa’ni-an (xxviii:13) ‘our sleep’ (RDP-*irip*-UO=1PE.GEN →)	*i-er’p-an=ian*
hæuugh ka äyht-ten (xxvi:27) ‘drinking cup’ (cup LK RDP-drink-UO →)	*häux ka ä-ĭt-ən*
’æymhgan (xxii:17) ‘tribute, customs’ (RDP-pay.tribute-UO →)	*ä-ĭm’x-ən*
’æilling-hen (xiii:13) ‘listening’ (RDP-listen-UO →)	*ä-ĭling’x-ən*
ni-ennagh (xii:23), **ni-innagh** (vii:28) ‘they were amazed’ (PST-ĭnax →)	*ni-ĕnax, ni-ĭnax*
ni-pa-ieroung-en tyn (xiv:19) ‘he let [them] sit down’ (PST-CAUS-LOC-*irung*-UO →)	*ni-pa-i-e-rung-ən*

1.6. A re-definition of **i**, **j** and **y** according to the syllabic length that they indicate

i, **j** and (to a less extent) **y** indicate a high front vowel *i*. Examples:

assi ‘no(t)’	*ăsi*
dmyllough (vi:13) ‘to follow’	*d‹m›ĭlux*
jna (xxvi:7) ‘woman; mother’	*ina*

j is a rather infrequent symbol; it can also stand for a palatal semivowel. Examples:

jouagh (viii:31) ‘herd’	*yuax*
ajajam (vi:26) ‘animals’	*ay-ayam*
rmaja tabe (v:47) ‘to greet’	*r‹m›aya tabe*

i is also used in final diphthongs:

vavoei (viii:30) ‘pig’	*vavuy*
alæi (xviii:7) ‘woe’	*aläy*
mapatei (xxii:24) ‘dead’	*ma-patay*

Sometimes, **i** occurs in free variation with **j** and **y**, and **j** occurs in free variation with **y**, as in the following examples:

jna (occurs 30x) and **ina** (10x) 'mother'	*ina*
jmagh (xxv:3, xxv:4; 2x) and **imagh** (xxiv:3, xxv:8, xxvi:30; 3x) 'oil'	*imax*
jngau (xix:12) (1x) and **'yngau'** (5x) 'voice'	*ĭngaw*
jrou (47x) and **irou** (146x) 'if'	*iru*
kytta (xiii:15) (120x) 'to see' and **pakita** (xxiv:1) (1x) 'to show'	*kĭta; pa-kĭta*
ringei (v:16) (12x) and **ryngei** (xxiii:3) 'work'	*ringey, rĭngey*

Some variant forms are highly exceptional, such as *pakita* (xxiv:1). It may be a typographical error, as it occurs only once as against *kytta* occurring 120 times.

This spelling variation in the gospel must be due to a combination of free variation, wrong spelling and different editorial policies (see *Introduction*), making it difficult to establish the exact nature of the vocoids behind **i**, **y** and **j**. These letters will be discussed individually below.

*1.6.1. **y** stands for a short high front vowel* ĭ

In Dutch spelling, double consonants indicate shortness of the preceding vowel. If Siraya **y** occurs in non-final syllables, it is, as a rule, followed by double consonants, as in **kytta** and **dmyllough**. It must stand for a short high front vowel. Length contrast is phonemic in at least one case: compare **rymma** (xiv:21) 'five' vs **rima** (iii:12) 'hand'.[55]

I write this vowel as *ĭ*, and I furthermore assume that **y** also stands for *ĭ* in final syllables and elsewhere where it is not followed by double consonants, as in the following examples:

madys (xiv:22) 'quick'	*ma-dĭs*
kmyta, **kytta** (xiii:15) (ix:22) 'to see'	*k‹m›ĭta*
yna (v:35), **ynna** (i:20) 'don't'	*ĭna*
ymi-æn (vi:12) '(1PE.GEN)'	*ĭmiän*
myligh (vi:24) 'to serve'	*m-ĭlix*

[5] Although these words are obviously historically related, they reflect a length opposition in PAn, for which David Zorc reconstructs *lĭma 'five' and *(qa-)lima 'hand' (David Zorc n.d.).

