Reviving Siraya:
A Case for Language Engineering

Alexander Adelaar
The University of Melbourne

Siraya is a language once spoken in Southwest Taiwan, which is being revived. Some Siraya data is inconsistent, requiring strategies as to how it will be implemented. I discuss some of these strategies in support of the revival attempt. The following issues deserve attention:

1. Siraya phonology includes a schwa ($ə$), although it is ignored in the original orthography. The choice here is between keeping this orthography and ignoring schwa, or re-establishing schwa and changing the orthography.

2. Siraya had maintained part of the original Proto Austronesian voice system. However, this system was losing some voice oppositions and was being re-aligned when Siraya was still spoken. Two approaches are possible: keeping the original Siraya voice system, or adapting to the tendencies to change, which were strong but had not yet taken their full course.

3. Siraya had at least three dialects, two of which are particularly useful for revitalization. In order to build a lexicon for a revitalized Siraya, should the vocabulary of these dialects be combined without further ado? Or should the words from one dialect phonologically be adjusted to the other? Is there a cause for revitalizing various dialects?

4. Siraya had "anticipating sequences", whereby a formal part (an initial consonant, a syllable, or two syllables) of the lexical verb is prefixed to the adverbial head. Anticipating sequences abound in one dialect but are absent in the other. As it is a rather complicated and irregular feature, should it be taught in modern Siraya classes? And if so, how should it be taught: in all its complexity, or in a somewhat simplified version? Or can it be ignored without causing too much structural imbalance to the grammar?

1 INTRODUCTION.[1] Siraya is a dormant language once spoken as a first language in the Tainan region in Southwest Taiwan. A form of it also became a lingua franca when, in the first half of the 17th century, the Dutch used it for liturgical, educational and administrative purposes in areas of Southwest Taiwan where other languages (or non-lingua franca varieties of Siraya) were spoken.

Siraya became a dormant language at the beginning of last century, although reportedly there are people today who still remember Siraya words (Huang, Macapili & Ta-

---

lavan 2013).[2] Furthermore, members of the Siraya community have for somewhat more than a decade been trying to revitalize the language and to obtain official tribal status for their group. While still denied by the central government, this status was granted by the government of Tainan City in 2005. An unofficial survey estimates the population at approximately 60,000. Siraya identity and aspects of the Siraya culture and belief system are still alive, and various Siraya villages have their own cultural association. Among modern Siraya cultural revitalization groups, some Christian ones concentrate on language reclamation, whereas other groups are more preoccupied with reviving the indigenous religion and cultivate traditional lyrical chanting, the linguistic content of which is largely beyond retrieval. In general, religion is no impediment to cooperation. The Tainan Pepo Siraya Cultural Association (TPSCA), chaired by Mrs. Uma Talavan, is based on the joint efforts of Christians and traditional believers and has become the leading activist group in the Siraya renaissance movement. Since its foundation in 1999, the TPSCA has been active in many projects, such as the organization of various summer camps for the promotion of Siraya language and culture (since 2007) as well as teacher-training seminars and language classes in four elementary schools (since 2009), the compilation of a trilingual Siraya-Chinese-English dictionary (Macapili 2008), and the development of language course material (in 2012) (Huang, Macapili & Talavan 2013).

The TPSCA uses a corpus consisting of 17th century Dutch materials and some vocabulary from wordlists collected by Japanese (see below). TPSCA activists were also able to collect some data from "rememberers" of individual words in 2007. Other sources do exist but the TPSCA considers their use too problematic because they were recorded in Chinese writing, which is phonemically imprecise (Huang, Macapili & Talavan 2013). In Siraya ethnic regions, Siraya is now being taught in 11 elementary schools and practiced in church via conversation and singing. The latter is often accompanied by Siraya traditional musical instruments.

Cooperation between the present author and the TPSCA has always been very cordial, but it has not been intensive. There are practical reasons for this. I started my investigations into Siraya grammar in 1991, almost a decade before the TPSCA was founded. Furthermore, my research was not based on fieldwork but on the analysis of literary texts, namely translations of the Gospel of St. Matthew (Gravius 1661) and the Heidelberg Catechism (Gravius 1662) into Siraya made by Dutch missionaries in the 17th century. There was also something of a "language barrier" inherent in the research data. On the one hand, the success of the Siraya revitalization project depends very much on the way this language is transferred to speakers of the local Chinese language[3], which I do not know at all but is a native language to TPSCA members. On the other, the successful analysis of the 17th century Siraya texts depended very much on knowledge of Austronesian comparative linguistics and the ability to check the 17th century Dutch liturgical texts on which these Siraya

[2] 1908 is often given as the year in which Siraya became dormant, although Huang, Macapili & Talavan (2013) point out that this is a *datum post quem* (i.e. the last date when Siraya was still known to be spoken, rather than the earliest date when it had became dormant).

[3] The most widespread variety of Chinese spoken in Taiwan is Southern Min or 'Taiwanese' (a variety of Hokkien Chinese).

translations were originally based. This is where my expertise came in, as I am trained in Austronesian comparative-historical linguistics and a native speaker of Dutch. Finally, the TPSCA members have clear aims in mind and are an extremely professional group when it comes to applied linguistics, curriculum design and public relations. They are much more successful in these fields than I am. I doubt that my usefulness extended much beyond sharing the linguistic analysis I was working on (in which I had a head start) and, on occasion, moral support. In summary, our aims were clearly complementary but did not overlap on a day-to-day basis. The conditions under which we interacted were much more relaxed than those involving First Nation languages in Canada (Czaykowska-Higgins 2009, Rice 2010, and Gerdts 2010), or native languages in Australia and most other countries where native languages have become marginalized because of colonization. The TPSCA has always welcomed the support of linguists, notwithstanding the scepticism they sometimes encountered from the linguistic establishment in Taiwan at the time they started their campaign (Huang, Macapili & Talavan 2013).

There are four categories of original Siraya language sources.

Materials produced by the Dutch in the 17th century consist of the Gospel of St. Matthew (Gravius 1661) and a Catechism (Gravius 1662) in one dialect (the Gospel dialect or "G dialect"), and a wordlist with very short dialogues in another dialect (the Utrecht Manuscript dialect or "UM dialect") which was rediscovered in an archive in the city of Utrecht in the Netherlands (Van der Vlis 1842). These data are limited but are workable as far as their amount and consistency is concerned. Adelaar 2011 is an analysis of the Gospel text and consists of a grammar, lexicon and texts.

Siraya-Chinese bilingual land contracts were drawn up throughout the 18th century until the beginning of the 19th century. Li 2010 is a comprehensive collection with translation and annotation of these contracts. Their content is formulaic and often repetitive, which makes them unsuitable for a language revitalization project.

At the beginning of the 20th century, the Japanese collected wordlists of a large variety of dialects. The quantity, reliability and general usefulness of these lists is limited. They were published by Ogawa (2006 [original manuscript 1917]) and by Tsuchida & Yamada (1991), who all conclude that the lists represent three main dialects (Siraya Proper, Makatau and Taivuan), each with their sub-dialects. According to Li, these three dialects are also reflected in the land contracts. He furthermore tries to identify Taivuan with the Gospel dialect and Siraya Proper with the UM dialect (Li 2010), but the evidence is ambivalent.

