Philological and Linguistic Analysis Working Together

Exploring Language Diversity in Japan

edited by
Paolo Calvetti

Edizioni
Ca'Foscari

e-ISSN 2724-2285 ISSN 2724-1203

Ca' Foscari Japanese Studies 24 Linguistics and Language Education 4

Philological and Linguistic Analysis Working Together

Ca' Foscari Japanese Studies
Linguistics and Language Education

Collana diretta da | A series directed by
Luisa Bienati, Paolo Calvetti,
Toshio Miyake, Massimo Raveri

24 | 4

Edizioni
Ca'Foscari

Ca' Foscari Japanese Studies
Linguistics and Language Education

Head office

Ca' Foscari Japanese Studies

Dipartimento di Studi sull'Asia e sull'Africa Mediterranea
Università Ca' Foscari Venezia
Palazzo Vendramin dei Carmini
Dorsoduro 3462
30123 Venezia

e-ISSN 2724-2285
ISSN 2724-1203

URL https://edizionicafoscari.unive.it/en/edizioni4/collane/ca-foscari-japanese-studies/

Philological and Linguistic Analysis Working Together

Exploring Language Diversity in Japan

edited by Paolo Calvetti

Venezia
Edizioni Ca' Foscari - Venice University Press
2024

Philological and Linguistic Analysis Working Together. Exploring Language Diversity in Japan
Edited by Paolo Calvetti

Scientific certification of the works published by Edizioni Ca' Foscari: the essay here published has received a favourable evaluation by subject-matter experts, through a double-blind peer review process under the responsibility of the Advisory Board of the series. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari, using a dedicated platform.

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari: il saggio qui pubblicato ha ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione doppia anonima, sotto la responsabilità del Comitato scientifico della collana. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari, ricorrendo all'utilizzo di apposita piattaforma.

Edizioni Ca' Foscari
Fondazione Università Ca' Foscari | Dorsoduro 3246 | 30123 Venezia
edizionicafoscari.unive.it | ecf@unive.it

1st edition August 2024
ISBN 978-88-6969-845-3 [ebook]
ISBN 978-88-6969-846-0 [print]

Cover design: Lorenzo Toso

This book has been published with the support of ISEAS, Kyoto

ISEAS | Scuola Italiana di Studi sull'Asia Orientale

Philological and Linguistic Analysis Working Together. Exploring Language Diversity in Japan / edited by Paolo Calvetti. — 1. ed. — Venice: Edizioni Ca' Foscari, 2024. — viii + 110 p.; 23 cm. — (Ca' Foscari Japanese Studies; 24, 4). — ISBN 978-88-6969-846-0.

URL https://edizionicafoscari.unive.it/it/edizioni/libri/978-88-6969-846-0/
DOI http://doi.org/10.30687/978-88-6969-845-3

Philological and Linguistic Analysis Working Together
Exploring Language Diversity in Japan
edited by Paolo Calvetti

Abstract

Japan's ecological variety seems to parallel the multitude of different languages and dialects attested in the Japanese archipelago. In addition to standard Japanese and its dialects, there exist other Japonic languages such as Ryūkyūan and Hachijō, as well as non-Japonic varieties like Ainu. In this volume, four articles explore the importance of a philological approach to sources for historical linguistics: "Adopting a Philological Approach Toward *Chishi* (地誌)" by Étienne Baudel; "Elements of Sakhalin Ainu Phonetics, Phonology, and Morphosyntax in Bronisław Piłsudski's Corpus of Ainu Folklore" by Elia Dal Corso; "The Language of Miyako Oral Traditions" by Aleksandra Jarosz; "Reflexes of Proto-Ryukyuan Mid Vowels in *Haedong Chegukki*" by Marc Miyake.

Keywords Japanese historical linguistics. Japanese philology. Proto-Ryukyuan. Hachijō language. Ainu phonetics. Ryukyuan language. Miyako oral tradition. Haedong chegukki.

Philological and Linguistic Analysis Working Together
Exploring Language Diversity in Japan
edited by Paolo Calvetti

Table of Contents

Philological and Linguistic Analysis Working Together

Exploring Language Diversity in Japan

Philological and Linguistic Analysis Working Together
Exploring Language Diversity in Japan
edited by Paolo Calvetti

Introduction

Paolo Calvetti
Università Ca' Foscari Venezia, Italy

In memory of Professor Silvio Vita
(1954-2023)

In October 2022, the ISEAS (Italian School of East Asian Studies – Scuola Italiana di Studi sull'Asia Orientale) in Kyoto organised an international workshop titled "Exploring Linguistic Diversity in Japan: How Philological and Linguistic Analysis Can Work Together" and dedicated it to the memory of language historian Professor Alexander Vovin (1961-2022), one of the most important figures in the West in recent decades, in Japanese historical linguistics studies. The event involved participants from several continents who had been academically connected with Alexander Vovin, and had made the relationship between philology and historical linguistics the main subject of their research.

