

# Word Lists and Dictionaries of Formosan Languages

*Paul Jen-kuei Li, Christopher Joby and Elizabeth Zeitoun*

5.1	Introduction	111
5.2	The 17th Century: Dutch and Spanish Word Lists	112
5.2.1	<i>Spanish Word Lists</i>	113
5.2.2	<i>Dutch Word Lists</i>	113
	5.2.2.1 Word Lists and Dictionaries of Favorlang	114
	5.2.2.2 Word Lists of Siraya	115
5.3	Late Qing and Japanese Periods	116
5.3.1	<i>Word Lists in the Qing Period</i>	116
5.3.2	<i>Reediting of Previously Published Word Lists</i>	117
5.3.3	<i>Word Lists and Dictionaries Compiled during the Japanese Occupation</i>	117
5.4	Word Lists and Dictionaries Compiled after 1945	118
5.4.1	<i>Word Lists</i>	119
	5.4.1.1 Individual Languages	119
	5.4.1.2 Comparative Word Lists	120
5.4.2	<i>Dictionaries</i>	121
	5.4.2.1 1970s to 1990s	122
	5.4.2.2 Early 21st Century	123
5.4.3	<i>Online Dictionaries</i>	125
	5.4.3.1 Online Dictionaries Sponsored by the Council of Indigenous Peoples	125
	5.4.3.2 Yami Online Dictionaries	126
	5.4.3.3 Austronesian Comparative Dictionary	127
5.5	Conclusion	127
	References	127

## 5.1 Introduction

Word lists and dictionaries have been compiled for Formosan languages for nearly four centuries, with the first known word lists dating back to the 1630s.

The distinction between “word lists” and “dictionaries” should be self-explanatory. According to the definitions given by the online edition of the *Merriam-Webster*, a word list is a collection of words arranged alphabetically (or thematically), while a dictionary is a reference source in print or electronic form listing words usually alphabetically and providing additional information on their form, pronunciation, function, etymology, meaning, and syntactic and idiomatic uses. In Formosan languages, the distinction between word lists and dictionaries is sometimes altered because of the perception that a compiler has about his or her work, and what certain scholars would treat as “large-scale word lists” are actually defined as “dictionaries” (see for instance Ferrell 1982 and Cauquelin 1991). Our discussion will not try to reclassify any of these original works. This chapter also does not intend to review all the resources available, nor to discuss the methods that should be adopted to edit useful lexicographic material, but, rather, to guide the reader to linguistically relevant or important work. More specifically, it aims to provide an overview of word lists and dictionaries compiled over the past 400 years during three different periods—(i) the Dutch and Spanish period in the 17th century (§ 5.2), (ii) the late Qing and Japanese period (§ 5.3), and (iii) the post-World War II period (§ 5.4)<sup>1</sup>—and explain the contexts in which they were compiled or published.

Overviews of Formosan lexicography have been written in Japanese by Tsuchida (1972) and in Chinese by P. Li (1995, 2007),<sup>2</sup> Huang et al. (2014), and T. Li (2016, pp. 55–93), and we have taken them into account in our discussion. T. Li (2016) provides a list of a hundred word lists and dictionaries of Formosan languages as of 2015, but if we were to consider both word lists designed as parts of individual studies (e.g., grammars) and those published in comparative studies, we would reach a far higher number (around 200).

## 5.2 The 17th Century: Dutch and Spanish Word Lists

During the colonization of Taiwan by the Dutch East India Company (1624–1662) and Spain (1626–1642), Christian missionaries compiled material that could aid them in understanding the indigenous languages, including word lists briefly reviewed below, which formed an essential component of their mis-

1 This division into three periods may be arbitrary, but it allows for a description within a historic and chronological framework.

2 Li's (2007) article was translated by Noah Sauvé and appeared in English in a subsequent publication (see Li 2013).

sion to convert indigenous people to Calvinist and Catholic Christianity. This was the first time that word lists were compiled in Formosan languages.

We first give a brief history of the materials produced by the Spanish (§ 5.2.1), which do not appear to have survived into the present, before turning to the Dutch word lists and dictionaries, which are still very useful for linguistic study and language revitalization (§ 5.2.2).

### 5.2.1 *Spanish Word Lists*

Spanish missionaries are known to have compiled two vocabulary lists. The first, entitled *Vocabulario de la lengua de los indios de Tanchui, en la Isla Hermosa* ‘Vocabulary of the language of the natives of Tamsui, on the island of Formosa’, was recorded by Fr. Jacinto Esquivel, who resided in Taiwan from 1631 to 1633 (Medina 1896, p. 29; Retana 1906, p. 100). Fr. Teodoro Quirós, who was in Taiwan from 1632 until 1642, wrote a grammar entitled *Arte de la lengua de Formosa* ‘Grammar of the language of Formosa’, along with a vocabulary list entitled *Vocabulario en la misma lengua* ‘Vocabulary in the same language’. It is not clear to which Formosan languages these titles referred, but it may be that *la lengua de los indios de Tanchui* and *la lengua de Formosa* are both varieties of Basay (Blust 2013, p. 50), spoken in the Spanish-controlled region of northern Taiwan (see T. Liu, this handbook, Chapter 45). No extant copies of these vocabularies have been found, but they may have been similar to the dictionaries compiled in the Spanish colony of the Philippines, such as the *Vocabulario de la lengua tagala* ‘Vocabulary of the Tagalog language’, published by the Franciscan Pedro de San Buenaventura in 1613. The origins of this model of dictionary may be traced to Antonio Nebrija’s (1492) Spanish–Latin lexicon, *Vocabulario español-latino* (Fernández Rodríguez 2014, p. 275).

### 5.2.2 *Dutch Word Lists*

Two languages were recorded during the Dutch period (1924–1662): Favorlang (later documented as Babuza), an extinct language spoken in the central-western plains of Taiwan, in the area of modern-day Yunlin County (P. Li, this handbook, Chapter 47), with two main dialects, Takkais and Ternern (both spelled in various ways), and Siraya, a dormant language spoken in southwest Taiwan, with at least three major dialects—Siraya (Proper) in coastal Tainan, Taiवान, to the north, in inland Tainan Plains, and Makatau farther south in the Kaohsiung and Pingtung areas (Adelaar, this handbook, Chapter 57).

The word lists recorded for these two languages differ in several respects, and the editions of each illustrate problems of reception. They are nevertheless valuable sources of linguistic data and are the oldest examples of lexicography known to have survived from Taiwan.

### 5.2.2.1 Word Lists and Dictionaries of Favorlang

The first Dutch missionary linguist to work on Favorlang was Simon van Breen, who was in Taiwan from 1644 until 1647 and established his base in Dalivo, in modern-day Dounan, Yunlin county, and is said to have compiled a lexicon (Campbell 1903, p. 242). While there is no trace of this material, it may well have formed the basis of the bilingual Favorlang–Dutch lexicon, with entries reflecting the two main dialects, which has been published several times. It was compiled around 1650 and was attributed to Gilbertus Happart, who resided in Taiwan from 1649 until 1653. The manuscript (Happart 1650), dated 1650, was identified in Batavia (Jakarta) in 1839 by Wolter (W.R.) van Hoëvell. In 1840, the English missionary Walter (W.H.) Medhurst published a Favorlang–English dictionary (Happart 1840), and two years later, van Hoëvell published a Favorlang–Dutch edition (Happart 1842). This had entries for 2,664 words, arranged with Favorlang first, and then Dutch, this particular order suggesting that the vocabulary was intended for the use of fellow missionaries studying the language. The range of entries in the lexicon strongly suggests the extensive involvement of Favorlang informants in the compilation of the lexicon, although their names and contribution are not mentioned in Dutch sources. In particular, the dictionary provides a lot of derived forms, some of which are illustrated with examples, though usually phrases (Happart 1840, 1842). One fundamental problem is that the current whereabouts of the original manuscript are unknown, but it seems that it contained two lists of words, a principal, longer list and a second, shorter list comprising only nouns, as shown in the editions published by Medhurst (Happart 1840) and van Hoëvell (Happart 1842).

These editions provide useful insights into the way in which Dutch missionaries went about compiling a bilingual lexicon. One feature of all surviving Favorlang material is that it is written in the Roman alphabet, but the missionaries used additional characters to represent sounds absent from the Dutch phonological system. One of these characters,  $\bar{g}$ , as in *a-eniḡ* ‘a word uttered quietly’, represents a pharyngeal ‘g’, which is absent from Dutch. Two other symbols are ‘, for the glottal stop, and  $\mathfrak{s}$ , for a voiceless sibilant (P. Li 2019, p. 9). Many Favorlang words include a hyphen. The precise function of this is unclear (ibid., p. 14, n. 18). In some cases, it functions as a morpheme boundary marker, e.g., *ma-chachod* ‘adulterer’ (< *chachod* ‘adultery’). In other cases, it may function as a syllable marker, e.g., *ma-a-chachod*, which also means ‘adulterer’, although usage is by no means consistent. Another distinctive feature of the orthography of the Favorlang words is the use of diacritics, e.g.,  $\acute{o}$ ,  $\acute{u}$ , and  $\acute{a}$ . Again, their precise function is unclear, and usage is inconsistent. For example, ‘silver money’ (Dutch *silver gelt*) is rendered in Favorlang as *so-è*, whereas ‘gold’ (Dutch *goût*)

is rendered as *so-ë o kakan* (Happart 1842, p. 377). As in the case of ‘gold’ (*goût*), diacritics are also used on several Dutch words, where they would typically be absent in other Dutch texts.