For an overview of recent publications dealing with the Siraya language data, the reader is referred to Li 2001, Tsuchida 2001 and Adelaar 2011. A publication deserving mention here is Jimmy Huang's PhD dissertation on Siraya language revitalization and identity politics (Huang 2010).

The 17th century materials are clearly more suitable for language revitalization than the later materials, which explains why the Siraya community has chosen them accordingly for this purpose. However, the data are by no means homogeneous and uniform. Apart from representing two distinct dialects, they also exhibit spelling inconsistencies and some unstable grammar, requiring clear strategy decisions before they can be used for language revitalization.

In this paper I discuss some of the main problems that I came across when I analysed the 17th century materials. I show various strategies that can be followed to solve these

problems, evaluating their pros and cons. In doing so I hope to facilitate the work of Siraya language activists and to contribute to the revitalization of Siraya.

In Section 2, I discuss the presence of a schwa, which is not indicated in the spelling but obviously existed as a distinctive phoneme in Siraya, to witness the various inconsistencies and some unnatural phoneme constellations that occur in the spelling. The question here is, should one follow the original spelling or re-establish the underlying phonemic distinction?

In Section 3, I investigate the existence of two non-agent voice suffixes, a feature which clearly reflects a morphemic feature in many other Austronesian languages and was inherited from Proto Austronesian. However, in the Gospel and Catechism, these suffixes are used rather inconsistently, and there are clear signs that they were in the process of becoming merged. Therefore, should this morphological opposition be maintained or ignored?

In Section 4, I deal with vocabulary building. The obvious problem with the 17th century materials is that they represent two dialects. Possible solutions are to combine vocabularies of these dialects without further ado, or to adjust the words from one dialect phonologically to the other dialect, which would then function as a basis for a single revitalized language. I also touch on the theoretical possibility of revitalizing several Siraya dialects.

Finally, in Section 5, I explain the workings of a rather unusual phenomenon in verbal morphology of the G dialect. I also discuss the need to maintain this phenomenon in a revitalized Siraya, given the fact that it is represented in only one of the dialects, it is somewhat unstable, and it is likely to cause difficulties to learners of Siraya. Is it worth maintaining this feature and teaching it in modern Siraya classes? If so, how should it be taught: in all its complexity, or in a somewhat simplified version? Or can it be ignored without causing harm to the internal structure of the language?

I give an account of the Siraya activists' position on these issues and make some concluding remarks in Section 6.

For a better understanding of the verb forms used in the sample sentences in this paper, it is useful to know that Siraya verbs belong to four formal classes.

Class 1 verbs are generally marked with *ka-*; however, when they are in active voice they have the prefix *ma-*, e.g. *ma-tukul* 'to be injust' and its nominal derivation *ka-tukul-an* 'injustice'. These verbs are usually stative.

Class 2 verbs have no affixation, e.g. *thabul* 'to travel', *tawrŭma* 'to get lost'.

Class 3 verbs have an infix *<m>* in active voice if they begin with a (non-labial) consonant, e.g. *k<m>an* 'to eat' (< *kan*). If they begin with a vowel they have *m-* prefixed, e.g. *m-ad* 'to bring' (< *ad*). These verbs are, as a rule, transitive.

Class 4 verbs are derived with *ma-* and *pa-* : *ma-* is prefixed if they are in the active voice, and *pa-* is prefixed in all other cases, e.g. *su* 'word' becomes *ma-susu* 'to say' vs. *pa-susu-ǝn* (same verb in patient form).

2. SCHWA IS PHONEMIC BUT IGNORED IN THE ORIGINAL SPELLING. Siraya must have had a phonemic schwa (mute *ǝ*). The fact that it is not indicated in the original spelling is most likely due to the fact that Dutch (and, for that matter, English) spelling has no dedicated symbol for this sound. Moreover, the 17th century missionaries working on Siraya were not aware that this language had a schwa as a separate distinctive unit. As a

consequence, they did not write it or wrote it as a free alternation between a short *i* and a short *u*. In the case of the suffix -Vn, they wrote it alternatingly as a 'e', 'i', 'u' or zero.

A. The lack of a schwa is seen in some written heterorganic consonant clusters that are phonotactically odd, implying the presence of a schwa between their components. This schwa is often supported by comparative evidence. Examples:

original spelling	comparative evidence	phonological spelling
tbung 'spouse'	PAn *Cəbung 'to meet'	*təbŭng*
matmoei 'full'	PAn *təmuy 'many; full'	*ma-təmuy*
pchag 'pounded rice'	PAn *bəRas[4]	*pəxax*
rbo 'inside'	PAn *ləbu 'interior'	*rəbo*

B. u and i both denote short vowels. They are sometimes in free variation, and when they are, they reflect a historical schwa, which is borne out by comparative linguistic evidence. Compare:

original spelling	comparative evidence	phonological spelling
'addim addum 'thorn'	--	*ădəm*
ninnim, nnum 'six'	PAn *nəm(nəm)	*nənəm, nəm*
'tdarim, 'td-darum 'to go down'	PAn *daləm	*t-darəm*
voukugh, voukig (UM) 'hair'	PAn *bukəS 'head hair'	*vukəx*

When occurring in closed syllables, the letters u and i are interpreted by Dutch speakers as the short vowels [ʏ] and [ɪ] respectively, which are both closer to schwa in their realisation than any of the other Siraya vowels. This explains why these letters often occur in alternation where they stand for a phonemic schwa.

C. The undergoer suffixes with final n exhibit considerable spelling variety in their vowels. We find a, e, i, u, and ' in this position. This vocalic chaos can somewhat be reduced:

The endings en, -in, -un and -'n all occur in patient forms and reflect the patient suffix –ən. In form and meaning, –ən corresponds closely to patient suffixes in other Austronesian languages and to Proto Austronesian *-ən. Examples:

original spelling	phonological spelling
ni-tnamsing-enhou (PST-believe-UV=2S.GEN)	
'was believed by you'	*ni-tna-m'sing-ən=hu*
kannin (eat-UV) 'to be eaten'	*kan-ən*
pa-kan-nun (CAUS-eat-UV) 'to be fed'	*pa-kan-ən*
ni-lpogh-un (xxi:35) (PST-kill-UV) 'was killed'	*ni-l'pux-ən*

---

[4] *b- > *p* is irregular.

<u>æillingigh-'noumi</u> (RDP-hear-UV=2P.GEN) 'heard by you'   *ä-ilingix-[ə]n=umi*

The ending -<u>an</u> often reflects the locative suffix *–an* but may also be the result of contraction of a root-final *-a* and the suffix *-ən*, e.g.

<u>ni-'æuloug han</u> (iii:6) (PST-baptise-UV) '(they) were baptised'   *ni-äwlux-an*
<u>ka-kytt'an-oumi</u> (RDP-look-UV=2P.GEN) 'seen by you'   *ka-kĭta-[ə]n=umi*[5]

An additional factor indicating that cases like <u>ni-'æuloug han</u> do contain a phoneme *a* in their suffix is that there are several other, allomorphic, suffixes *-an* in which *a* never alternates with other vowels. This will be discussed in Section 3.