The workshop hosted a stimulating discussion among scholars whose research focused on the geographical space of Japan, albeit with different linguistic subjects or approaches. Afterwards, with the scientific coordinator of ISEAS, Professor Silvio Vita, and in collaboration with the researchers who had participated in the workshop, it was decided to organise a publication presenting the results of their research. The ISEAS Scientific Committee, chaired by Professor Giorgio Amitrano, welcomed the idea of publishing this volume in Ca' Foscari Japanese Studies, being Ca' Foscari University of Venice one of the institutions supporting ISEAS activities.

This volume brings together a selection of four works: "Adopting a Philological Approach Toward *Chishi* (地誌)" by Étienne Baudel; "Elements of Sakhalin Ainu Phonetics, Phonology, and Morphosyntax in Bronisław Piłsudski's Corpus of Ainu Folklore" by Elia Dal Corso;

"The Language of Miyako Oral Traditions" by Aleksandra Jarosz; "Reflexes of Proto-Ryukyuan Mid Vowels in *Haedong chegukki*" by Marc Miyake.

The first essay proposes an analysis of sources dating back to the Edo period that describe Hachijō, a small endangered Japonic variety spoken on three islands (Hachijō-jima, Hachijō-kojima, and Aogashima) located south of the Izu archipelago, approximately 300 km south of Tōkyō. The considerable number of available sources from the Edo period provides a detailed description of the Hachijō language over a period of about 100 years. The philological analysis of these sources may be of fundamental importance for reconstructing any diachronic changes in this language variety, which is considered the sole remaining form of Eastern Old Japanese.

The essay by Elia Dal Corso presents linguistic data from a collection of traditional Ainu folklore published in 1912 by Bronisław Piłsudski, compares them with more recent data collected in the 1960s-1970s, and analyses the existing differences between the East Sakhalin and West Sakhalin Ainu subgroups. In particular, the author focuses on phonemic inventory and phonotactic rules, some phonological alternations, and verbal morphology, highlighting a number of important differences between the two Ainu dialect sub-groups of East Sakhalin and West Sakhalin. Following the general discourse of this volume, the author has underlined how the description of data from old sources can contribute to the advancement of the historical linguistics of East Asian languages.

The third contribution by Aleksandra Jarosz points out that, although many Ryukyuan language varieties do not possess an established 'literary language', there is ample room for using the sources of phonographic transcriptions of oral texts like, for example, those of the Miyako language, such as songs, chants, narratives, proverbs, etc., dating back to the eighteenth century as the basis of what the author defines as "Miyako philology". The language of the Miyako oral traditions displays numerous archaisms and stylisations and a diachronic and comparative analysis of this data provides insights into the otherwise impossible history of the language, which also has interesting implications for the study of the Japonic languages in general.

The last article, by Marc Miyake, relies on the *Haedong chegukki*, a Korean work from the early sixteenth century, which provides important information for the reconstruction of Proto-Ryukyuan phonology, adding new elements to the results achieved so far by other scholars, and questioning some aspects of these earlier studies. Miyake, starting from a survey of hangul transcriptions of Proto-Ryukyuan mid vowel reflexes in *Haedong chegukki* Okinawan, comes to propose the reconstruction of *əj and *a(:)j in Proto-Ryukyuan and a later date for the realisation of aspiration before non-low vowels. As the author himself admits, the elaboration of such theories requires

a comparative analysis of the Old Okinawan of *Omoro sōshi*, the native spelling of the Old Okinawan of *Omoro sōshi* and the modern Ryukyuan varieties. Nevertheless, this contribution reaffirms the importance of the interplay between philology and linguistics, especially in historical linguistics.

I am, therefore, particularly pleased that a selection of the papers originally conceived for the Kyoto workshop organised by ISEAS is now being printed and will contribute to the studies of historical linguistics related to the Japanese archipelago and promoting the support philological studies can provide to linguistic research.

It deeply saddens me that, just as we were about to start editing this volume, Professor Silvio Vita, my long-time friend and colleague, passed away prematurely. The editing of the collection of essays, which I was to share with Silvio, was therefore carried out by me alone, but it is to him that I dedicate this book.