#### 5.2.2.2 Word Lists of Siraya

There are several reports of Dutch missionaries compiling vocabularies and dictionaries of Siraya, although only one of these survives. The first Dutch minister in Taiwan between 1627–1631 and 1633–1637, Georgius Candidius, compiled a vocabulary of all the native words “with which [he was] familiar” (Campbell 1903, p. 97). A letter written by the Dutch consistory at Zeelandia records that Candidius’s successor, Robertus Junius, compiled *een groot vocabulaer* ‘a large vocabulary’ (Campbell 1903, pp. 236–237, Grothe 1884–1891, IV, p. 76). No vocabulary that can be directly attributed to either Candidius or Junius is known to have survived, but one Siraya vocabulary, called the Utrecht Manuscript (henceforth UM), does. It is a 55-page manuscript consisting of a Siraya–Dutch word list and four short bilingual dialogues and is an early example of Dutch bilingual lexicography (Adelaar 2006). The manuscript is written on Chinese rice paper in quarto. It was acquired by Utrecht University professor Sebald Rau, at an auction in 1761. The auction was of manuscripts owned by Hubertus Reland, the son of another Utrecht University professor, the orientalist scholar Adriaan Reland (*latine* Hadrianus Relandus, 1676–1718). It had previously been owned by Adriaan, an avid collector of manuscripts from the Middle East and Asia (Jaski 2021, pp. 334–335, Joby 2021). It has the title *Vocabularium Formosanum* ‘Formosan vocabulary’, added later in a different ink, possibly by Adriaan Reland (Pytlowany & Van Hal 2016, p. 26). We currently do not know who compiled the UM, but it consists of 1,072 words in Siraya Proper (van der Vlis 1842, p. 443), divided into two columns, with Siraya words on the left and their Dutch equivalents on the right. These two columns have the Latin headings *Formosana* ‘Formosan (words)’ and *Belgica* ‘Dutch (words)’, respectively, which were possibly added by Reland. In most cases, each entry has one Siraya word and one Dutch word, with Dutch nouns typically given an article. The word list can be divided into two parts. The first, larger part groups words by theme such as God and nature, the house, time, and food. The second part gives verbs in alphabetical order with no attempt to provide root forms, so many of these verbs begin with the actor-voice prefix *m(a)-* or have the infix *<m>* (Joby 2021).

### 5.3 Late Qing and Japanese Periods

Blust (2013, p. 50) mentions that “following the expulsion of the Dutch from Taiwan in 1662 no other language materials were collected until the Japanese colonial occupation of 1895–1945.” This is also the position adopted by T. Li (2016, p. 55), who asserts that Formosan lexicography started with the arrival of the Japanese. While it is true that officials and scholars of the Qing Empire had little interest in the languages and cultures of the indigenous peoples, a fact that reflects their general lack of interest in the island (Price 2019, pp. 75–108), these statements need to be mitigated, as some lexical items were transcribed in Chinese characters, and word lists were also compiled by foreign visitors in Taiwan during the Late Qing period (§ 5.3.1). This section also reviews the republication of earlier word lists (§ 5.3.2) and the compilation of word lists and dictionaries during the Japanese period (§ 5.3.3).

#### 5.3.1 *Word Lists in the Qing Period*

Although they were recorded only sporadically or as part of historical descriptions of native songs (Huang 1724), lexical items were written down in Chinese during the Qing dynasty by Zhou (1717) from Siraya, L. Liu (1741) from various ethnic groups, and Zhu (1765) from Makatau. All these words have been quoted in full with transliterations and arranged alphabetically by the Japanese historian Naojirō Murakami (Murakami 1933). Naoyoshi Ogawa (1999) discusses L. Liu’s (1741) word list and compares it with various Formosan languages and a few extra-Formosan languages.

In the late Qing period, Taiwan started to open again, and after the 1860s, Western explorers started to undertake long trips around Taiwan. They not only wrote observations and published their notes as articles in bulletins and journals but also undertook linguistic fieldwork. Two word lists are noteworthy, the first by Bullock (1874) and the second by Taintor (1874). The first is a comparative list of 179 lexical items thematically ordered, including Tsui-hwan (Thao), Sekhwan (Pazeh), Bu-hwan (Seediq), Pepo-hwan (Siraya), Favorlang, and Pelam (Puyuma), with English glosses. The second is a list of about 300 lexical items in Kabaran (Kavalan) and Yukan (Atayal), divided into ten categories (numerals, human beings and relations, body parts, domestic life, time, flora, fauna, miscellaneous nouns, pronouns/adjectives/adverbs, and verbs) organized alphabetically. Along with the English glosses, comparisons are made with Malay and a number of western Austronesian languages. Taintor’s (1874) word list of Kavalan and Atayal is very useful for a comparison of different dialects at an earlier stage (see Li & Tsuchida 2006, p. 11).

### 5.3.2 *Reediting of Previously Published Word Lists*

In 1896, the Scottish missionary William Campbell published a Favorlang–English edition of the lexicon that combined the two lists mentioned in § 5.2.2.1, while attempting to rectify some of the errors in the earlier editions (Campbell 1896, pp. xviii–xix). In 2003, Paul J. Li published an English–Favorlang edition of the lexicon, based on notes compiled by the Japanese linguist Naoyoshi Ogawa, which in turn were based on the Campbell edition (Ogawa 1930a, 2003).

As mentioned above, Naojirō Murakami arranged data extracted from Chinese sources (see § 5.3.1). He also published the vocabulary of the UM and provided English translations of the entries (Murakami 1933, pp. 154–202). However, he based his edition on the van der Vlis (1842) edition rather than the original UM manuscript, and, in total, nearly one in ten (103 or 9.6%) of Murakami's entries contain an error. A third edition of the UM was published by Shigeru Tsuchida (Tsuchida 1998), and while there were attempts to correct several of the transcriptions of van der Vlis/Murakami and to fill in the gaps in Murakami's English translations, Tsuchida did so without reference to the UM, thus leaving intact the mistakes introduced in the earlier editions. Joby (2021) attempted to correct these errors by using the manuscript itself as his source and provided an online database with corrections and comparative data.

### 5.3.3 *Word Lists and Dictionaries Compiled during the Japanese Occupation*

The Japanese believed that to successfully govern Taiwan as a colony, they had to understand the people and their languages. They carried out large-scale surveys and very meticulous studies of different aspects of the societies of the indigenous peoples, including their environment (land and forestry), their traditional customs, and their languages. Anthropologists and linguists collected and published word lists for different Formosan languages, which were intended, among others, for Japanese policemen, so that they could learn these languages.

The data collected by the Japanese anthropologist Inō (1897, 1899) not only served as a basis for the vocabulary list given in Davidson (1903) but also provided the only details that we have for certain languages before they became extinct. Mori (1910) and Abe (1930), who were also anthropologists, published data for Paiwan, Amis, and Bunun. Abe (1930) also published vocabulary lists on Puyuma, Rukai, Kavalan, and Yami. Their data were transcribed in Katakana. At about the same time, Ogawa (1930b, 1931, 1933), who was a linguist, compiled pocket dictionaries of about 6,000 lexical entries each of Paiwan, Atayal, and

Amis, with Japanese translations and transcriptions in Katakana and Romanized script. The major drawback of these word lists and dictionaries is that they were “Japanese-centric,” listing Japanese vocabulary with equivalents in the Formosan languages, but not the other way around, thus leaving out important lexical items.

In the meantime, the Russian scholar Nikolaj Nevskij, who studied in Russia and Japan, worked on the Tsou language in 1927. His major publication (Nevskij 1935), translated and published in Chinese in 1993, contains a word list written in the Cyrillic alphabet.

Two comparative word lists have been compiled. The first appears in an appendix in Ogawa & Asai’s (1935) pioneering work. The second consists of word lists that Ogawa put together for all Formosan languages and dialects, collected from 163 major sources (with data from the same language/dialect extracted from more than one source). He listed 286 lexical items, arranged in 14 semantic fields (body parts, kinship terms, natural phenomena, objects, time words, directions, animals, plants, material culture items, stative verbs, active verbs, personal pronouns, interrogatives, and numerals) and by language (including Atayal, Seediq, Thao, Bunun, Tsou, Saaroa, Kananavu, Rukai, Pawai, Puyuma, Amis, Kavalan, Basay/Ketangalan, Saisiyat, Taokas, Papora, Babuza, Pazeh, Hoanya, Siraya, and Yami, in that order). This comparative list was further edited by P. Li (Ogawa 2006).

#### 5.4 Word Lists and Dictionaries Compiled after 1945

Blust (2013, p. 50) mentions that at the end of World War II, a second period of scholarly neglect ensued, broken only by the important grammar of Tsou by Tung (1964). Two things need to be noted: while many word lists were compiled (§ 5.4.1), very few dictionaries were published between 1970 and 2000, and those that were published were compiled by missionaries or foreign anthropologists and linguists. One of the main reasons why dictionaries were not produced locally was the lack of manpower: P. Li was usually the sole Taiwanese researcher working full time on the Formosan languages.