The above cases clearly demonstrate the presence of a phonemic schwa. The decision between following the original spelling or re-establishing the underlying phonemic distinction involves two considerations. On the one hand it is clear that, although it would alter the original spelling, re-introducing schwa with a dedicated symbol would bring written Siraya more in alignment with its original phonology. Moreover, it would reduce Siraya spelling inconsistency. On the other hand, we do not always know where to put schwas. Should we introduce them without corroborating evidence from other Austronesian languages? And should we break up every strange-looking consonant-cluster with a schwa? A possible way out of this would be a compromise, introducing a schwa wherever we have comparative-historical linguistic evidence of its existence in Siraya, and not interfering in other cases.

3. THE VOICE SYSTEM: HOW MANY NON-AGENT VOICES? Proto Austronesian, and today many languages in and around the Philippines and Taiwan, have a voice system combining an agent voice with several non-agent voices. They typically maintain several of the original three non-agent voices inherited from the Proto Austronesian voice system, which, in simplified form, had the following markers (differentiated for indicative and subjunctive):

1. Proto Austronesian (simplified)

	Indicative	Subjunctive
AV	*\<m>	*\<m> + -a
UV	*-ən	*-aw
LV	*-an	*-ay
IV	*Si-	*-anay

It is clear that Siraya has lost the original Proto Austronesian instrument voice. According to Tsuchida (2000), Siraya maintained actor -, object - and locative voice of the

---

[5] Note that while *kĭta-n* occurs 75 times, there is also one instance of *kĭta-ən-hu* testifying of the vowel contraction that has taken place in all other instances.

original Proto Austronesian system. As far as the matching non-indicative markers are concerned, he does not qualify them as subjunctive and he distinguishes imperative and future markers:

	Indicative	Imperative	Future
AV	{m}-	{m}-+-a/ä	{m}-+-ah/-äh
UV	-en	-au/-äu	-auh/-äuh
LV	-an	-ei	-eih, -aneih

On the other hand, Adelaar (2011) only distinguishes between actor- and object-voice; whereas IV had disappeared, UV and LV were still formally distinguished but they were in the process of merging. Furthermore, there is a great overlap between imperative and subjunctive affixes: the imperative and subjunctive would not be distinguished if it were not for the presence of two dedicated imperative markers, *-i* and *-u* (which have cognates in other Austronesian languages):

	Indicative	Imperative	Subjunctive
AV	{m}-	{m}-+-a	{m}-+-a
UV	zero, -ən, -an	-ey, -aw[6], -i, -u	-ey, -aw, -aney

Tsuchida observes that UV and LV markers have an unclear distribution and there is 'no serious difference between them'. However, in Adelaar (2011), I argue that their distribution is not entirely random and that the Siraya voice system is in transition. It is undergoing restructuring whereby *-an* and *-aw* have become obsolete and *-ən* and *-ey* are being re-aligned as the only productive matching pair of non-actor voice markers, one for indicative verbs, the other for subjunctive ones. Today I would go even further and argue that in contrast to *-an* and *-ən*, *-aney* may still have maintained a separate role as a conveyer voice (CV), although the distinctions were sometimes getting blurred (see further below).

Tsuchida is right in so far that *-an*, *-ən*, *-aw*, *-ey*, *-aney* and ø sometimes alternate in unpredictable ways, as shown in the following instances.

Compare the variation between verb forms with and without *-ən* in sample sentences (1) and (2):

(1) *ni-paki-valey nein ta rawey lam ki Rarenan tĭn*
PST-seek-find 3P.GEN NOM child with DF Mother 3S.GEN
'they found the child with its Mother' (ii:11)

(2) *ăsi nein ni-paki-valey-ən ki mamang*
NEG 3P.GEN PST-seek-find-UV DF whatever
'they did not find anything' (xxvi:60)

---

[6] No instances of *-aney* with imperative meaning are found in the Gospel text.

Compare the variation between *-an* and *-ən* in sample sentences (3) and (4):

(3) *Päx-s'hŭt-a   äpak,   ătaral-ən=kaw   ki   varaw*
think-trust-SJ   son   forgive-UV=2S.NOM   DF   sin
'have faith, Son, your sins are forgiven to you' (ix:2)

(4) *ka-ilx-ən ......   tu   ătaral-an   ki   varaw   tu   Näy=ra*
v1-strong-UV   LOC   forgive-UV   DF   sin   LOC   earth=ADV
'the power...to forgive sins on earth' (ix:6)

In (5), (6) and (7), there seems to be no difference in meaning between *pää-tunun-ey*, *pää-tunun-ey* and *pää-tunun-aney*:

(5) *pää-tunun-ey   ta   Alak   ki   kaäwlung*
give-pass.on-SJ.UV   NOM   offspring   DF   human.being

tu   rima   ki   kaäwlung.
LOC   hand   DF   human.being
'the Son of Man shall be betrayed into the hands of men' (xvii:22)

(6) *pää-tunun-aw   ta   Alak   ki   kaäwlung*
give-pass.on-SJ.UV   NOM   child   DF   human.being

tu   rima   ki   <u>tama-ka-varaw</u>.
LOC   hand   DF   sinner
'the Son of Man shall be betrayed into the hands of sinners' (xxvi:45)

(7) *ni-hawat-ən   tĭn   ta   kidi   ka   ma-riang,*
PST-look.out-UV   3S.GEN   NOM   time   LK   AV1-good

ka   măno-no   ka   pää-tunun-aney   tĭn   tĭni-än.
LK   RDP-when   LK   give-pass.on-SJ.CV   3S.GEN   3S-OBL
'he looked for the right moment to deliver him' (xxvi:16)

Finally, in (8) the semantic difference between *väut-aw* and *väut-aney* is obscure, and it is not clear how both contrast syntactically with *ääw-väut-ey*:

(8) *ka   mamang   ka   väut-aw=mumi   tu   Näy,*
LK   whatever   LK   bind-SJ.UV=2P.GENLOC   earth

väut-aney-l-ato   tu   <u>tunun ki vŭlŭm</u>
bind-SJ.CV-LIGATURE-PRF   LOC   heaven

Ra	*ka*	*mamang*	*ka*	*ääw-va-väut-ey-mhu*	*tu*	*Näy,*
ADV	LK	whatever	LK	AS(undo)-RDP-bind-SJ.UV-2S.GEN	LOC	earth

ääw-va-väut-ey	*tu*	*tunun ki vùlù-vùlùm*.
AS(undo)-RDP-bind-SJ.UV	LOC	heaven

'whatever you will bind on earth, will be bound in Heaven; and whatever you will loosen on earth will be loosened in Heaven' (xvi:19)

However, the randomness is not pervasive. As far as the variation between *-an* and *-ən* is concerned, this only happens when these suffixes are voice markers. There are also many nominal *-an* suffixes, which are homonymous with the voice suffix *-an*, and these rarely alternate with *-ən*:

• The collective suffix *-an* occurring with nouns never alternates with *-ən* e.g. *asu-asu-an* 'dogs' (< *asu* 'dog'); *vato-to-an* 'stony places' (< *vato* 'stone).
• The oblique noun marker *-an* also invariably appears as *-an*, e.g. *ĭau-an* 'me (+oblique)'; *Isaac-an* 'Isaac (+oblique)'.
• feature nouns (that is, nouns derived from class 1 verbs) usually have *-an*, e.g. *ka-tiktik-an* 'righteousness'; *ka-patey-an* 'death'; *ka-tukul-an* 'inquity, injustice'; *kä-wäx-an, kä-wax-an* 'life'. Nevertheless, there are some counterexamples, e.g. *ka-irang-ən* 'goodness'; *kaläwhäw-an* (6x) vs *ka-läwhäw-ən* (5x) 'hell'; *ka-harum-an* (occurring 3x), *ka-harum-'n-* (2x) and *ka-harum-ən* (4x) 'mercy'.