Philological and Linguistic Analysis Working Together
Exploring Language Diversity in Japan
edited by Paolo Calvetti and Giuseppe Pappalardo

The Language of Miyako Oral Traditions

Aleksandra Jarosz
Nicolaus Copernicus University in Toruń, Poland

Abstract Although Ryukyuan languages with the exception of Okinawan are essentially unwritten, they are rich in oral traditions whose language displays unique characteristics compared to the spoken register. This applies also to the Miyako language, with its wealth of traditional texts transmitted orally, including songs, chants, narratives, proverbs, aphorisms and riddles, and the oldest phonographic written records dating to eighteenth century. Research of the linguistic aspects of such texts should be of great value, both as a specialized register in its own right and in terms of a contribution to the broad historical-comparative research of Japonic languages. The present chapter provides the profile of Miyako as a modern spoken Japonic language (§ 1), presents theoretical essentials concerning the Miyako oral literature as well as an outline of the history of its documentation and research (§ 2), addresses general characteristics of the literary language (§ 3) and discusses some of its conservative features (§ 4). It introduces the basics of what can be labeled 'Miyako philology' while arguing for the importance of pursuing this field.

Keywords Oral literature. Miyako. Ryukyuan. Songs. Narratives.

This research was supported by the funding "Excellence Initiative – Debuts IV", grant no. [4101*66, 03.01*2356] as part of the project "Case marking and valence patterns in Ryukyuan: a typological and comparative study".

Edizioni Ca'Foscari

Ca' Foscari Japanese Studies 24
Linguistics and Language Education 4
e-ISSN 2610-9395 | ISSN 2610-8976
ISBN [ebook] 978-88-6969-845-3 | ISBN [print] 978-88-6969-846-0

Peer review | Open access
Submitted 2023-09-18 | Accepted 2024-01-11 | Published 2024-08-28

DOI 10.30687/978-88-6969-845-3/001

1 Miyako-Ryukyuan: The Profile of a Japonic Language

Miyako-Ryukyuan (Miyako) is one of minority Japonic languages spoken in the Okinawa prefecture in Japan. It comprises the area of the Miyako island group with its seven traditionally inhabited, small islands centered around the Miyako main island. Genetically, Miyako is a representative of the South Ryukyuan/Sakishima subgroup of the Ryukyuan group, together with the languages of Yaeyama and Yonaguni.

Like in the case of other Ryukyuan languages, there is no census data available concerning the demographics of the Miyako language. Rough estimates propose that at the moment the population counts c. 20,000 speakers, a definite majority of them being in their sixties or older (cf. Jarosz 2024). The language is therefore seriously endangered, with few speakers in the productive age. It has no official state-level recognition and no formally regulated presence in the administrative or public spaces, which, again, resembles the situation of all other minority Japonic languages.

The regional varieties, or topolects, of Miyako-Ryukyuan are in general mutually intelligible and share most of the essential characteristics (although there are some studies which counter this perspective, cf. Takubo 2021, 65). Nevertheless, they are perceived by the speakers as clearly distinct, mostly due to marked phonological, but also grammatical and lexical differences. The vitality of specific topolects, and therefore the prospect of their survival into next generations, is varied. Some topolects are considered more robust than others. The Irabu varieties (Sawada/Nagahama, Kuninaka) are recognized among the Miyako locals as those with relatively high proportions of female speakers, whereas Ikema with its subvariety of Sarahama are reported to have individual fluent second language speakers among young (below 30) males working in the fishing industry, where they interact with older fishermen who still primarily use Ikema in their workplace communication with one another (cf. an interview with a Sarahama speaker in Linguapax Asia 2019). On the other hand, there are varieties on the verge of extinction, with only several speakers at a very advanced age left in depopulated areas. The outlook is particularly bleak for remote insular topolects such as Minna, Ōgami and Kurima.

Typologically, Miyako matches in many respects the usual profile of a Japonic language, the key characteristics including: the SOV and modifier-head constituent order; dependent-marking; nominative-accusative alignment; agglutinative and mostly synthetic morphology; finite and non-finite verb morphology; systemic, morphological encoding of information structure values (topic and focus), chiefly on nominals; mora and pitch accent-based prosody.

Miyako is also noted for a number of features which are rather unique in the context of Japonic languages. Although, as mentioned,

all Japonic languages are predominantly synthetic, the synthetic tendencies in Miyako are above average due to its relatively extensive grammaticalization processes. This is observed in nominal morphology, in which vowel-initial case (accusative *-u*) and information structure markers (topic *-a*) are subject to a range of phonotactic principles which prevent the breach of the language-specific distributional rules (cf. Takubo 2021), as well as in aspect marking, in which the auxiliary verb from the original construction appears in a variant word form as a suffix hosted by the original lexical verb.

pstu-a => *psto*: person-TOP
midum-a => *midumma* woman-TOP
usɨ-u => *ussu* cow-ACC

progressive *ɕ-i: uz* do-CVB PROG.NPST > *ɕ-u:z* do-PROG.NPST
resultative *aʑʑ-i utsɨ* say-CVB RSL.NPST > *aʑʑu:tsɨ* say-RSL.NPST