*1.6.2. **i** stands for a long high front vowel* i *except base-finally after* e*, where it stands for a palatal semivowel*

In contrast to **y**, **i** almost never occurs before double consonants.[6] Since it does not, and since **y** is a short vowel, **i** must be a long high front vowel. I maintain it as *i*, as already demonstrated in some previous examples, e.g. **sivægh**, **sivæh** 'wheat' → *siväx*; **pihgik** 'altar' → *pixik*; **rima** 'hand' → *rima*.

Base-finally after a vowel, however, it stands for a semivowel, which I write as *y*. Supplementary evidence for analyzing base-final **i** as a semivowel is found in the reduplication of word bases ending in a diphthong. Full reduplication of a word base follows a CVCV-CVCVC pattern, with deletion of the last consonant in the first sequence of the reduplication. If a base ends in a diphthong *-ay* or *-aw*), it is the coda of the diphthong which is deleted, which therefore is structurally a consonant. Examples:

mapatey (xx:24) 'dead; die'	**mapatapatey** (UM) 'to scuffle, brawl'	*ma-pata-patay*
ringei (v:16) 'work'	**ringa-ringei** (xxiii:5) 'works'	*ringa-ringey*
—	**vouna-vounei** (xvii:15) 'often'	*vuna-vunay*
—	**hmouwahouwau** (xiv:30) 'to cry'	*hɯmɤua-huaw*

*1.6.3. **j** stands for a palatal semivowel* y *and (sometimes) for a high front vowel* i

j basically represents a palatal semivowel, whether it stands alone, as in **jouagh**, **ajajam** and **rmaja** (see above), or it is preceded by **i**, as in the following instances:

vaija (v:13) 'salt'	*vaya*
keija-en (v:31) 'wife'	*k-äya-en*
kavei-joung-an (xix:9) 'fornication'	*ka-veyung-an*

Wherever **j** and **ij** appear intervocalically and stand for a semivowel, I write *y*. **j** may occasionally indicate a vowel and is then often in variation with **i** or **y**. It occurs as a vowel in the following forms:

---

[6] There are exceptions, such as **littou** (vi:13) 'Devil', which however occurs only once and has a variant **Lyttou** occurring at least 27 times.

jdarinoughan (xii:32) 'century; world'[7] (1x) vs **yddarynnoughan** 'id.' and **myddarynnough** 'eternal' (together 12x)	*i-da-rinux-an, m-i-da-rĭnux*
jlalimough-an (x:22) 'the end' (1x)	*i-la-limux-an*
jmagh (xxv:3,4) (2x) vs **imagh** (3x) 'oil'	*imax*
Tama-jmuh (xvii:24), **Tama-jmug** (xxi:31), **Tama-jmugh** (xxi:32) (together 3x) vs **Tama-imugh**, **Tama-imuh**, **Tama-imigh** (together 7x) 'tax collector'	*Tama-imix, Tama-imŭx*
jna (30x) vs **ina** (10x) 'woman; mother'	*ina*
jngau (xii:19) (1x) vs **yngau** (5x) 'voice'	*ĭngaw*
jrou 'if, when' (47x) vs **irou** (146x)	*iru*
jtou- 'to be at' (xxvi:36, xxvi:38, xxvii:11, xxvii:33) (4x) vs **itou**- (83x)	*itu-*

Forms with **j** referring to a vowel are less frequent than their counterparts having **i** or **y**, but there are exceptions, such as **jlalimough-an** and **jna**. The former is not significant as it occurs only once. The case of **jna**, however, is remarkable. A possible reason for its frequent spelling with **j** is to contrast it with a near-homonym **ynna** 'don't', which is occasionally spelled **ina** (2x). I write *i* wherever **j** alternates with **i** and stands for a vowel. I write *ĭngaw* (with short *ĭ*) for **jngau**.