Finally, *-an* (/-*än*) and *-ən* are also not in free variation if they constitute the final syllable of a lexical root, as in *k<m>an* 'to eat', *ĭmiän* 'we (exclusive)', or *than* 'profit'. The alternation between *-an* and *-en* is also hardly ever seen in the numeral *kĭtiän*, which occurs 53 times with *-än*, and only once with *-ən*.

The *-an* (/-*än*) and *-ən* alternation is clearly not caused by a loss of vowel contrast between *a* and schwa in general or even in last syllables: it is basically limited to verbal derivations, and most typically to non-actor voice markers.

The loss of contrast between *-ən*, *-an* and *-aney* is semantic. It is a new development in the Siraya morphosyntax, distinguishing it from Proto Austronesian. While individual non-actor verbs usually select the same set of non-actor suffixes (as can be expected), they tend to favor a combination of *-ən* and *-ay* in this function, whereas *-an* and *-aw* are becoming unproductive. In terms of frequency, combinations can be listed as follows:

1. *-ən* and *-ay*: this is the most frequent combination, occurring among others with *ma-susu* 'to say', *pa-nanang* 'to give a name', *vana-vana* 'to tell', *ĭnang* 'to refuse', *tna-m'sing* 'to believe';
2. *-ən* and *-aw*: a few frequently occurring verbs combine these, including *kan* 'to eat', *ĭt* 'to drink', *kalang* 'to know', *ma-i-alak* 'to give birth' and *pa-darang* 'to send (someone) away'
3. *-an* and *-ey*: *äwlux* 'to baptise' is the only verb which is always compatible with *-an* and *-ey*.

4. *-an* and *-aw*: There are no verbs with this combination (except for *xĭlingix* 'to hear', which combines with any of the four suffixes although its most usual UV form is *ĭlingix*-ən).[7]

The instability shown between *-an* and *-ən* and between *-ey* and *-aw*, as well as the fact that there are many verbs regularly combining (indicative) *-ən* with (subjunctive) *-ey*, indicate that there was a re-alignment in progress. Assuming that this pairing of *-ən* and *-ey* was really gaining ground, it clearly demonstrates that the Siraya voice system was not a simple continuation of the Proto Austronesian one.

As indicated above, *-aney* marks CV. It almost invariably expresses a directional movement or an act of receiving, including receiving presents, knowledge and infliction. In the Gospel text, these semantic roles transpire more clearly than those inherited from Proto Austronesian that are marked with *-an* and *-ən*. The CV only appears in subjunctive mood. Examples:

(9) *muhul-aney=kaw*   *tu*   *ka-läwhäw-an*
     cast-SJ.CV=2S.NOM   in   v1-deep-NOM
     'you will be cast down unto hell' (xi23)

(10) *Ka-vahĭr-aney=kamu ki ĭmid ki tau   aley   ki   Nanang=au*
     v1-hate-SJ.CV=2P.NOM DF all   DF person reason DF   name=1S.GEN
     'You will be hated by all for My Name's sake' (x:22)

(11) *Mang   ta   p'hä-aney-mumi   ìau-an?*
     what   NOM   give-SJ.CV=2P.GEN   1ST-OBL
     'What would you like to give me?'(xxvi:15)

(12) *ti ma-mang ta   r<m>au-a   ĭau-an tu ämäx ki   kaäwlung,*
     whoever   NOM   <AV3>deny-SJ   1S-OBL LOC front DF   human.being

     teni   ta   rau-aney=mau   tu   ämäx   ki   Raraman=au
     3S   NOM   deny-SJ.CV=1S.GEN in   front DF   Father=my
     'Whoever denies me in front of the people, him I will deny in front of my Father' (x:33)

(13) *U-ku'-aw='to   pa-tä-täutäux-aney,   kamang ta   ăta,*
     MOT-be.at/move-SJ.UV=PRF V4-RDP-teach-SJ.CV   what   NOM   PROX

     Ka-harŭm-an   ta   ka-muy-ən=au,   ăsi   pa-p'xik-an.
     v1-have.mercy-UV NOM   v1-desire-UV=1S.GEN   NEG   CAUS-sacrify-UV
     'But go and learn what this means, "I desire mercy, and not sacrifice."' (ix:13)

---

[7] The occurrences are as follows: 21 x with *-ən*, 3x with *-ey*, 3x with *-aw*, 1x with *-an*.

In Adelaar (2011), I did not acknowledge the existence of a CV. The Siraya voice system presented in this source should be adjusted as follows:

	Indicative	Imperative	Subjunctive
AV	{*m*}-	{*m*}-+-*a*	{*m*}-+-*a*
UV	zero, -*ən*, -*an*	-*ey*, -*aw*[8], -*i*, -*u*	-*ey*, -*aw*
CV			-*aney*

Compared to Proto Austronesian, Siraya verbal syntax shows the following changes:
1. The Proto Austronesian voice system with multiple undergoer-oriented voices has partly broken down. The difference between undergoer voice and locative voice was basically lost. The distinction of a third non-actor voice was maintained as a conveyer voice but it was marked in subjunctive mood only.
2. A new alignment between -*ən* and -*ey* is evolving.
3. The old -*ən*/-*aw* and *an*/-*ey* alignments still exist but are becoming obsolete, and forms like *äwlux-an* and (less so) *kan-aw* are anomalies. In the few cases where the voice suffixes -*an* and -*aw* are still used they have almost become lexicalised: the formal oppositions with -*ən* and -*ey* are still there, but they have basically lost their meaning.

Schematically, this historical process looks as follows:

PAn			-->	Siraya in origin		-->	Siraya as it tended to evolve	
	IND	SJ		IND	SJ		IND	SJ
AV	-ø	-a		AV	-ø	-a		AV	-ø	-a
UV	-ən	-aw		UV	-ən	(-aw)		UV	-ən	-ey
LV	-an	-ay		LV	(-an)	-ey	
IV	Si-	-anay		CV	-ø	-aney		CV	-ø	-aney

IS -*AN* HIGHLIGHTING A NON-CORE ARGUMENT? Some verbs with -*an* or (subjunctive) -*aney* seem to have the function of highlighting a location or another non-core argument. When this highlighting occurs, the patient is still subject. Examples:

(14) *ni-padipax-an   nein   tu   vlung   tĭn*
      PST-spit-UV   3P-GEN   in   face   3S.GEN
      they spat him in the face' (xxvi:67)

(15) *aley   ka   ni-ärax-an   ki   vato=ra*
      reason   LK   PST-found-UV   DF   rock=ADV
      'because [the house] was founded upon a rock' (vii:25)

---

[8] No instances of √-*aney* with imperative meaning are found in the Gospel text.