In some instances, these synthesizing tendencies are coupled with an increased fusionality or lead to lexicalizations. Thus, the conventional 'good night' greeting, *nivvju:ratɕi*, is ultimately derived from an analytic honorific construction *nivvi uramatɕi = nivv-i ur-ama-tɕi* sleep-CVB PROG-HON-IMP. Similarly, a 'how do you do?' question *dzo:kara:mma* has a more analytic counterpart *dzo:-kar-i ur-ama-z-na* good-VRB-CVB be-HON-NPST-YNITR, with the former omitting the auxiliary verb 'to be' *uz* by attaching the honorific suffix *-ama* directly to the verbalized adjectival stem.

The original constructions underlying such honorific or aspectual expressions are attested and synchronically still recoverable. In contrast, there is no evidence that the underlying forms of the topic and accusative marking were ever permitted and actually used at any past stage of the language.

Another distinctive feature of Miyako is its relatively complex syllable structure and distribution patterns. While standard Japanese is known for its syllabic simplicity with a strong predominance of open syllables and a consistent avoidance of consonant clusters, and most other Japonic varieties follow this pattern, Miyako enjoys a richness of syllable types and a range of moraic consonants (usually nasals and fricatives) which can take any position in a syllable, including the nucleus. This results in frequent consonant clusters: *ftsɨ* 'mouth', *pstu* 'person', as well as entire vocabulary items consisting only of a syllabic consonant:[1] *vv* 'to sell', *mm* 'sweet potato', *zz* 'gruel'. The presence

1 The consonant in such items is necessarily long (geminated) to satisfy the minimality constraint – a rule that a phonological word needs to be at least two morae long, pervasive among Ryukyuan languages.

of geminated/lengthened variants of voiced consonants, including sonorants, is one more unusual characteristic for a Japonic language.

In contrast to these innovative structural properties, Miyako is also recognized as a language rich in conservative features, such as the retention of Proto-Japonic *p, preservation of the two Proto-Japonic valence markers *-(r)aje and *-(r)are (cf. Jarosz 2023, and also § 4.2.1. below), and perhaps most famously, a thick layer of archaic vocabulary of Proto-Japonic ancestry shared with pre-modern Japanese which has since been lost in standard/common Japanese as well as in most Japonic topolects. Examples include adjectival stems such as *apara* 'beautiful' (OJ *apare*),[2] *kupa* 'hard' (OJ *kəpa*); native numerals in the row of tens and above, e.g. *misu* '30', *jusu* '40', *isu* '50', *mumu* '100' (OJ *miso*, *jəso*, *iso*, *momo* respectively); first person pronoun *a*, found in a limited number of topolects such as Irabu-Nakachi, Tarama, and Kurima (OJ *a*, *are*); verbs such as *udzɨm ~ vdzɨm* 'to bury' (OJ *undumu*), *uguna:iz* 'to gather' (OJ *ungonaparu*).

Due to these unique developments as well as retention of archaic features, Miyako is perceived as a highly valuable language in terms of not just synchronic/typological studies of the Japonic family, but also diachronic/historical studies (cf. e.g. Nakama 1992). Scholars have appreciated the diachronic and synchronic importance of the language of Miyako since the early twentieth century and the inception of modern linguistic studies in Japan, followed by the discovery of Sakishima languages as valid objects of a linguistic scrutiny. The narrative of Miyako-Ryukyuan as providing crucial material for comparative studies, linguistic and cultural alike, was shared by the influential minds of the pre-war Japan: Kunio Yanagita, Fuyū Iha, as well as the Russians Nikolay Nevskiy and Evgeniy Polivanov (cf. Jarosz 2015). The importance of studying oral traditions of the Miyako islands, which was expected to reflect still more archaic layers of an already "archaic" language, was emphasized in particular by Nevskiy, whose initial main purpose of the fieldwork in Miyako was to study folk songs in Miyako (Jarosz 2015, 144), preferably as remembered by the oldest members of the community and as reflected in the archaic Miyako song language (Katō 2011, 137-40).