I also write *i* for **j** and **y** in **jdarinoughan**, **yddarynnoughan**, **myddarynnough** and **jlalimough-an**. These vowels represent a location-oriented prefix which is generally written with **i** (316 times). It is written with **y** (/**j**) in only 18 cases (less than 3 percent), 17 of which are long words having the **y** or **j** in syllables preceding the antepenultimate.[8] The short vowels in these 17 cases must be due to their position relatively far away from the root, which (presumably) contains the stressed syllable.

In initial position the semivowel *y* is rare, with only two instances: *yuko* (ix:36) 'sheep' (a Hokkien loanword) and *yuax* (viii:31) 'herd'.[9]

In **Joep-an** 'ghost', **j** does not stand for a semivowel. Variant forms with **i** are much more frequent, and words related to Joep-an show that their base is *(h)ĭup*, cf. **Joep-an** (i:18), **I-oep-an** (iii:11) 'ghost', **mioup** (xxv:8) 'to blow out', **pa-i-oup** (xii:20) 'to blow out something'; cf. also PAn *Seyup 'to blow'.

---

7 Apparently, the basic meaning of *i-da-rinux-an* is 'infinity (of time and space)', and that of *m-i-da-rinux* 'endless (in space or time)'.

8 In the remaining case, **'ni-pa-yh-patei'** (xiii:7) '(they) choked (them)', **'y'** appears in the antepenultimate itself.

9 Also a Hokkien loanword?

Nor is **j** a semivowel in **jauwan** '(1S.GEN)': it is basically a prefix *ǐ-* which is also found in other independent and oblique pronouns for the 1st and 2nd person, such as the independent pronouns *ǐmhu* '2S.NOM', *ǐmumi* '2P.NOM', *ǐmǐta* '1PE.NOM' (and their oblique counterparts *ǐmhu-än*, *ǐmumi-än*, *ǐmǐtä-n*) (see Table 2 in §4.2.1).

j also occurs once in **matt'moej** (xiv:20) 'full' and once in **hejrou** (xviii:3) 'if'. These forms are normally written **mat(t)moei** and **heirou** respectively and I consider **j** here a typo.

1.7. **ou** and **o** stand for respectively *u* and *o*

ou stands for a (long) high back vowel. Examples:

rarouha (xviii:16) 'two'	*ra-ruha*
tatouro (xviii:16) 'three'	*ta-turo*
pourough (ii:16) 'country'	*purux*
arough (iv:15) 'opposide side'	*arux*
sat ka'-ætou-han (viii:13) 'one hundred'	*saat ka-ätux-an*
ta toulou-an (xiii:40,52) 'oven'	*ta-tulu-an*

ou is sometimes in free variation with **o**. Examples:

tatouro (xviii:16) 'three' and **katatourou** (xxi:35) 'third'	*ta-turo, ka-ta-turo*
koulamogh (xi:8) and **koulamough** (ix:16) 'clothing'	*kulamux*
siouro (v:26) and **si-ourou** (v:19) 'ahead, the first, foremost'	*si-uro*

However, **o** seems to prevail in word-final position, and **ou** in penultimate syllables and in closed final syllables. Final **o** developed from PAn (or PF) *-u(ø/q/h/H), as is shown in the following words:

Siraya	PAn (or PF)	My spelling
ti ano (xii:47, xiii:19) 'someone' (2x), **ano** '(vocative particle)' (1x)	*n-anu 'what'	*ti ano, ano*
vahto (xxiii:37) (24x), **vahtou** (xxi:44) (1x) 'stone'	*baCu	*vato*
touro (19x), **tourou** (9x)	*təlu	*turo*
siouro (19x), **si-ouro** (6x), **siourou** (3x), **si-ourou** (5x) 'ahead, first, foremost'	*qulu	*si-uro*