The early 2000s represent a turning point in that more substantial dictionaries were compiled and published (§ 5.4.2), many of them by indigenous people, and primary data started to be made available through online dictionaries, allowing access to a more substantial amount of linguistic information (§ 5.4.3).

### 5.4.1 *Word Lists*

Most word lists concern individual languages, including cross-dialectal variation (§ 5.4.1.1); a few offer comparative data (§ 5.4.1.2).

#### 5.4.1.1 Individual Languages

There are three types of word lists on individual languages that are discussed below: (i) those appended to grammars, (ii) those published along with phonological and/or historical studies, and (iii) those containing reediting data collected in earlier years.

Word lists can be found for individual languages as appendices of grammars and texts (e.g., Tung 1964, Zeitoun & Lin 2003, Rau & Dong 2006, De Busser 2009, Adelaar 2011, Formosan Languages Series 1–16 2018, see for instance Huang & Hayung). Tung's (1964) pioneering work on Tsou includes a glossary of 2,000 lexical items (Tsou–English, pp. 443–547 and English–Tsou, pp. 548–620), arranged alphabetically with parts of speech indicated. Derived forms (which are not always glossed) are cross-referenced to the root form, and dialectal differences are also indicated. A list of errata was provided by P. Li in the 1991 reedition of this work, correcting typographical errors, adding a (phonemic) glottal stop when it was not recorded intervocally or initially before a vowel, and changing nonsyllabic semivowels to glides. Adelaar (2011, pp. 296–400) compiled a Siraya lexicon arranged alphabetically, based on the Gospel of Saint Matthew and the Catechism in Siraya (Gravius 1661, 1662), with glosses in English and information concerning the word class and relevant grammatical function given at the beginning of the entries. At the end of his dissertation on Takivatan Bunun, De Busser (2009) provides a Bunun–English word list (pp. 541–645), word classes being indicated for each lexeme, and an index of roots (pp. 646–685). Both lists are ordered alphabetically. Each of the 16 monographs of the Formosan Languages Series (2018) includes a list of about 500 lexemes (without any grammatical description or attempt to distinguish grammatical from content words) organized alphabetically with a reverse list.

These lists were compiled at different times, and they differ in their orthography. Tung (1964) and De Busser (2009) mainly used IPA symbols. P. Li in his errata of Tung's (1964) word list switched to the Romanized orthography and used *ʔ* to represent /ʔ/ and ? for the glottal stop /ʔ/. In the Formosan Language Series (2018), the authors adopted the orthography promulgated by the Ministry of Education and the Council of Indigenous Peoples in December 2005, so the glottal stop, for instance, is represented as '.

In the 1970s and the 1980s, linguists focused mostly on the synchronic and diachronic phonologies of the Formosan languages. Reconstructions and sub-

grouping hypotheses and word lists were usually given in appendices. They consist of word lists of fewer than 1,000 lexical items from a single language or from dialects of the same language, usually organized thematically and/or alphabetically. They were published on Puyuma (Suenari 1969, Ting 1978, Tsuchida 1980), Rukai (P. Li 1977), Seediq (Yang 1976), Saisiyat (P. Li 1978), Atayal (P. Li 1981), Paiwan (Ho 1978), and Bunun (P. Li 1988). Most of these vocabularies were recorded in the IPA.

There are more recent word lists, compiled for different purposes. A Kaxabu word list divided into 24 semantic fields was compiled with translations in Taiwanese Southern Min along with Chinese characters for the purpose of language preservation and revitalization (see Phuann 2015). This was based on a preliminary (but unpublished) list of vocabulary recorded by Tsuchida (1969) on a closely related dialect, Pazeh. This is a large-scale word list containing over 2,000 entries, though some are repeated (sometimes with some discrepancies) across different fields.

Goderich (2020) provides an appendix of 1,100 lexical items reconstructed to Proto-Atayal with cognate forms in seven Atayal dialects (Matu'uwal, Plngawan, Squliq, Skikun, Klesan, Matu'aw, and S'uli), and Shibata (2020) offers a word list of nearly 500 lexical items in Bunun (five dialects) with reconstructed forms in Proto-Bunun. These studies attempt to reassess the internal relationships of the Atayal and Bunun dialects, respectively.

Word lists of about 1,000 lexical items for Basay and Trobiawan were collected by Asai in 1936–1937 and were translated into English and reorganized alphabetically (Basay–English and English–Basay) by Moriguchi (Tsuchida, Yamada & Moriguchi 1991, pp. 203–257). This was the first time that lexical data from this language were made available.

#### 5.4.1.2 Comparative Word Lists

Four comparative word lists are worth mentioning. The anthropologist and linguist Raleigh Ferrell worked on Formosan languages in the early 1960s, and his monograph (Ferrell 1969) provides a comparative word list of Formosan languages. Each of the lexical items is marked in such a way that the source is clearly indicated. The word list includes data on ten languages based on his own fieldwork (C'uli' Atayal, Seediq, Budai Rukai, Pazeh, Paiwan, Katripul Puyuma, Amis, Isbukun Bunun, Kavalan, and Yami), six cited from published and unpublished sources (among others F. Li et al. 1956, Wei & Lin 1952, Miyamoto 1932, Ogawa 1944, Ogawa & Asai 1935, Suenari 1969, Tung 1964, Taintor 1874, Thomson 1873) and two from Dutch sources (Gravius 1661, Happart 1840, p. 75). Tsuchida (1982) published a comparative vocabulary of 426 lexical items for the Western Plains languages including Taokas, Papora, Babuza, Hoanya, and

Pazeh based on Ogawa's, Asai's, and several other scholars' field notes. P. Li (1997, pp. 513–554) collected about 500 lexical items in a comparative word list including Tsou, Kanakanavu, Saaroa, (Tanan, Budai, Maga, Tona, and Mantauran) Rukai, and (Nantou and Kaohsiung) Isbukun Bunun. The ordering of the lexical items is similar to that found in other sources, i.e., these word lists are organized thematically—nouns precede verbs and other function words, such as pronouns, interrogatives, and numerals, if they are included. P. Li (2004, pp. 1483–1530) provides 235 lexical items in a comparative word list including 13 languages, among which Atayal (3 dialects, Squliq, Skikun, and Mayrinax), Seediq (4 dialects, Paran, Toda, Hecuo, and Truku), Tsou, Kanakanavu, Saaroa, Rukai (5 dialects, Tanan, Budai, Maga, Tona, and Mantauran), Bunun (3 dialects, Isbukun, Takituduh, and Takbanuaz), Paiwan (3 dialects, Butanglu, Stimul, and Tjuabar), Puyuma (3 dialects, Nanwang, Lower Pinlang, and Katripul), Thao, Saisiyat (2 dialects, Taai and Tungho), Pazeh, Amis (3 dialects, Fata'an, Farang, and Sakizaya), and Kavalan.

These four comparative word lists include an index (see Ferrell 1969, pp. 79–81, Tsuchida 1982, Li 1997, pp. 555–573, 2004, pp. 1531–1532), which makes them convenient to search through.

#### 5.4.2 *Dictionaries*

Up to the late 1990s, very few substantial dictionaries had been published (§ 5.4.2.1). More substantial dictionaries were published in the first two decades of the 21st century (§ 5.4.2.2), but to date, there are no adequate dictionaries for Atayal, Seediq, Saisiyat, Bunun, Tsou, Kanakanavu, Saaroa, and Rukai, according to Blust & Trussel (2013, p. 495).

Most of these dictionaries were recorded in the Roman alphabet and to a lesser extent in the IPA (e.g., Cauquelin 1991, 2015). Most are arranged in alphabetical order based on the source language, though only a few contain a reverse index, and a few (e.g., Duris 1970, 1988, Flatuhez 1970) are ordered according to the French alphabet. To the best of our knowledge, there are no monolingual dictionaries, which are seen as “an important step toward linguistic autonomy [and ...] a worthy goal for communities involved in language maintenance and revitalization of their languages” (Hinton & Weigel 2002, pp. 164–165). Most dictionaries are bilingual (usually either in English or in Chinese), which makes them available to only a limited number of interested people, and only five dictionaries are trilingual, providing translations of the lexical items and/or examples from the source language in both English and Chinese (see Fey 1986, Li & Tsuchida 2001, 2006, Rau et al. 2012, 2018).

All in all, these dictionaries, though representing important references, include various infelicities regarding the display and organization of the

data, word-class citations, illustrative examples, reverse index, etc., as further discussed below.

#### 5.4.2.1 1970s to 1990s

In the 1970s and the 1980s, most dictionaries were compiled by missionaries, many of whom were based in eastern Taiwan. They had the advantage of being quite fluent in the relevant Formosan language, though they did not pay particular attention to dialectal variants.