2 Miyako Oral Literature: Outline and Sources

Overall, pre-modern Ryukyuan languages can be considered unwritten. A notable exception is Okinawan, with a focus on Shuri-Okinawan, the language of the erstwhile Ryukyu Kingdom elites. Nevertheless, even the Okinawan corpus, which incorporates predominantly the

2 OJ phonological values in essence follow Miyake (2013) and Vovin (2020).

ritual song collection *Omorosōshi* おもろさうし (*omoro* songbooks; compiled in sixteenth–seventeenth centuries), *kumiodori* (*kumiwudui*) plays 組踊 (ensemble dance) and *ryūka* poetry 琉歌 (Ryukyuan poems; eighteenth century onwards) has not led to an establishment of a stable orthography nor writing conventions which would be shared community-wise and readily reproducible at a large scale. The mass literacy introduced together with compulsory education following the annexation of the Kingdom to Japan (1872-79) has always concerned Japanese and Japanese only (cf. Ogawa 2015). Any attempts at writing Ryukyuan with Ryukyuan speakers set as target readership have been a result of local bottom-up efforts, often in the context of language documentation and/or revitalization, and it is only very recently that the local administration has undertaken initiatives to regulate the writing conventions for respective Ryukyuan languages, using the Japanese *kana* syllabaries as the default writing system, the result being a document with orthographic recommendations published in 2022 (Okinawa-ken Bunka Kankō Supōtsu-bu 2022).

The situation of Miyako is no exception. While being an essentially unwritten language, Miyako boasts a rich and diverse corpus of oral literature,[3] which has a prominent place in the history of Japan's folklore. Some of the song genres are considered to retain formal layers which are still more ancient than those reflected in the oldest preserved Japanse chronicles, the eight-century *Kojiki* 古事記 (Records of Ancient Matters) and *Nihon Shoki* 日本書紀 (The Chronicles of Japan) (Uemura 1997, 318). Miyako is in particular known for its long epic songs called *ajagu* アヤグ or *a:gu* アーグ, whose content either relays the stories of famous legendary or historical figures, such as Yonaha Sedo Toyomi-oya 与那覇せど豊見親 (fourteenth century) and Nema Ikari 根間いかり (fifteenth century, both names japanized; Inamura 1962), or sings the plight of common people, mostly women. This genre is so representative of Miyako that the label has also come to be used to refer to Miyako folk songs in general (cf. Hokama, Shinzato 1972, 333).

Following a simplified version of classifications of the Miyako oral literature by Hokama and Shinzato (1972, 333-7) and Shinzato (2005, 12-28) as well as my own observations rooted mostly in the study of Nikolay Nevskiy's records (Jarosz 2015; 2022), one can divide the corpus into poetry and prose. Poetry is understood here as rhythmic texts which follow formal conventions such as versification and parallel phrases. Poetry is represented by the following subgenres: chants,

3 In fact, even after the introduction of mass literacy in the Miyakos, e.g. in the Karimata village it was still prohibited to write down those genres of Miyako songs which contained religious/taboo content and were used in ritual contexts, such as *ni:ri* ニーリ (cf. Karimata 2011, 150, 155; Okuhama 2022, 182).

such as medical magic formulas or children's counting chants; religious songs, used in prayers and rituals, some of which are considered to be almost as old as the Japonic settlement in the Miyako islands,[4] with genres such as *pjaːɕi*, *tukurufum*, *fusa*, *taːbi* and *niːri*; epic songs, which include the *aːgu* in the narrow sense as well as *kuitɕa*/*kuitɕa aːgu*, performed as group song and dance; and lyrical songs called *taugani*/*toːgani* (pronunciation dependent on the region), usually sung on festive occasions and sometimes improvised, whose equivalent in Tarama is called *ɕunkani*. Unlike Japanese *waka* 和歌 (native Japanese poems) (5-7-5-7-7) or Okinawan *ryūka* (8-8-8-6), the Miyako songs do not have a fixed metrum, although they have a strong tendency to use four- and five-mora-count verses (Inamura 1962).

Prose includes more or less fixed forms such as folk tales, proverbs, riddles, and worldview-expressing aphorisms, often pertaining to some sort of belief/superstition or reliant on some kind of metaphor.

Despite its status of an unwritten language, there are also some early modern records of Miyako which date back to eighteenth century. These are the eight epic songs included in the Ryukyu Kingdom chronicle *Yōsei Kyūki* 雍正旧記 (Chronicles of the Yōsei Era, 1727), which is part of a Miyako-related chronicle trilogy unofficially referred to as *Miyakojima Kyūki* 宮古島旧記 (Sunagawa 1983; Shinzato 2005, 368). The songs are recorded in what can be labeled *ryūkyūgana* 琉球仮名:[5] an orthography based on the Japanese *hiragana* script with the addition of semantic *kanji* sinograms, first used for the purpose of notation of Old Okinawan e.g. in the *Omorosōshi* (Serafim, Shinzato 2021) and in some early documents such as the kingdom's decrees collected in *Danake Monjo* 田名家文書 (Documents of the House Dana) (Hokama 1971, 33; Tartarini 2009, 44).