(16)   *ru*   *i-rua*   *paki-valey-ən*   *tĭn*   *ta*   *kuko*   *ka* *ni-uap-an*
        when   LOC-arrive seek-find-UV   3S-OBL   NOM   empty   LK PST-sweep-UV

        ki   *uap*   *tmura*   *pa-ka-tabung*   *ta*   *ăna*
        DF   broom   also   CAUS-V1-beautiful   NOM   DIST

'when he came he found it [=the house] empty, swept with a broom and decorated' (xii44)

In sentence (14), the fact that Jesus is spat *in the face* adds considerable drama to the act. In (15), a house is much stronger if built *on a rock*. In these cases, it seems that suffixation of *-an* has the effect of foregrounding the location where the act takes place. Sentence (16) is different in that here it is not the location but the instrument (a broom), which is highlighted.

A problem with this interpretation is that the Siraya text corpus is small and there are not many instances with *-aney* to verify it with. Moreover, sentence (16) is somewhat out of line with the other sample sentences because it foregrounds an instrument rather than a location, which makes the description ambiguous.

For revivalists of Siraya, the problems involving voice marking will pose several dilemmas. In the case of the variation between *-an* and *-ən*, can the original morphological opposition be maintained? Or would it be better to run ahead of the merger that must have been going on and ignore the opposition, especially since it had become unproductive and was phonologically inadequately expressed? In the case of *-aney* one may be on firmer ground as there seems to be a clear correlation between its form and function. But then again, the use of this suffix is confined to subjunctive mood. It is therefore not particularly salient. The case of *-an* as a possible highlighter of non-core arguments is even more precarious: should this function be acknowledged on the basis of the handful of sentences that are included in the Siraya corpus? Or can it best be ignored, seeing that it is weakly attested and of marginal importance in the overall verbal system? Given the above morphological picture, my personal advice would be to create a voice system in which *-ən* is maintained as the default indicative undergoer suffix, *-ey* as the default subjunctive undergoer suffix, and *-aney* as the subjunctive recipient suffix. The voice markers *-an* and *-aw* can be used optionally in those cases where they also occur in the original texts, but they should not be enforced rigorously. However, this is a matter for the burgeoning speech community to decide upon. It depends on how they want to strike a balance between authenticity and practicability.

4. THE LEXICAL SOURCES REPRESENT SEVERAL DIALECTS. If a dormant language is to be revived, one would like to use as much of its original vocabulary as possible. But, as is not uncommon with dormant languages, the 17th century Siraya materials represent two dialects.

In such a case one could combine the vocabularies of these varieties without any adjustments, but such rudimentary dialect mixing is in fact more problematic than it seems and can easily lead to undesirable outcomes.

There are two more alternatives: either the vocabulary from one dialect is adapted to the phonological developments of the other, or both dialects are compared from a historical

perspective and the etymological forms of their joint vocabularies are used. I will discuss some of these possibilities further below, after I have given an overview of the most important phonological differences between the G dialect and the UM dialect.

In section 4.2, I will discuss a fourth option: revitalizing both dialects.

4.1 SOME PHONEMIC DIFFERENCES BETWEEN THE G DIALECT AND UM DIALECT.

Proto Siraya *ð >   G *d/r*
                         UM *s*

In many cases, the UM dialect has *s* corresponding to *d* or *r* in the G dialect. These *d* and *r* sounds are basically the same phoneme: *d-* appears word-initially and in consonant clusters, and *r* appears elsewhere, although there are various cases where *r* also appears word-initially. Compare:

Utrecht Manuscript	Gospel	proto Siraya
sosoa 'two'	*ruha*, *ru-ruha*, *du-ruha* 'id.'	*ðu-ðuha 'id.'
salom 'water'	*ralum*, *-d'lŭm* 'id.'	*ðalum 'id.'
x<m>isong 'to labour'	*m-irŭng* 'id.'	*x<m>iðung 'id.'
masimdim 'to be dark'	*ma-rimdim* 'id.'	*ma-ðəmðəm 'id.'
	na da-rimdim 'thoughts'	*ða-ðəmðəm 'id.'

Proto Siraya *x >   G ø- -ø-, -x
                         UM *x* (in all positions)

Proto Siraya had a (voiced) velar fricative *x[9], which was maintained in all positions in the UM dialect but was lost in initial and intervocalic position in the G dialect. Examples:

Utrecht Manuscript	Gospel	proto Siraya
vaxo 'new'	*vahäw* id.	*vaxo 'id.'
waxi 'day; sun'	*wäi* 'id.'	*waxi 'id.'
xmisong 'to labour'	*m-irŭng* 'id.'	*x<m>iðung 'id.'
xaap 'seed'	*ääp, ähäp* 'id.'	*xaap 'id.'

---

[9] The choice of *x* as a symbol for what must have been a voiced velar fricative is admittedly inaccurate from a phonetic perspective. However, it is more practical than a dedicated IPA symbol for the purposes of language revitalization and the development of a standard spelling.

Proto Siraya *a, *u   >   UM *a*, *u*

                                        G *ä*, *äw*

Note in the G dialect that the presence of word-final *x* as well as the presence of a historical *x* in all positions entails palatalisation of some adjacent vowels: usually, *a* becomes *ä*, and *u* becomes *äw*. These vowels remain unaffected in the UM dialect. Palatalisation also happens across morpheme boundaries. It is not always predictable[10] but it is most likely to occur in three environments:

1) *a* or *u* are in the vicinity of *x*. Examples:

pasănax (UM) 'tree'	*päränäx* id.
pulax (UM) 'barren country, desert'	*puläx* id.
xiltax (UM) 'thunder'	*'ltäx* 'thunder'
ma-xanix (UM) 'it's good'	*mä-änix* 'beautiful'
rĭx 'mind' + *-uhu* '2s.GEN'	*rĭx-äwhu* 'your mind'
pa-ka-rämäx 'let shine' + *-aw* '(SJ, UV)'	*pa-ka-rämäx-äw* 'let shine!'

2) *a* or *u* are in the vicinity of a palatalised *ä*. Examples:

ma- (stative prefix) + *-uax-*, *-uäx-* 'living'	-->	*mä-uäx* 'to live'
pihä (xiv:7) 'to give' + *-a* (+ SJ)		*piä-ä* 'give!'
pasănax (UM) 'tree'		*päränäx* id.
pa-ka-rämäx 'let shine' + *-aw* '(SJ, UV)'		*pa-ka-rämäx-äw* 'let shine!'
puläx 'barren country, desert'		*päwlä-päwläx* id.
m-u- (AV-MOT-) + *pänäx* 'outward'		*m-äw-pänäx*, *m-u-pänäx*
		'to appear in the open'

3) *a* or *u* were historically in the vicinity of *x*, which is still extant in UM. The effect of this palatalisation is fairly regular; note however that the combination of *x* loss and vowel palatalisation may sometimes entail a very different word shape.