Dictionaries that were never published (and are thus not widely available) include those on Takivatan Bunun by Flahutez (1970) and Duris (1987, 1988) and those on Amis by Duris (1969, 1970) and Pourrias (1996). They contain many lexical items (up to 8,000 entries) but do not distinguish between dialectal varieties. Furthermore, certain phonemic distinctions (e.g., confusion between *q*—transcribed as *kh* in word-initial position and *hk* in word-medial and word-final positions—and *k* in Takivatan Bunun, and the pharyngeal stop vs. the glottal stop in Amis) are not made, while in other places, phonetic symbols are added that are not required.

Two dictionaries published by missionaries are worth mentioning, the first on Truku Seediq by Ferdinando Pecoraro (1977) and the second on Amis by Virginia Fey (1986), both recorded in Romanized script. Pecoraro lived in Wanrong township, Hualien county, for 17 years and collected a large amount of data on Truku Seediq. His Taroko [Truku]–French dictionary includes about 2,000 roots with over 4,000 derivations, illustrated with numerous examples, but it has no index. Though very knowledgeable in this language, Pecoraro was not always able to catch the phonemic distinction between (i) the velar stop *k* and the uvular stop *q*, (ii) the velar fricative *x* and the glottal fricative *h*, and (iii) the lateral *l* and the trill *r*. Fey's (1986) dictionary contains more than 3,000 lexical items with examples, translations in English and Chinese, and illustrations for certain words. She also misses certain phonemic contrasts, such as the distinction between the glottal stop and the pharyngealized stop, for example, e.g., Central Amis *'op'op* 'frog' vs. *qopo* 'assemble'.

Other dictionaries were published on Atayal (Egerod 1980, 1999), Paiwan (Ferrell 1982), and Puyuma (Cauquelin 1991). Søren Egerod's (1980, 1999) and Josiane Cauquelin's (1991) dictionaries were recorded in the IPA, and Ferrell's (1982) data in Romanized script.<sup>3</sup>

3 Ferrell's (1982) orthography was adopted with minor modifications in the (2005) version simultaneously promulgated by the Council of Indigenous Peoples and the Ministry of Education.

Egerod was a prominent linguist who started working on Atayal in the early 1960s and acquired a good understanding of Atayal grammar. His dictionary is based on the Sqliq dialect, as spoken in Taoyuan county, and includes about 5,000 words, with about 2,500 major lexical entries, and 2,500 derived words. It was first published in 1980 and revised in the early 1990s. The second edition was published posthumously in 1999. The first edition had fewer entries but did have an index. The second edition was expanded but did not contain an index. There are three other things to note. First, though very valuable, Egerod's dictionary contains some translation errors due to misunderstandings between the informants and the investigator, and some important lexical items are also missing, e.g., *qlup-an* 'hunting territory'. Second, while Egerod's transcriptions were very reliable, he did not recognize the phonemic status of the glides. In the first edition, the word 'past' is recorded as *uaial* rather than *wayal*. This was changed in the second edition, but Egerod kept the transcriptions of long high vowels as geminates *ii* and *uu* rather than a combination of a vowel followed by a glide as in *iy* and *uw*; compare *kgii* vs. *kgiy* 'hemp' and *uruu* vs. *uruw* 'valley'. Third, the arrangement of the dictionary is not very practical—even with the English index, it is inconvenient to search for lexical items—and many examples are not given in full, i.e., examples sometimes consist of a half clause or sentence.

Ferrell's (1982) dictionary is based on the Kuljaljau dialect, though the data were collected in the field in Tuban, Taitung county, in southeast Taiwan. It contains about 5,000 lexical entries and derivations, but very few examples, although it provides an especially large number of plant-related terms.

Cauquelin's (1991) Puyuma dictionary was modeled on Ferrell's dictionary, with about 4,000 lexical entries and derivations from the Nanwang dialect but very few examples. It was written in French, and thus not really accessible until its expanded reedition in English in 2015 (see § 5.4.2.2). The strength of Cauquelin's work is that all her interviews were conducted in Puyuma.

#### 5.4.2.2 Early 21st Century

During his one-year stay at Academia Sinica in 1994, Robert Blust stressed the importance of compiling dictionaries, a process that had been neglected for about 30 years. His call seems to have been heard, since in the early 21st century, there was an increase in the production of substantial dictionaries, written in English and/or Chinese.

P. Li and Tsuchida produced two dictionaries, the first on Pazeh (Li & Tsuchida 2001) and the second on Kavalan (2006), including data collected by both compilers as well as lexical items recorded by the Japanese linguists Ogawa and Asai. Both dictionaries include about 3,000 entries, followed by derived forms, arranged in alphabetical order and with many examples.

Blust (2003) produced one of the largest dictionaries on Thao with unusually rich language data and a 64-page English–Thao index. Blust always worked with a translator in the field—which complicated the interview process—but collected a lot of language data because of his background in Austronesian languages and his rich experience acquired through extensive fieldwork. He adopted a different format—roots are organized alphabetically, and each derived form is numbered and can be cross-referenced in the text—and a slightly distinct Romanized spelling, which might, in some cases, be confusing. For instance, *c* does not represent the alveolar affricate /ts/ as is conventional, but rather the interdental fricative /θ/, as a reflex of PAN \*C. Verbs are given with their various derivations, and there are numerous examples. Whenever possible, Blust provides the PAN etymology of the word in question and the source language (usually Bunun) if it is a loanword. He worked with the last most competent speaker (on and off from 1994 to 1999), and perhaps because of the context in which the lexical items were collected (i.e., with translators), quite a few native plant and bird names are missing.

Blust's dictionary is only surpassed in size by Namoh Rata's (2019) Amis–Chinese dictionary,<sup>4</sup> which is the longest dictionary of any Formosan language, with 32,000 entries, a total of 1,770,000 words and derivations, illustrated with many examples, and a Chinese–Amis index. Even though it contains a carefully edited English–Amis index, the dictionary is mostly in Chinese and thus unfortunately not accessible to those who cannot read Chinese. Namoh Rata's (2019) orthography is usually consistent and correct, but in some places, he missed the distinction between the pharyngeal stop and the glottal stop, as in previous Amis dictionaries.

Cauquelin (2015) revised her earlier Puyuma–French dictionary and published an expanded version with copious examples, and data exclusively transcribed in the IPA. Rau et al. (2018) are also to be commended for having published a Yami Bible dictionary (in print and online), indexing all the vocabulary found in the New Testament.

Other dictionaries include Egli (2002) for Paiwan, Tudaw Pisaw et al. (2006) for Truku Seediq, Zeng (2008) for Central Amis, and Tiang (2016) for Bunun. Egli's (2002) Paiwan dictionary is based on the Tjuabar dialect. It is much more copious than Ferrell's dictionary, with entries being supported by various examples, with translations in German. The layout leaves a lot to be desired, though: Egli did not completely understand the complex verbal morphology of Paiwan and was thus unable to systematize it in a way that would have allowed the

4 A slightly expanded edition was published in 2019.

examples to be well-categorized grammatically. Tudaw Pisaw was one of six indigenous priests who compiled a dictionary of Truku Seediq (Tudaw Pisaw et al. 2006), which contains about 3,000 lexical entries with derivations, though these are not always displayed in a consistent manner. To give but one example, *psrimal* ‘five times’ appears as a main lexical category (p. 534), but it should be treated as a subentry of *rima* ‘five’ (p. 669). A second problem is that the examples given are not as numerous as those found in Pecoraro (1977). A third drawback is its orthography. Stress falls regularly on the penult, but the penultimate *a* is left out of the orthography, e.g., *hma* ‘hand’ (instead of *həma*), creating inconsistencies with vowel-less syllables bearing stress. Zeng (2008) contains many more lexical entries including derivations (10,908 in total) than Fey (1986), but with no illustrative examples due to the publisher’s editorial decisions. Tiang’s (2016) Bunun dictionary contains about 3,000 lexical entries, arranged alphabetically, with illustrated examples. It exhibits two shortcomings: (i) entries of the same verb marked for voice differently (e.g., *maun* ‘eat (AV)’ and *kaun* ‘eat (UVP)’ appear under different entries and are not cross-referenced, and (ii) the entries are arranged alphabetically, but the symbol “*ng*”, which stands for the velar nasal, is treated as *n + g* rather than as a single phonetic symbol, so forms starting with *ng* are mixed with those starting with *n*.

### 5.4.3 Online Dictionaries

In this section, three different online dictionaries are discussed: the Council of Indigenous Peoples online dictionaries (§ 5.4.3.1), the Yami online dictionaries, (§ 5.4.3.2) and the Austronesian Comparative Dictionary (ACD) (§ 5.4.3.3).

#### 5.4.3.1 Online Dictionaries Sponsored by the Council of Indigenous Peoples

Between 2007 and 2014, the Council of Indigenous Peoples sponsored the development of 16 dictionaries (one per ethnic group) with translations in Chinese for all the officially recognized ethnic groups (Amis, Atayal, Bunun, Kanakanavu, Kavalan, Paiwan, Puyuma, Rukai, Saaroa, Saisiyat, Sakizaya, Seediq, Thao, Truku, Tsou, and Yami), compiled by research teams composed of a group of native speakers along with (usually) one specialist on the language in question. Each dictionary contains about 2,000 roots with derived forms (around 4,000 to 5,000) plus sentence examples. A particular aspect of online dictionaries is that they provide audio files of each word and sentences. Most are of good quality, though some do not sound natural because younger speakers voiced the recordings.