Due to the differences in phoneme inventory and phonotactics between Old Okinawan (sixteenth-seventeenth centuries) and pre-modern Japanese, resulting in underrepresentation of some Okinawan sounds/sound sequences and overrepresentation (heterography) of others, especially those related to close vowels and palatalization-related syncope (vowel loss), the accuracy of these orthographic conventions is limited. Difficulties in the decipherment of *ryūkyūgana* notations are compounded by their fossilization: the notations did not change over the centuries to reflect e.g. the progression of vowel

4 Shinzato (2005, 24) estimates that most of the few *niːri* songs passed down until this day were already in use at least since mid-fifteenth century.

5 In reference to Japanese *rekishiteki kanazukai* 歴史的仮名遣い (historical *kana*-spelling), Hokama (1971, 33) calls this orthographic convention *Okinawa-no rekishiteki kanazukai* 沖縄の歴史的仮名遣い (Okinawan historical *kana*-spelling).

raising in Middle Okinawan (eighteenth-nineteenth centuries; cf. e.g. Hattori 1978-79; 1979b).

Ryūkyūgana is therefore difficult to process to a modern reader even when it is used in notation for Okinawan, for which it was devised. The difficulties are only compounded when *ryūkyūgana* is employed to transcribe another language, such as Miyako, which has a still different phoneme inventory and syllable structure (cf. § 1). As a consequence, the decipherment and interpretation of the songs recorded in *Yōsei Kyūki* is impossible without specialist knowledge of the song register of the Miyako language and the diachronic stages which may be reflected in these pre-modern texts.

Modern academic studies of Miyako oral traditions began in the late nineteenth century with *Miyakojima-no Uta* 宮古島之歌 (*Songs from the Miyako island*), a collection of 89 pieces (with additional 11 songs from Yaeyama) recorded by Risaburō Tajima, released in 1893 (Shinzato 2005, 368). Apart from the eight songs which are a reproduction of the *Yōsei Kyūki* records, the collection is a result of Tajima's own fieldwork. Again, the usability of the volume is, however, diminished by Tajima using the *ryūkyūgana* conventions. Although Tajima devised some diacritic markings to indicate Miyako-specific sounds (such as using the upper right circle <˚> to represent a moraic consonant) and appended phonographic readings to some of the sinograms used in the text, this is still insufficient to provide ready, unambiguous phonological interpretation of the transcripts. *Miyakojima-no Uta* also does not include any complete translations or interpretations of the songs, only providing glossing and commentaries at random intervals, with many of the songs left without any comments at all.

Thus, while being of a high documentational value, Tajima's *Miyakojima-no Uta* is not an academic result per se. It did, however, serve as a basis for a number of studies of Tajima's successors, most notably Fuyū Iha and Nikolay Nevskiy (Shinzato 2005, 369-70).

Nevskiy, who conducted his Miyako fieldwork in the 1920s, can be considered the father of modern Miyako literary studies. He was the first to transcribe the Miyako language material in a phonetic (IPA-based) notation, and he also provided detailed linguistic and content analysis of many of his recorded texts. Some of the songs and stories that he documented also have full Russian and/or Japanese translations appended. His records have been released or re-released in Nevskiy and Oka (1971), Nevskiy (1978, Japanese edition 1998; 1996), and are discussed in detail in Jarosz (2015; 2022).

Systematic research of Miyako songs coupled with insights into their historical and ethnographic background was also initiated in the 1920s by the native Miyako scholar Kōnin Kiyomura. Although Kiyomura's records are equipped with translations into Japanese and commentaries, their usability is yet again limited by the employment

of an obsolete *hiragana*-based notation system. Nevertheless, despite retaining some of the non-phonological conventions of the early modern *ryūkyūgana*, Kiyomura's notation is unquestionably better adjusted to representing the sounds of the Miyako language, and with some effort it can be deciphered based on the knowledge of modern spoken Miyako language (cf. Kiyomura 2008 [1927]). Such *kana*-based notations of Miyako literary material continued to prevail in the research conducted by other Miyako locals, such as Kempu Inamura (e.g. Inamura 1977).

Documentation and study of Miyako oral literature saw a genuine surge in the post-war period, predominantly through the efforts of Shuzen Hokama, Kōshō Shinzato and Eikichi Hateruma. Their collections of songs include phonetic as well as syllabic notations and complete Japanese translations. Hokama is also credited as the most active contributor to theoretical/classifying research of Miyako poetry. Cf. Shinzato (2005, 367-87) for a historical overview of research on the Miyako oral traditions as well as a detailed bibliography up to the year 2003.