UM	Gospel
xamax (<u>gamach</u>) 'blood'	*ämax* id.
xiltax 'thunder'	*'ltäx* id.
xapit 'get up'	*äpit* id.
raxang 'rib'	*räang* id.
nixaha 'sister'	*niähä* id.

---

[10] However, historically palatalization of *a* and *u* is predictable in that it occurs in the vicinity of a Proto Siraya *x descending from PAn *R (a velar fricative); in cases where Proto Siraya *x reflects PAn *S (a sibilant), no palatalization takes place (Adelaar. In press).

maxanix 'it's good'	*mä-änix* 'beautiful'
waxi 'day; sun'	*wäi* id.
vaxiox 'stormwind'	*bäyux* 'thunderstorm'
vaxo (<u>vacho</u>) 'new'	*vahäw* id.
tataxof 'blanket'	*ta-taäwf* 'to reveal'
taxo 'shoulder'	*pa-ka-taäw-ən* 'laid on the shoulders'
xuma 'village, town'	*äwma* id.
vuxox 'enemy'	*väux* id.
uxla 'snow'	*äwla* id.
uxing 'candle'	*äwing* id.
muxo (<u>moucho</u>) 'whale'	*mäwäw* id.
muxax (<u>mougag</u>) 'to live'	*m-äwäx* id.

As mentioned at the beginning of this section, there are three ways to solve the circumstance that the lexical data belong to more than one dialect. One way is by combining the lexical sources without further ado (and cause dialect mixing); a second possibility is to combine the sources but adapt the lexicon of one dialect to the phonology of the other through reconstruction (using the comparative method). A third solution is to combine the sources but adapt the words of both dialects to the phonology of their common stock language.

Combining all available vocabulary without adaptation of some of the words will lead to a dialect mix, which is not particularly representative for the language once spoken. In some cases the result will even be unpractical and messy, for instance if there is much overlap between the vocabularies of several dialects and the revived language ends up with many pairs of matching words which somehow look similar and have meanings that are identical, or almost identical. Finally, a crude form of dialect mixing will sometimes lead to different dialects contributing words that have different morphology but share the same root. This would blur the lexical relationship between these words even further. To avoid such problems, phonological adaptation to one dialect only or to Proto Siraya seems the most reasonable option. If one choses for adapting all words to the phonology of one dialect only, the UM dialect would seem the most appropriate because its phonology and word structure are more regular.

Compare the following word pairs, where information from the G dialect is used to derive a matching form which has different morphology in the UM dialect and is marked with two preceding asterisks. In all these cases, phonological information in both dialects is also sufficient for the reconstruction of a corresponding Proto Siraya etymon (marked with one preceding asterisk):

UM	**Gospel**	**Proto Siraya**
xmisong 'to labour'	*m-irŭng* 'id.'	*x<m>iðung 'id.'
**tama-isung 'worker'	*tama-irŭng* 'worker'	*tama-iðung 'id.'
ma-sarey 'wicked, angry'	---	*ma-ðarey 'id.'

**sarey 'anger'	*rarey* 'anger'	*ðarey 'id.'
**sarey-rey-ing 'lamentation'	*rarey-rey-ing* 'lamentation'	*ðarey-rey-ing 'id.'
**ka-sarey-ang 'scolding, dispute'	*ka-rarey-an* 'scolding, dispute'	*ka-ðarey-an 'id.'
x\<m\>a-xirax 'to sell, trade'	---	*x\<m\>a-xirax 'id.'
**xa-xirax-ang 'market place'	*a-ĭräx-an* 'market place'	*xa-xirax-an 'id.'
**ma-xangal 'worthy, valuable'	*m-angal* 'worthy, valuable'	*ma-xangal 'id.'
waxi ki ka-xa-xangal-ang 'festive day'	---	*ka-xa-xangal-an 'festivity'
x\<m\>imix 'to value, appraise'	---	*x\<m\>imix
**tama-ximəx 'tax collector'	*tama-imix*, *tama-imŭx* 'tax collector'	*tama + *ximəx 'id.'

4.2 THE OPTION TO REVITALIZE SEVERAL REGIONAL DIALECTS. Theoretically there is also the option to revitalize more than one Siraya regional dialect. Similar projects are under way for Kāi Tahu Maori, the dialect once spoken on Kāi Tahu (New Zealand's South Island), and Ume Saami in Sweden. The revival of Kāi Tahu consists mainly in changes in names and translations of some place names and institutions; it is apparently reinforced by regional individualism (Tahu Potiki Online). The revival of Ume Saami is motivated by the fact that already standardized versions of Saami languages (including Ume Saami itself) are unintelligible to Ume Saami speakers (Larsson unpublished).

However, in the case of Siraya, such conditions do not apply[11], and there is little cause for regional dialect revival. To begin with, there is no popular demand for it among Siray-ans. Furthermore, in contrast to the Kāi Tahu and Ume Saami cases, there is no standard dialect which is derived from one particular spoken regional variety and risks to over-shadow other regional varieties that are spoken or at least remembered. The notion that there are regional dialects is a scholarly one, and it is not alive among those who identify as Sirayans. Finally, as far as the 17th century Siraya data is concerned, it is linguistically very difficult to link up the G dialect and UM dialect with particular Siraya regions because the evidence is minimal and ambivalent (compare Li 2010). From a sheer practical linguistic point of view one may add that if it comes to the potential to revitalize a fully functioning language, only the G dialect stands a chance, as this is the only dialect provided with suffi-cient grammatical and lexical information. For the UM dialect there only is a long wordlist and some very short and simple dialogues. For other dialects (documented in Ogawa's

---

[11] Note however that as far as language engineering is concerned, the Ume Saami case is not unlike that of Siraya. In his attempt to revitalize this language, Larsson makes a careful evaluation of how the various individual Ume Saami dialects and sources can be used for standardization. East Ume Saami still has a speech community, whereas West Ume Saami is almost dormant but has a much richer lexicon on record. Larrson proposes to take the phonology, morphosyntax and limited vocabulary of East Ume Saami as a basis and to combine it with the lexicon of West Ume Saami after adapting it to the phonology of East Ume Saami (Larsson 2013:14–15).

wordlists) the data is even less, both in terms of quantity and quality. Strictly speaking, a small corpus may be no reason to refrain from revitalization: however, the present case is a rather academic one in the absence of any demand for it.

For the sake of completeness I also mention an effort to revitalize the Siraya language through the expertise of shamans. The endeavour is based on the assumption that "shamans, as exponents of their traditional culture, might have some vestigial language knowledge", and in trance they may be able to transfer linguistic knowledge from their ancestral god Alizou (Fang & Nathan 2013). Fang & Nathan point out that the notion to elicit language through trance raises interesting questions about revitalization methodology, community participation and the acquisition of community approval (2013). However, no analysis of the data thus obtained has yet been published, and its linguistic status remains unclear. It remains to be investigated whether the elicited utterances represent a Siraya variety or a register, and whether they are still adequate for linguistic interpretation in the first place.