The dictionaries were completed over a two-year period and have been online since 2016. The data quality of these dictionaries is quite uneven. A brief

comparison is made between two online dictionaries of similar length, those for Atayal and Rukai, as based on the most representative dialects, Squiliq and Budai, respectively (henceforth “Atayal” and “Rukai”):

- (i) in Atayal, word classes are indicated for nearly all lexical items, but not in Rukai;
- (ii) under each Atayal headword, derivations are given, cross-referenced in the text, and if two words are homophonous, subscripts are used; for Rukai, derivations are not cross-referenced with their nonderived counterpart (i.e., the root) when they appear in the text, and the relation between related words is not clearly explained;
- (iii) in Atayal, the source of loanwords is clearly indicated; it is not in Rukai;
- (iv) in Atayal, affixes are not listed as entries, but in Rukai, they are presented as if they were free words, which is, of course, incorrect and rather misleading. The prefix *ara-* ‘use’, for instance, is treated as a free morpheme, when it is not. None of the voices of the derived verb forms is indicated in Atayal, which might make the reading of the examples more difficult for the layperson.

Despite their many shortcomings, these online dictionaries are a welcome addition to the materials available for Formosan languages. The compilation of these dictionaries was also a landmark in the close collaboration between indigenous and nonindigenous specialists.

#### 5.4.3.2 Yami Online Dictionaries

Different Yami dictionaries—all compiled by Victoria D. Rau and her team (including Meng-Chien Yang with the long-standing help of Maa-neu Dong and Hui-huan Chang), as part of her Yami Language Documentation project—can be found online.

The Yami dictionary project webpage contains the Yami online dictionary, funded by the Council of Indigenous Peoples (see §5.4.3.1), a learner’s dictionary, a participatory wiki dictionary, and a culture dictionary. The online dictionary includes 2,200 entries, each illustrated with examples. The wiki dictionary is a platform directed by Meng-Chien Yang, with a similar structure to that of Wikipedia. The trilingual cultural dictionary (also found in print, see Rau et al. 2012) features over 2,000 roots and nearly 4,000 derived forms, with examples drawn from the texts contained in the Yami Digital Archive Corpus.

The Yami Bible dictionary includes over 1,200 roots and 7,600 stems and nearly 9,000 examples based on the New Testament. This is the sole Formosan dictionary of this kind.

### 5.4.3.3 Austronesian Comparative Dictionary

As of its final update in June 2020, the Austronesian Comparative Dictionary (ACD), compiled by Blust (Blust & Trussel 2020), includes over 8,000 sets of reconstructions for nine hierarchically ordered protolanguages based on Blust's (1999) subgrouping hypothesis and is an unrivaled tool (Blust & Trussel 2013). Relational databases allow the user to search for different types of entries (e.g., cognate sets, proto-form indexes, and words), each entry beginning with the protolanguage to which the etymon is assigned, followed by the reconstructed form marked by an asterisk and its gloss, and the next lines listing the reflexes that support the reconstruction, with the language abbreviations on the left, reflexes in the middle, and glosses on the right. One noteworthy detail is that there is a "Formosan" entry set (with 179 reconstructed forms) that refers to cognate sets distributed among different Formosan language groups.

## 5.5 Conclusion

This chapter has provided an overview of the lexicographic work that has been undertaken on the Formosan languages over the past 400 years. These word lists and dictionaries have been produced in an attempt to understand, document, and preserve the Formosan languages. It is hoped that in the future, we will be able to produce a comprehensive and up-to-date report about which method to adopt in compiling useful, user-friendly, multilingual word lists or dictionaries in order to determine the amount of data that should be collected and for what purpose.

## References

- Abe, Akiyoshi 安倍明義 (1930). *Fan gengo kenkyū* 蕃語研究 [Study of languages of the native tribes]. Taihoku 臺北 [Taipei]: Bango Kenkyūkai 蕃語研究會 [Research group].
- Adelaar, Alexander (2006). Siraya dialogues. In Henry Y. Chang, Lillian M. Huang and Dah-an Ho (Eds.), *Streams converging into an ocean: Festschrift in honor of Professor Jen-kuei Li on his 70th birthday*, (pp. 665–685). Language and Linguistics Monograph Series W-5. Taipei: Institute of Linguistics, Academia Sinica.
- Adelaar, Alexander (2011). *Siraya: Retrieving the phonology, grammar and lexicon of a dormant Formosan language*. Berlin: De Gruyter Mouton.
- Adelaar, Alexander (this handbook). Siraya. In Paul Jen-kuei Li, Elizabeth Zeitoun and

- Rik De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*. Leiden: Brill.
- Anonymous. (ca. 1635). *Vocabularium Formosanum* [Utrecht Manuscript]. MS University Library, Utrecht. Depository-S MAG: Hs 1 E 23.
- Blust, Robert (1999). Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-Kuei Li (Eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics*, (pp. 31–94). Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No. 1. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert (2003). *Thao dictionary*. Institute of Linguistics, Academia Sinica, Language and Linguistics Monograph Series A5. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Blust, Robert (2013). *The Austronesian languages*. Asia-Pacific Linguistics Open Access Monographs A-PL 008. Canberra: The Australian National University. Retrieved from [openresearch-repository.anu.edu.au/bitstream/1885/10191/6/Blust-2013-AustronesianLanguages.pdf](http://openresearch-repository.anu.edu.au/bitstream/1885/10191/6/Blust-2013-AustronesianLanguages.pdf).
- Blust, Robert and Stephen Trussel (2013). The Austronesian comparative dictionary: A work in progress. *Oceanic Linguistics* 52(2), 493–523.
- Bullock, T.L. (1874). Formosan dialects and their connections with the Malay. *China Review* 3, 38–46.
- Campbell, Rev. William (1896). *The articles of Christian instruction in Favorlang-Formosan Dutch and English from Vertrecht's manuscript of 1650 with Psalmanazar's dialogue between a Japanese and a Formosan and Happart's Favorlang vocabulary*. London: Kegan Paul, Trench, Trübner & Co.
- Campbell, Rev. William (1903). *Formosa under the Dutch: Described from contemporary records, with explanatory notes and a bibliography of the island*. London: Kegan Paul, Trench, Trübner & Co.
- Cauquelin, Josiane (1991). *Dictionnaire puyuma-français* [Puyuma–French dictionary]. Paris and Jakarta: Ecole Française d'Extrême-Orient.
- Cauquelin, Josiane (2015). *Nanwang Puyuma–English dictionary*. Language and Linguistics Monograph Series No. 56. Taipei: Institute of Linguistics, Academia Sinica.
- Davidson, James W. (1903). *The island of Formosa, past and present: History, people, resources, and commercial prospects—tea, camphor, sugar, gold, coal, sulphur, economical plants, and other productions*. Reprint. Taipei: Southern Materials Center (SMC).
- De Busser, Rik (2009). *Towards a grammar of Takivatan Bunun: Selected topics*. PhD dissertation. Bundoora: La Trobe University.
- Duris, P. Antoine (1969). *Dictionnaire amitsu-français*. [Amis–French dictionary]. Hualien: Privately published manuscript.

- Duris, P. Antoine (1970). *Dictionnaire français-amitsu*. [French–Amis dictionary]. Hualien: Privately published manuscript.
- Duris, P. Antoine (1987). *Lexique de la langue bunun en usage à Ma-Hoan* [Lexicon of the Bunun language in use in Ma-yuan]. Hualien: Privately published manuscript.
- Duris, P. Antoine (1988). *Vocabulaire français-bunun*. [French–Bunun lexicon]. Hualien: Privately published manuscript.
- Egerod, Søren (1980). *Atayal–English dictionary*. Scandinavian Institute of Asian Studies, Monograph Series No. 35. Curzon Press. Malm, Denmark: Curzon Press.
- Egerod, Søren (1999). *Atayal–English dictionary*. Historisk-filosofiske Skrifter 20. Copenhagen: Det Kongelige Danske Videnskabernes Selskab [The Royal Danish Academy of Sciences and Letters]. (2nd ed.)
- Egli, Hans (2002). *Paiwan Wörterbuch. Paiwan–Deutsch, Deutsch–Paiwan* [Paiwan dictionary: Paiwan–German, German–Paiwan]. Wiesbaden: Otto Harrassowitz.
- Fernández Rodríguez, Rebeca (2014). Traducción de términos religiosos en los vocabularios filipinos (1565–1800) [Translations of religious terms in Philippine vocabularies]. In Otto Zwartjes, Klaus Zimmermann and Martina Schrader-Kniffki (Eds.), *Lingüística Misionera V* [Missionary linguistics V], (pp. 273–294). Amsterdam: John Benjamins Publishing Co.
- Ferrell, Raleigh (1969). *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Institute of Ethnology, Academia Sinica Monograph No. 17. Taipei: Institute of Ethnology, Academia Sinica.
- Ferrell, Raleigh (1982). *Paiwan dictionary*. Pacific Linguistics C-73. Canberra: The Australian National University.
- Fey, Virginia (1986). *Amis dictionary*. Taipei: The Bible Society.
- Flahutez, Marcel (1970). *Dictionnaire français-bunun* [French–Bunun dictionary]. Paris: Privately published manuscript.
- Goderich, Andre (2020). Atayal phonology, reconstruction and subgrouping. PhD dissertation. Hsinchu: National Tsing Hua University.
- Gravius, Daniël (1661). *Het Heylige Evangelium Matthei en Johannis ofte Hagnau Ka D'llig Matiktik. Ka na sasoulat ti Mattheus, ti Johannes appa. Overgeset inde Formosaansche tale, voor de Inwoonders van Soulang, Mattau, Sinckan, Bacloan, Tavocan en Tevorang* [The Holy Gospel according to Matthew and John or the True and Holy Gospel, according to the Books of Matthew and John: Translated into the Formosan language, for the inhabitants of Soulang, Mattau, Sinckan, Bacloan, Tavocan, and Tevorang]. Amsterdam: Michiel Hartogh.
- Gravius, Daniël (1662). *Patar ki tna-'msing-an ki christang, Ka Tauki-papatar-en-ato tmæu'ug tou Sou Ka Makka Si-Deia: Ofte 't Formulier des Christendoms: Met de Verklaringen van dien inde Sideis-Formosaansche Tale*. [Document of the Christian faith into the Sideia language: Or the Formulary of Christianity: With explanations of this in the Sideia-Formosan language]. Amsterdam: Michiel Hartogh.