Shirō Hattori, the eminent Japanese linguist and the demiurge of historical Japonic linguistics, wrote in reference to Hokama and Shinzato (1978):

> These traditional [Miyako] songs retain old vocabulary which parallels that of the *Omorosōshi*. It is important to record such pieces. This could become a source which rivals the *Omorosōshi*. (Hattori 1979a, 432)

In so doing, he recognized the intrinsic value of the linguistic material contained in the traditional Miyako songs. Almost half a century following this evaluation by Hattori, a systematic linguistic study of Miyako songs and other literary material is yet to be attempted, a fact which emphasizes the importance and urgency of such work. Obscurity of the song language at times even to the singers has been an obstacle which hampered the progress and accuracy of the research of the Miyako literary traditions (cf. Shinzato 2005, 374; 378).

With most of the documentation focus targeted at the Miyako songs/poetry, the documentation of prose as represented by narratives has lagged behind. Nevertheless, there is some material available, starting again from Nevskiy's research (cf. Nevskiy, Oka 1971; Jarosz 2015; 2020a). The pinnacle of folk tale documentation has to be a two-volume collection of 231 narratives from the Gusukube region on the Miyako main island (Fukuda et al. 1991), written in the hybrid *kanji-kana* script with complete Japanese translations appended.

Miyako also enjoys a relatively rich assortment of publications on proverbs. There are hundreds of items, accompanied by explanations and interpretations, in collections using a *ryūkyūgana*-based

orthography (Yoshimura 1974, a volume dedicated to the Hirara topolect versions of proverbs) as well as phonetic notation (Arazato 2003, an excellent, linguistically and ethnographically detailed work, with each of the 115 proverbs provided in multiple regional variants; Nevskiy 1978; 1998, and a digest of Nevskiy's manuscript collection of proverbs in Jarosz 2015, 403-11).

3 Linguistic Characteristics of Miyako Oral Literature

Due to the conventionalized nature of oral literature as opposed to everyday spoken language, the texts representing the former are expected to be stylized and include features distinct from the latter. This applies also to the prose genres, despite their form in essence mimicking the usual spoken language. Folk tales employ markers signaling their beginning, such as *nkja:n-du* 'once upon a time' (lit. 'in the past was it', past-FOC), and ending, employing hearsay (such as *antsi-nu panasi-tsa* 'the story goes like that, they say', like.that-GEN story-HRS) or fixed phrases such as *mmja: usaka* 'the end!'. Riddles follow a mostly fixed structure of question and answer, using a marked word order distinct from that employed in the spoken language (Jarosz 2015, 411-24). Proverbs tend to use parallel phrases, synthetic or antithetic, which gives them a resemblance to poetry, and when they are represented by complex sentences, they are often conditional or sequential (Jarosz 2015, 403-11).

Expectedly, a much higher degree of formalization is encountered in the language of songs (henceforth song language). Apart from the already mentioned (§ 2) characteristics of versification and parallelisms, the Miyako song language represents a special register which affects all levels of language structure: phonology, grammar and vocabulary. In general, this register is more archaic, as it reflects the earlier stages of the language, with forms not encountered anymore in everyday spoken language. This is obviously rooted in the intergenerational oral transmission of pieces which may date hundreds of years back.

Many of the retained forms are longer – e.g. they may not reflect syncopes or apocopes, they may preserve earlier grammaticalization stages, or be overall more analytic – which prevented damaging the metric structure of the piece. At other times this conservativeness may have resulted from a concern toward lyrical faithfulness, approaching the form as the component at least as instrumental to the identity of the piece as the meaning, to the degree that such faithfulness to the form eventually obscured the meaning to the later generations of speakers. This is in particular true of songs with religious connotations, in which the form itself carried ritual and spiritual importance and any alterations risked bringing about the

wrath of gods (cf. Karimata 2011, 155). Such retentions are exemplified by the progressive aspectual forms of verbs in a ritual Karimata song from the genre *ta:bi*, titled *Jama-nu fuɕiradzɨ* (Okuhama 2022), which are more analytic and less grammaticalized than their synthetic spoken language counterparts (cf. § 1): *tate-jor-i* build-PROG-CVB, *tor-i-jor-i* take-CVB-PROG-CVB. Both forms also reveal non-raised Proto-Ryukyuan mid-vowels, another archaic feature of the song register (§ 4.1.). The initial glide /j/ of the progressive auxiliary, typically absent from the modern spoken forms (*uz* ~ *ui* ~ *ul*, with the exception of Tarama, where the initial has undergone fortition j > b: *bul*), is preserved in the song register, highlighting the status of the progressive forms as structures built from two free word forms rather than a host and an affix.

Nevertheless, not all distinct song language forms are archaisms compared to the spoken language. On the contrary, metric considerations permit the type of synthesis that is foreign to the spoken variant of Miyako, and namely a synthesis across morpheme boundaries pertaining to word forms representing incompatible lexical categories and without any motivation for lexicalization. Cf. the following examples from epic songs *Ni:ma-nu ɕu:*[6] (The Lord of Niima; Nevskiy 1978, 108) and *Pstujumja: a:gu* (Song of a daughter-in-law; Nevskiy 1978, 124).