5. MAINTAINING THE GOSPEL DIALECT'S 'ANTICIPATING SEQUENCES': ENRICH-MENT OR BURDEN? Complex verb constructions involve an auxiliary and a following infinitive verb. In these constructions, the G dialect usually prefixes a formal part (the first one or two syllables, or the initial consonant) of the infinitive verb to the auxiliary. Sometimes the prefixed element is the initial part of another verb with the same meaning, or it is an element that has an iconic meaning corresponding to that of the infinitive verb. There is some variability in the formal parts that are prefixed. I call these elements 'anticipating sequences'. Note that they are absent in the UM dialect data, and that complex verb constructions are therefore more simple in the UM dialect. Examples:

The anticipating sequence is a formal element of the lexical verb, such as *kmi-* (< *k<m>ĭta*) or *mu-* (< *m-umxa*):

(17)   *Raraman-uhu*   *ka*   **kmi**-*dung*   **k<m>ĭta**
        Father-2sGEN   LNK   AS-do in secret   <AV>look,watch
        'your Father Who sees in secret…' (vi:18)

(18)   **mu**-*ìməd-kamu*   *kawa*   **m-umxa**   *ki*   *àta*
        AS-all-2sNOM   maybe   AV-understand   DF   this
        'do you understand all this?' (xiii:51)

The anticipating sequence is semantically related to the lexical verb (*pää-* is also a bound verb meaning 'to give' or 'to pass on'):

(19)   **pää**-*ìməd-ey-(m)au-kaw*   *p-u-daäux*
        AS-all-ᴜᴠ.SJ-1sGEN-2sNOM   CAUS-MOT-pay
        'I will pay you everything' (xviii:26)

The anticipating sequence is iconically related to the lexical verb (assuming that *ää* is an onomatopoeia referring to basic reading rehearsals in a literacy program)[12]:

(20)   ***ää**-la-likux*   ***s<m>ulat***
        AS-RDP-do again   <AV>to write
        "It is written again."

Anticipating sequences do not always appear where they are expected: *k-* (< *k<m>an*) is prefixed to *ĭməd* and to *k'da* in (21) but not to *ruruk* in (22).

(21)   *Ni-k-ĭməd*   *k<m>an*   *ka*   *mi-bangtaw*   *ta*   *neni,*
        PST- AS(eat)-all   <AV3>eat   LK   INCH-satiated   NOM   3P

        ka   *ni-ara*   *nein*   *ki na*   *pipi*   *ka ni-k'da*   *k<m>an*
        LK   PST-take   3P.GEN   DF PART   crumb   LK PST-AS(eat)leave   <AV3>eat
        'They all ate and were satisfied, and they took up the fragments that were left' (xiv:20)

(22)   *Ni-irua*   *ta*   *ay-ayam*   *ka*   *subŭx,*
        PST-arrive   NOM   RDP-animal   LK   sky

        ka   *ni-r<m>uruk*   *k<m>an*   *ki*   *ăna.*
        LK   PST-<AV3>devour   <AV3>eat   DF   DIST
        'The birds came and devoured them' (xiii:4)

Likewise, *m-ăta-* (< *ătaral*) is prefixed to *ra-rĭx* in (23) but not to *dĭs* in (24):

(23)   *ru*   *ăsi-kamu*   *m-ăta-ra-rĭx*   *ta*   *mäix*   *bua-vual*
        if,when   not-2P.NOM   AV1-AS(forgive)-RDP-neck   NOM   each   RDP-body

        ătaral   *ki*   *täiapara-umi*   *ki*   *varaw*   *nein.*
        forgive,abandon   DF   brother-2P.GEN   DF   sin   3P.GEN
        'If not each of you from the depth of your heart forgive the sins of your brother'
        [lit. forgive your brother concerning his sins] (xviii:35)

(24)   *du*   *ma-dĭs*   *ătaral*   *ki*   *ra-reyhul,*
        if,when   AV1-quick   forgive,abandon   DF   RDP-net
        'While quickly leaving their nets behind...' (iv:20)

    One might ask whether this rather remarkable grammatical phenomenon is worth maintaining, given that it is only represented in the Gospel dialect, it does not always occur,

---

[12] The Dutch missionaries in West Taiwan established large literacy programs for the local population.

and it is very likely to cause difficulties to Siraya learners. Three approaches are conceivable. First, speakers maintain anticipating sequences wherever they are found in the Gospel dialect. In that case, they will have to deal with various anticipating sequences that are unclear in meaning and valency. Second, speakers dispense with anticipating sequences: this would in fact be emulating the UM dialect. Finally, speakers maintain a reduced number of easily recognisable anticipating sequences that convey a clear semantic notion, and use them systematically in well-prescribed situations. This third approach is attractive in that it will preserve a special feature unique to West-Formosan languages while using a system that is more user-friendly than the one we find in the textual material inherited from 17th century Dutch missionaries.

6 CONCLUDING REMARKS. The topics dealt with in the previous sections clearly show that some Siraya data are inherently messy and need engineering before they can be used for revitalization purposes.

Recognizing schwa as a phoneme in its own right will restore an inherent feature of Siraya phonology. Moreover, it will also reduce the spelling inconsistency in Siraya texts and bring out the phonotactic structure of Siraya.

The roles of voice suffixes lack transparency in Siraya texts. Formally, *-ən* and *-an* reflect Proto Austronesian undergoer *-ən* and locative *-an*. However, by interpreting these suffixes as default undergoer and locative voice markers (which they are in many other Formosan languages), we would ignore the clear Siraya tendency to merge undergoer and locative voice. To adopt this merging tendency seems a more efficient option than to re-establish a distinction that was historically real but had already become unproductive in the 17th century.

Combining vocabularies from different source dialects without further adaptation is awkward and may create an unrepresentative phonological mix. It also carries the risk of creating unnecessary doublets and obscuring lexical and morphological connections. One way to avoid this is by adjusting the vocabulary of one dialect (presumably the G dialect) to the phonology of the other (the UM dialect). Another - probably more complicated - solution would be to adjust the vocabularies of the several dialects to the phonology of their common stock language (Proto Siraya).

Anticipating sequences are typical for some Formosan languages in Southwest and Central Taiwan. In Siraya, they only occur in the Gospel dialect, and they do not seem to be very rooted in the grammar. Their formal connections and meanings are not always obvious. One could ignore them and follow the grammar of the Utrecht Manuscript, which does not have these sequences. However, this would constitute the loss of a rather emblematic feature of the G dialect, which in other respects is the most important source we have for the structure of Siraya in general. Rather than eliminating anticipating sequences from the grammar, one could also rationalize the system by which they are governed by only maintaining those ones that are productive and convey a clear meaning, and to apply them consistently in clearly defined situations.