- Grothe, J.A. (1884–1891). *Archief voor de geschiedenis der Oude Hollandsche Zending* [Archive for the history of the Old Dutch Mission], vols. 1–5. Utrecht: C. van Bentum.
- Guérin, M. (1868). Vocabulaire du dialecte Tayal ou aborigène de l'île Formose [Vocabulary of the Tayal or aboriginal dialect of Formosa Island]. *Bulletin de la Société de Géographie 5e Série* [Bulletin of the Geographical Society, Fifth Series] XVI, 466–507.
- Happart, Gilbertus (1650). *Woord-boek der Favorlangsche taal, waarin het Favorlangs voor; het Duits achter gestelt is*. MS.
- Happart, Gilbertus (1840). *Dictionary of the Favorlang dialect of the Formosan language by Gilbertus Happart, written in 1650, translated from the Transactions of Batavian Literary Society*, by W.H. Medhurst. Batavia: Parapatian.
- Happart, Gilbertus (1842). *Woord-boek der Favorlangsche Taal, waarin het Favorlangs voor; het Duits achter gestelt is door Gilbertus Happart* [Dictionary of the Favorlang language, with Favorlang words first and then Dutch, by Gilbertus Happart, edited by Walter van Hoëvell]. Batavia: Bataviaasch Genootschap van Kunsten en Wetenschappen.
- Hinton, Leanne and William Weigel (2002). A dictionary for whom? Tensions between academic and nonacademic functions of bilingual dictionaries. In William Frawley, Kenneth C. Hill and Pamela Munro (Eds.), *Making dictionaries: Preserving indigenous languages of the Americas*, (pp. 155–170). Berkeley, CA. University of California Press.
- Ho, Dah-an 何大安 (1978). Wuzhong paiwan fangyan de chubu bijiao 五種排灣方言的初步比較 [A preliminary comparative study of five Paiwan dialects]. *Zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* 中央研究院歷史語言研究所集刊 [Bulletin of the Institute of History and Philology, Academia Sinica] 49(4), 565–681.
- Huang, Lillian M., Yu-yang Liu and Tali' Hayung 黃美金、劉宇陽、吳新生 (2014). Cong taiyayu cidian tan taiwan yuanzhuminyu cidian zhi bianzhan 從泰雅語詞典談臺灣原住民族語詞典之編纂 [On compiling Taiwan Indigenous language dictionaries: An illustration of Atayal dictionaries]. *Taiwan yuanzhumin zu yanjiu jikan* 臺灣原住民族研究季刊 [Taiwan Journal of Indigenous Studies] 7(2), 73–118.
- Huang, Lillian M. and Tali' Hayung 黃美金、吳新生 (2018). *Taiyayu yufa gailun* 泰雅語語法概論 [A sketch grammar of Atayal]. Taiwan nandao yuyan congshu 2 臺灣南島語言叢書 2 [Series on Formosan Languages 2]. Xinbei 新北 [New Taipei]: Yuanzhuminzu weiyuanhui 原住民族委員會 [Council of Indigenous Peoples]. (2nd ed.)
- Huang, Shujing 黃叔璥 (1724). Taihai shicha lu 臺海使槎錄 [Notes of envoy to Taiwan]. *Historical written documents on Taiwan* 臺灣文獻叢刊 [Historical written documents on Taiwan]. Reprinted in Shuhui Lin 林淑慧 (2004). *Taiwan wenhua caifeng: Huang shujing jiqi taihai shicha lu* 臺灣文化采風：黃叔璥及其《臺海使槎錄》研究 [Research on Huang Shujing's "Notes of envoy to Taiwan"]. Taipei 臺

- 北 [Taipei]: Wanjuan lou tushu gongsi 萬卷樓圖書公司 [Wanjuan lou Book Company].
- Inō, Kanori 伊能嘉矩 (1897). Seiban goi sono-ichi (taiwan tsushin dai jugo kai) 生蕃語彙其の一 (臺灣通信第15回) [Vocabulary of Formosan native tribes (Correspondence from Taiwan, No. 15)]. *Tōkyō jinrui gakkai zasshi* 東京人類學會雜誌 [*The Journal of the Anthropological Society of Tokyo*] 12 (133), 270–272.
- Inō, Kanori 伊能嘉矩 (1899). Taiwan ni okeru “peipo”-zoku no gaisetsu (taiwan tsushin dai niju-hachi kai) 臺灣ニ於ケル「ペイポ」族ノ概説 (臺灣通信第28回) [Outline of the Peipo tribes in Taiwan (Correspondence from Taiwan, No. 28)]. *Tōkyō jinrui gakkai zasshi* 東京人類學會雜誌 [*The Journal of the Anthropological Society of Tokyo*] 14 (154), 126–136.
- Jaski, Bart (2021). The manuscript collection of Adriaan Reland in the University Library of Utrecht and beyond. In Bart Jaski et al. (Eds.), *The Orient in Utrecht: Adriaan Reland (1676–1718), Arabist, cartographer, antiquarian and scholar of comparative religion*, (pp. 321–361). Leiden: Brill.
- Joby, Christopher (2021). Revisions to the Siraya lexicon based on the original Utrecht Manuscript: A case study in source data. *Historiographia Linguistica* 48(2/3), 177–204. <https://doi.org/10.1075/hl.00084.job>
- Li, Fang-kuei, Chi-lu Chen and Mei-chün Tang 李方桂, 陳奇祿, 唐美君 (1956). *Shaoyu jilue* 邵語記略 [Notes on the Thao language]. *Guoli taiwan daxue kaogu renlei xuekan* 國立臺灣大學考古人類學刊 [*Bulletin of the Department of Archaeology and Anthropology, National Taiwan University*] 7, 23–51.
- Li, Paul Jen-kuei (1977). The internal relationships of Rukai. *Bulletin of the Institute of History and Philology, Academia Sinica* 48(1), 1–92.
- Li, Paul Jen-kuei (1978). A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 49(2), 133–199.
- Li, Paul Jen-kuei (1981). Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52(2), 235–301.
- Li, Paul Jen-kuei (1988). A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 59(2), 479–508.
- Li, Paul Jen-kuei 李壬癸 (1995). Taiwan nandao yuyan de cidian bianzhan jishu jiantao—jian pingjie xianyou de jibu cidian 臺灣南島語言的詞典編纂技術檢討—兼評介現有的幾部詞典 [On the lexicography of Formosan languages]. *Taiwan nandao minzu muyu yanjiu luwenji* 臺灣南島民族母語研究論文集 [*Papers on the natives languages in Taiwan*], (pp. 1–18). Taipei 臺北 [Taipei]: Jiaoyubu jiaoyu yanjiu weiyuanhui 教育部教育研究委員會 [Ministry of Education].
- Li, Paul Jen-kuei 李壬癸 (1997). Cihui bijiao biao 詞彙比較表 [Table of comparative lexicon]. In Paul Jen-kuei Li (Ed.), 李壬癸 (編), *Gaoxiongxian nandao yuyan* 高雄縣南島語言 [*The Formosan languages of Kaohsiung County*], (pp. 509–573). Gaoxiongxian wenxian congshu xilie 7 高雄縣文獻叢書系列7 [Kaohsiung County

- Documents Series 7]. Fengshan 鳳山 [Fengshan]: Gaoxiong xian zhengfu 高雄縣政府 [Kaohsiung County Government].
- Li, Paul Jen-kuei (2004). Basic vocabulary for Formosan languages and dialects. In Paul Jen-kuei Li, *Selected papers on Formosan languages*, (pp. 1483–1532). Language and Linguistics Monograph Series C-3. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei 李壬癸 (2007). Taiwan nandao yuyan de cidian bianzhuān jishu jiantao 臺灣南島語言的詞典編纂技術檢討 [On the lexicography of Formosan languages]. In Shih-hui Lin 林蒨慧 (Ed.), *Yuanzhuminzu yuyan fazhan luncong—lilun yu shiwei* 原住民族語言發展論叢—理論與實務, (pp. 1–27). Taipei 臺北 [Taipei]: 原住民族委員會 *Yuanzhuminzu weiyuanhui* [Council of Indigenous Peoples].
- Li, Paul Jen-kuei (2013). Techniques for compiling dictionaries of Formosan languages. In Paul Jen-kuei Li, translated by Noah Sauvé, *The ethnic groups, languages and migration of the Formosan natives*, (pp. 260–309). New Taipei: National Academy for Education Research, Republic of China (Taiwan) and Bookman Books, Ltd.
- Li, Paul Jen-kuei (2019). *Text analysis of Favorlang*. Language and Linguistics Monograph Series No. 61. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei (this handbook). Favorlang. In Paul Jen-kuei Li, Elizabeth Zeitoun and Rik De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*. Leiden: Brill.
- Li, Paul Jen-kuei and Shigeru Tsuchida (2001). *Pazih dictionary*. Language and Linguistics Monograph Series A2. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Li, Paul Jen-kuei and Shigeru Tsuchida (2006). *Kavalan dictionary*. Language and Linguistics Monograph Series A-19. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Thoi-yen 李台元 (2016). *Taiwan yuanzhuminzu yuyan de shumianhua licheng* 臺灣原住民族語言的書面化歷程 [The literation of Taiwanese aboriginal languages]. Taipei 臺北 [Taipei]: Zhengda chubanshe 政大出版社 [Chengchi University Press].
- Liu, Dorinda Tsai-hsiu (this handbook). Basay. In Paul Jen-kuei Li, Elizabeth Zeitoun and Rik De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*. Leiden: Brill.
- Liu, Liang-bi 劉良璧 (1742). *Chongxiu fujian taiwan fuzhi* 重修福建臺灣府志 [Topographical and historical description of Taiwanfu]. Taiwan yinghang 臺灣銀行 [Bank of Taiwan]: Taiwan wenxian congkan di 74 zhong 臺灣文獻叢刊第74種 [Taiwanese Literature Series No. 74].
- Medina, José Toribio (1896). *La imprenta en Manila desde sus origenes hasta 1810* [Printing in Manila from its beginnings until 1810]. Santiago de Chile: José Toribio Medina.
- Miyamoto, Nobuto 宮本延人 (1932). Taokasu-zoku no goi タオカス族の語彙 [Taokasu vocabulary]. Nanpō dōzoku 南方土俗 [The Ethnographical Journal of Southeast Asia and Formosa] 2(1), 61–62.

- Mori, Ushinosuke 森丑之介 (1910). *Bunun banngo shū* ぶぬん 蕃語集 [*Bunun vocabulary*]. Taihoku 臺北 [Taipei]: Minsei-bu keisatsu honbu minju-bu 民政部警務本署 蕃務課 [Office of the Taiwan Governor-General].
- Murakami, Naojirō (1933). *Sinkan Manuscripts*. Memoirs of the Faculty of Literature and Politics, Taihoku Imperial University 2(1). Taipei: Taipei (Taihoku) Imperial University.
- Nebrija, Antonio de (1979 [1492]). *Diccionario Latino-Español*. Estudio preliminar por Germán Colón y Amadeu & J. Soberanas. Barcelona: Puvill.
- Nevskij, Nikolaj A. (1935). *Материалы по говорам языка Су* (*Materials po govoram jazyka Su*) [Materials on the Tsou language]. Moscow: Труды Института Востоковедения (Trudy Instituta Vostokovedenija) [Proceedings of the Institute of Oriental Studies].
- Ogawa, Naoyoshi 小川尚義 (1930a). *Fuaborang go ni tsuite* ファボラング語について [On the Favorlang language], *Gengo to bungaku* 言語と文學 [*Language and Literature*] 4, 33-40 (1930:11).
- Ogawa, Naoyoshi 小川尚義 (1930b). *Paiwan goshū* パイワン語集 [*Paiwan vocabulary*]. Taihoku 臺北 [Taipei]: 臺灣總督府 [Gouvernement Général of Taiwan] (1930:3).
- Ogawa, Naoyoshi 小川尚義 (1931). *Atayaru goshū* アタヤル語集 [*Atayal vocabulary*]. Taihoku 臺北 [Taipei]: 臺灣總督府 [Gouvernement Général of Taiwan]
- Ogawa, Naoyoshi 小川尚義 (1933). *Ami goshū* アミ語集 [*Amis vocabulary*]. Taihoku 臺北 [Taipei]: 臺灣總督府 [Gouvernement Général of Taiwan]. (1933:6)
- Ogawa, Naoyoshi 小川尚義 (1944). *Indonesia-go ni okeru Taiwan takasago-go no ichi* インドネシア語における臺灣高砂語の位置 [The position of the Formosan aboriginal languages in the Indonesian languages]. In *Taiheiyō Kyokai* 太平洋協會 (Ed.), *Taiheiyō chiiki no hitobito to bunka* 太平洋圈民族と文化 [*Peoples and cultures in the Pacific region*, vol. 1], (pp. 448-494). Tokyo 東京 [Tōkyō]: Kawade Shobō Shinsha 角川書店 [Kawade Shobo, Ltd.].
- Ogawa, Naoyoshi 小川尚義 (1999). *Taiwan-fu-shi ni de detaru shigeru-go* 臺灣府誌に出でたる蕃語 [Aboriginal vocabulary in the *Gazetteer of Taiwan Province*]. *Taiwan genjūmin kenkyū* 臺灣原住民研究 [*Studies of Taiwan Aborigines*] 4, 159-186.
- Ogawa, Naoyoshi (2006). *A comparative vocabulary of Formosan languages and dialects by Naoyoshi Ogawa*. Asian and African Lexicon Series, No. 49. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Ogawa, Naoyoshi (original author) and Li, Paul Jen-kuei (editor) (2003). *English-Favorlang vocabulary*. Asian and African Lexicon Series No. 43. Tokyo: Tokyo University of Foreign Studies.
- Ogawa, Naoyoshi and Erin Asai 小川尚義、浅井惠倫 (1935). *Gengo ni yoru taiwan takasago zoku densetsu shū* 原語による臺灣高砂族伝説集 [*The myths and traditions of the Formosan native tribes (texts and notes)*]. Taihoku 臺北 [Taipei]: Taihoku

- Teikoku Daigaku Gengo-gaku Kenkyū-shitsu 臺北帝國大學言語學研究室 [Institute of Linguistics, Taihoku Imperial University].
- Pecoraro, Ferdinando M. (1977). *Essai de dictionnaire taroko-français* [Truku–French dictionary]. Cahier d'Archipel 7. Paris: Société pour l'Etude et la Connaissance du Monde Insulindien (S.E.C.M.I.).
- Phuann, Ing-Lik 潘永歷 (Tun Kunay) (2015). *Gehawu yu fenlei cidian* 噶哈巫語分類辭典 [A classified Kaxabu dictionary]. 董忠司、陳熾庄、潘正浩、林鴻瑞、涂文欽、朱恩成. Nantouxian pulizhen 南投埔里鎮 [Puli, Nantou]: Nantouxian pulizhen shoucheng shequ fazhan xiehui 南投縣埔里鎮守城社區發展協會 [Development Association of Shouchen Community, Nantou County].
- Pisaw, Tudaw, Yudaw Dangaw, Iyuq Ciyang, Yuki Lowsing, Jiru Haruq, Lowsi Rakaw and Iyuq Ciyang (2006). 太魯閣族語簡易字典 [Simple Truku dictionary]. Hualien 花蓮 [Hualien]: Hsiulin xiangongsuo 秀林鄉公所 [Government of Xiulin County].
- Pourrias, Louis (1996). *Dictionnaire amis-français* [French–Amis dictionary]. Hualien: Privately published manuscript.
- Price, Gareth (2019). *Language, society, and the state: From colonization to globalization in Taiwan*. Boston, MA and Berlin: De Gruyter Mouton.
- Pytlowany, Anna and Toon van Hal (2016). Merchants, scholars and languages: The circulation of linguistic knowledge in the context of the Dutch United East India Company (VOC). *Histoire Épistémologie Langage: Une autre langue globale? Le néerlandais comme langue scientifique dans l'espace extra-européen (XVIIe–XIXe siècles)* [History: Epistemology: Language: Another global language? Dutch as a scientific language in non-European sphere (17th–19th centuries)] 38(1), 19–37. doi: 10.1051/hel/2016380102
- Rata, Namoh 吳明義 (2019). *Ameizuyu cidian* 阿美族語辭典 [Amis dictionary]. Taipei 臺北 [Taipei]: Nantian shuju 南天書局 [SMC Publishing Inc.]. (2nd ed.)
- Rau, Der-Hwa Victoria, Maa-Neu Dong and Ann Hui-Huan Chang (with Daniel E. Rau and Gerald A. Rau) (Eds.) (2012). *Yami (Tao) dictionary*. Taipei: National Taiwan University Press.
- Rau, Der-Hwa Victoria, Maa-Neu Dong and Hui-Huan Chang (Eds.) 何德華、董瑪女、張惠環 (編) (2018). *Dawuyu shengjing cidian* 達悟語聖經詞典 [A Yami New Testament dictionary]. Xinbei 新北 [New Taipei]: Caituan faren shengjing gonghui 財團法人臺灣聖經公會 [The Bible Society in Taiwan].
- Rau, Victoria D. and Maa-neu Dong (2006). *Yami texts with reference grammar and dictionary*. Language and Linguistics Monograph Series A-10. Taipei: Institute of Linguistics, Academia Sinica.
- Retana, Wenceslao Emilio (1906). *Aparato bibliografico de la historia general de Filipinas* [Bibliographical apparatus of the general history of the Philippines], vol. 1. Madrid: Imprenta de la Sucesora de M. Minuesa de los Ríos.
- Shibata, Kye (2020). *A reconstruction of Proto-Bunun phonology and lexicon*. Hsinchu: National Tsing Hua University.