It is of course up to the language revivalists in the Siraya community how they will resolve the above issues. How have they dealt with these issues so far? In some cases their approach has been to simplify the grammar, but they did this less by inclination than per force as it was due to a lack of adequate descriptions at their disposal. In several discus-

sions they have voiced their preference to maintain the authentic flavour of Siraya wherever possible in view of the practical constraints they face in teaching the language. They use an open tool for modification approach and consider as they go solutions for structural difficulties such as how to apply the voice affixes and anticipating sequences. In the case of schwa, the Siraya revivalists use a conservative orthography, and they do not represent this phoneme. However, they are keenly aware of its existence as well as of other orthographic inconsistencies. They are discussing the creation of an adapted orthography which would also indicate this phoneme. This orthography would no doubt affect the entire lexicon. It could solve some phonological problems arising when the G dialect is enriched with vocabulary belonging to other dialect sources.

References

Adelaar, Alexander. 2011. *Siraya: Retrieving the phonology, grammar and lexicon of a dormant Formosan language*. Berlin: De Gruyter Mouton.

Adelaar, Alexander. In press. Proto Siraya phonology: A reconstruction. *Cahiers de Linguistique Asie Orientale in Paris*.

Czaykowska-Higgins, Ewa. 2009. Research models, community engagement and linguistic fieldwork: Reflections on working with Canadian indigenous communities. *Language Documentation & Conservation* 3(1). 15–50. http://hdl.handle.net/10125/4423.

Fang, Meili & David Nathan. 2013. Divining Siraya: Sources of language and authority in documentation and revitalisation. Paper presented at the International Workshop on "Special genres" in and around Indonesia, Linguistic Dynamics Science Project. Tokyo University of Foreign Studies, Tokyo. 17–19 February, 2013.

Gerdts, Donna B. 2010. Beyond expertise. The role of the linguist in language revitalization programs. In Lenore A. Grenoble & N. Louanne Furbee (eds.), *Language documentation: Practice and values*, 173–192. Philadelphia, PA: John Benjamins.

Gravius, Daniel. 1661. *Het heylige Euangelium Matthei en Johannis. Ofte Hagnau ka d'llig matiktik ka na sasoulat ti Mattheus ti Johannes appa*. Amsterdam: Michiel Hartogh.

Gravius, Daniel. 1662. *Patar ki tna-'msing-an ki Christang ofte 't Formulier des Christendoms*. Amsterdam: Michiel Hartogh.

Li, Paul Jen-kuei. 2001. Formosan languages: The state of the art. In David Blundell (ed.), *Austronesian Taiwan: Linguistics, history, ethnology, and prehistory*, 45–67. Berkeley, CA: Phoebe A. Hearst Museum of Anthropology, University of California, Berkeley; Taipei: Shung Ye Museum of Formosan Aborigines.

Li, Paul Jen-kuei. 2010. *Studies of Sinkang manuscripts. Languages and Linguistics Monograph Series 39*. Taipei: Institute of Linguistics, Academia Sinica.

Huang, Jimmy. 2010. Language revitalization and identity politics: An examination of Siraya reclamation in Taiwan. University of Florida dissertation.

Huang, Chun (Jimmy), Edgar Macapili & Uma Talavan. 2013. Heritage linguistics and language activism: A conversation with the Siraya. Paper presented at the 3rd International Conference on Language Documentation and Conservation (ICLDC3). The University at Hawai'i at Mānoa, Honolulu, HI. 28 February–3 March, 2013. http://hdl.

handle.net/10125/26136.

Larsson, Lars-Gunnar. 2013. Language variation in the revitalization process. Paper presented at Australex 2013: Endangered Words, and Signs of Revival. The University of Adelaide, Australia, Adelaide City Centre. 25–28 July, 2013.

Macapili, Edgar L. 2008. *Siraya glossary. Based on the Gospel of St. Matthew in Formosan (Sinkan dialect). A preliminary study.* Tainan: Tainan Pe-po Siraya Cultural Association.

Ogawa, Naoyoshi. 2006 [original manuscript 1917]. *A comparative vocabulary of Formosan languages and dialects.* Edited by Paul Jen-kuei Li & Masayuki Toyoshima. Asian-African Lexicon Series No.49. Tokyo: Tokyo University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa.

Rice, Keren. 2010. The linguist's responsibilities to the community of speakers: Community-based response. In Lenore A. Grenoble & N. Louanne Furbee (eds.), *Language documentation: Practice and values*, 25–36. Philadelphia, PA: John Benjamins.

Potiki, Tahu Leslie Karetai Kingi. n.d. Otakou representative. http://otakourep.ehclients.com. (18 September, 2013.)

Tsuchida, Shigeru. 2000. Lexical prefixes and prefix harmony in Siraya. In Videa P. de Guzman & Byron W. Bender (eds.), *Grammatical analysis: Morphology, syntax and semantics: Studies in honor of Stanley Starosta*, 109–128. Honolulu, HI: University of Hawai'i Press.

Tsuchida, Shigeru. 2001. Japanese contribution to the linguistic studies of the Formosan indigenous languages. In David Blundell (ed.), *Austronesian Taiwan: Linguistics, history, ethnology, and prehistory*, 68–94. Berkeley, CA: Phoebe A. Hearst Museum of Anthropology, University of California, Berkeley; Taipei: Shung Ye Museum of Formosan Aborigines.

Tsuchida, Shigeru & Yukihiro Yamada. 1991. Ogawa's Siraya/Makatao/Taivoan Comparative Vocabulary. In Shigeru Tsuchida, Yukihiro Yamada & Tsunekazu Moriguchi (eds.), *Linguistic materials of the Formosan Sinicized Populations I: Siraya and Basai*, 1–194. Tokyo: The University of Tokyo, Department of Linguistics.

Van der Vlis, Christianus Jacobus. 1842. Formosaansche woorden-lijst, volgens een Utrechts Handschrift [Word list of Formosan according to a manuscript from the archives of Utrecht]. *Verhandelingen van het Bataviaasch Genootschap* 18. 437–488.

Alexander Adelaar
karlaa@unimelb.edu.au

CONVENTIONS AND ABBREVIATIONS USED IN THIS PAPER

Siraya data in italics are in near-phonemic spelling (cf. Adelaar 2011); Siraya data that are underlined represent the original orthography.

List of abbreviations:

1s	first person singular
1PE	first person plural exclusive
1PI	first person plural inclusive
2s	second person singular
2P	second person plural
3s	third person singular
3P	third person singular
ADD	additive ('and', 'also')
ADV	adversative ('but', 'however')
AV	actor voive
AV(1/3/4)	AV affixes resp. belonging to class 1,3,4 verbs
AS	anticipating sequence
CAUS	causative
COM	comitative
CV	conveyer voice
DF	default case marker
DST	distal
GEN	genitive
id.	= 'having the same meaning as previous gloss'
IMP	imperative
INCH	inchoative
INCL	inclusive
INDEP	independent
IND	indicative
INV	inversive
IV	instrument voice
LK	linker
LOC	locative
LV	locative voice
MOT	motion
NEG	negator
NOM	nominative
OBL	oblique
PA	personal article
PAn	Proto Austronesian
PRF	perfective
PRX	proximal
PST	past

RDP	reduplication
ST	stative
SJ	subjunctive
SJ.UV	portemanteau suffix combining SJ and UV
UV	undergoer voice
V(1/4)	prefixes forming resp. class 1 and 4 verbs