- Suenari, Michio (1969). A preliminary report on Puyuma language. *Bulletin of the Institute of Ethnology, Academia Sinica* 27, 141-164.
- Taintor, Edward C. (1874). Formosa: Comparative vocabulary of the Kabarab (Pepo) and Yukan (Savage) dialects of North-Eastern Formosa. In Edward C. Taintor, *The aborigines of northern Formosa: A paper read before the North China Branch of the Royal Asiatic Society, Shanghai 18th June, 1874*, (pp. 37-54). Shanghai: Customs Press.
- Thomson, J. (1873). Notes of a journey in Southern Formosa. *Journal of the Royal Geographical Society* 43, 97-107.
- Tiang Istasipal 余榮德 (Yu, Rong-de) (2016). *Danqun bunongzu yuci cidian* 丹群布農族語詞辭典 [A dictionary of Takivatan Bunun]. Hualian 花蓮 [Hualien]: Yichuan xiaomi zuu dili chubanshe 一串小米族語獨立出版社 [Millet Indigenous Languages Independent Publisher].
- Ting, Pang-hsin 丁邦新 (1978). Gu beinanyu de nice 古卑南語的擬測 [Reconstruction of Proto-Puyuma phonology]. *Zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* 中央研究院歷史語言研究所集刊 [Bulletin of the Institute of History and Philology, Academia Sinica] 49(3), 321-392.
- Tsuchida, Shigeru (1969). Pazez vocabulary. MS.
- Tsuchida, Shigeru (1972). Takasago-zoku shogo no jisho togoi-shū 高砂族諸語の辞書と語彙集 [Dictionaries and word lists in Formosan languages]. *Gakutō* 學鐙 [Gakuto] 69(5), 42-45.
- Tsuchida, Shigeru (1980). Puyuma-go (tamarakao hōgen) goi: Fu, gohō gaisetsu oyobi tekisuto ブユマ語 (タマラカオ方言) 語彙-付・語法概説およびテキスト [Puyuma (Tamalakaw dialect) vocabulary, with grammatical notes and texts] In Kuroshio Bunka no kai 黒潮文化の会 (編) [Black Current Cultures Committee] (Eds.), *Kuroshio no minzoku, bunka gengo* 黒潮の民族、文化、言語 [Peoples, cultures and languages along the Black Current], (pp. 183-307). Tōkyō 東京 [Tokyo]: Kadokawa 角川書店 [Kadokawa Shoten].
- Tsuchida, Shigeru (1982). Subclassification of Amis dialects, along with a comparative vocabulary of Amis dialects, 593 items from 20 villages. Invited talk at the Institute of History and Philology, Academia Sinica, 2 April 1982. Taipei: Institute of Linguistics, Academia Sinica.
- Tsuchida, Shigeru (1998). English Index of the Siraya vocabulary by van der Vlis. *Studies of Taiwan Aborigines* 3, 281-310.
- Tsuchida, Shigeru, Yukihiro Yamada and Tsunekazu Moriguchi 土田滋、山田幸宏、森口恆一 (1991). *Taiwan heihozoku no gengo shiryō no seiri to bunseki* 臺灣平埔族の言語資料の整理と分析 [Linguistic materials of the Formosan sinicized populations, vol. 1: Siraya and Basai], 1-94. Tōkyō 東京 [Tokyo] Tōkyō Daigaku 東京大学 [University of Tokyo].
- Tung, T'ung-ho with the assistance of S.H. Wang, T.K. Kuan, T.F. Cheng and Margaret M. Yan (1964). *A descriptive study of the Tsou language, Formosa*. Institute of History

- and Philology, Academia Sinica Special Publications, No. 48. Taipei: Institute of History and Philology, Academia Sinica. (List of errata by Paul Jen-kuei Li in the 1991 reedition of this work).
- Vlis, Christianus Jacobus van der (C.J.) (1842). Formosaansche woorden-lijst, volgens een Utrechtsch handschrift: Voorafgegaan door eenige korte aanmerkingen betreffende de Formosaansche taal [Formosan word list, according to a Utrecht manuscript: Preceded by some short observations on the Formosan language]. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, dl. 18, [Proceedings of the Batavian Society for the Arts and Sciences, part 18], (pp. 437–488). Batavia: Bataviaasch Genootschap van Kunsten en Wetenschappen,
- Wei, Huei-lin and Lin Heng-li 衛惠林、林衡立 (1952). Caozuzhi 曹族誌 [Ethnography of the ethnic group Tsou]. *Taiwanshen tongzhigao juanba zhouzhi* 臺灣省通誌稿卷八同青誌 [Historical manuscripts of Taiwan province, vol. 8]. Nantouxian nantou 南投縣南投 [Nantou County, Nantou]: Taiwanshen wenxian weiyuanhui 臺灣省文獻委員會 [Commission of Historical Research of Taiwan].
- Yang, Hsiu-Fang 楊秀芳 (1976). Saideke wushe fangyan de yinyun jieyou 賽德克霧社方言的音韻結構 [Phonological structure of Paran: A dialect of Sediq]. *Zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* 中央研究院歷史語言研究所集刊 [Bulletin of the Institute of History and Philology, Academia Sinica] 47(4), 611–706.
- Zeitoun, Elizabeth and Hui-chuan Lin (2003). *We should not forget the stories of the Mantauran: Memories of the past*. Language and Linguistics Monograph Series A4. Taipei: Institute of Linguistics, Academia Sinica.
- Zeng, Si-qi 曾思其 (2008). *Taiwan Amis yu changong cizhui yu dongci cigen shouce* 臺灣 Amis 語常用詞綴與動詞詞根手冊 [A handbook of productive affixes and verb roots in Amis]. Beijing 北京 [Peking]: Minzu chubanshe 民族出版社 [Minority Languages Publisher].
- Zhou Zhongxuan 周鐘瑄 (1717). *Zhu luo xianzhi* 諸羅縣志 [Topographical and historical description of Tsulohsien]. *Fengshi dibace* 風俗志 第8冊 [Ethnology, vol. 8].
- Zhu Shijie 朱仕玠 (1765). *Xiao liuqiu manzhi juanshi* 小琉球漫誌卷十 [A desultory account of Hsiaoliuchiu, vol. 10, 1765]. *Xiadanshuishe jiyu* 下淡水社寄語第10冊 [Words of Lower Tamsui tribe 10].

### Online Resources and Open Access Data

- Blust, Robert and Stephen Trussel (2020). *The Austronesian comparative dictionary*. Retrieved from [trussel2.com/acd/](http://trussel2.com/acd/).
- Council of Indigenous Peoples 原住民族委員會 (2021). *Yuanzhuminzu yuyan xianshang cidian* 原住民族語言線上辭典 [Online dictionaries of Indigenous languages]. Retrieved from [e-dictionary.ilrdf.org.tw/](http://e-dictionary.ilrdf.org.tw/).
- Council of Indigenous Peoples and Ministry of Education 原住民族委員會和教育部 (2005). *Yuanzhuminzu yuyan shuxie xitong* 原住民族語言書寫系統 [Orthographic

- systems for indigenous languages]. Retrieved from [ilrdc.tw/research/rwview/rwystem.php](http://ilrdc.tw/research/rwview/rwystem.php).
- Joby, Christopher (2021). Revisions to the Siraya lexicon based on the original Utrecht Manuscript. Retrieved from [neerlandistiek.nl/ebooks/](http://neerlandistiek.nl/ebooks/).
- Merriam-Webster dictionary. Retrieved from [merriam-webster.com/dictionary/dictionary](http://merriam-webster.com/dictionary/dictionary).
- Yami dictionary of the New Testament. Retrieved from [ccunix.ccu.edu.tw/~lgrau/Yami%20Bible%20Web/0809-mix.html](http://ccunix.ccu.edu.tw/~lgrau/Yami%20Bible%20Web/0809-mix.html).
- Yami dictionary project. Retrieved from [yamibow.cs.pu.edu.tw/corpus\\_ch.htm](http://yamibow.cs.pu.edu.tw/corpus_ch.htm) and [yamionto.cs.pu.edu.tw/tao\\_dict/lexicon/main.htm](http://yamionto.cs.pu.edu.tw/tao_dict/lexicon/main.htm).