3-1. Naha-minatu-gamj-u:-kur-a uja-minatu-gamj-u-kur-a
Naha-harbor-LIM-ACC-send-VOL father-harbor-LIM-ACC-send-VOL
'I will see you off up to the Naha harbor, up to the main harbor.'

3-2. Ban-tɕu:mu-i naru-tɕu:mu-i tungara
1SG-QUOT.think-IMP RFX-QUOT.think-IMP friend
'Think that this is me, think that this is myself, friend.'

In 3-1., the compound nouns inflected for the combined limitative-accusative cases (*Naha-minatu-gamju:*, *uja-minatu-gamju:*) prosodically host the verb *ukura*, which is the head of the phrase. The initial vowel of *ukura* is deleted (apharesis), a process which would be unusual in the spoken language. Similarly in 3-2., the pronominal forms *ban* and *naru*, followed by the quotative marker *-ti*, blend with the head of the verb phrase, *umui*. Although the vowel /i/ of the quotative marker is apparently deleted as a result of crasis with the verb-initial /u/, the /u/ is lengthened to rhythmically compensate for the loss of a mora

6 For the sake of consistency, all excerpts in Miyako from other sources have been retranscribed to match the IPA-based phonological transcription used by myself.

carried by the /i/ – an occurrence which confirms the treatment of *bantɕu:mui* and *narutɕu:mui* as single word forms.

The result of both 3-1. and 3-2. are combined nominal-verbal word forms which are essentially polysynthetic in their nature, unlike any tendencies observed in spoken Miyako.

Lexically, Miyako songs have a corpus of vocabulary which is only used in this lyrical register. This vocabulary need not be archaic, i.e. unambiguously represent items which had existed in a proto-language but have fallen out of use in modern spoken Miyako. The corpus includes possible loanwords, usually from Okinawan, such as *kumi* 'rice' or Tarama *tsɨbul* 'head', an indication of the prestige of Okinawan in the Ryukyu Kingdom area as the language of authority and art (cf. Gillan 2015), but also perhaps from Japanese (via Okinawan; Gillan 2015, 690-2), like the word *atama* 'head' encountered in songs from the Tarama island.

Expressions apparently reserved for the song register include pronouns such as reflexive Tarama *juf* and Karimata first person *jana* and *ja:ra*; common nouns such as *mata* 'treetop', Hirara *junimuz* 'sandy field', *umizbi:* 'father', or Shimajiri *pudamma* 'mother'; verbs like Tarama *sɨmil* 'to wash one's hands or face'; temporal expressions such as Hirara *junu:rja:* 'at the same time in the future; exactly a year after'; attributes such as *naja* 'famous, renown' and *sɨmu* 'nether, below, southern' (e.g. in reference to the Yaeyama islands), including prefixes like beautifying *ma-* 'real, true, genuine' and honorific *mi-*; poetic references to objects or locations such as *ma-dama muz* 'the Shuri castle' (the king's residence in the capital city), literally 'true-jewel hill'. *Madama muz* itself is likely a loan from the Old Okinawan *omoro* vocabulary (cf. Hokama, Okinawa Kogo Daijiten Henshū Iinkai 1995, 609), and so is the composite beautifying prefix *ma-* (cf. Serafim, Shinzato 2021, 133-5). For a larger list of song register-specific lexicon, cf. Jarosz (2022, 70).

A characteristic element of the song language register, shared with other Ryukyuan lyrical traditions (Gillan 2015, 693), are the repetitive rhythmic fillers called *hayashi*. They can be made up of nonsense syllables (e.g. *jo*, *ju*) or strings of syllables (e.g. *ju:i*, *ho:i*, *hi:ja:ro:gahi:*) inserted at the end of a verse or between verses with an apparent purpose of fixing the rhythmical structure of the piece. Some of the *hayashi* have actual recognizable meaning, and their repetitions serve as a kind of refrain, like *so:dan so: mina-to: pstujumja* 'let's have a talk, daughters-in-law' in aforementioned *Pstujumja: a:gu* (Nevskiy 1978, 122-4), or *ju:-nu pairu* 'may there be prosperity' in *Nagapsɨda* (Nevskiy 1978, 128).

4 Conservative Features

4.1 Phonology

A range of examples from the Miyako song language represent earlier stages of the language, displaying features that are known to be archaic through comparative data and/or the expected directionality of sound change, and which have been lost in the modern spoken language. Table 1 lists such examples together with explanations of the equivalent developments in the spoken register.

Table 1 Conservative phonological features of the Miyako song register

Conservative feature	Example items	Spoken register	Song register
bilabial approximant between two open vowels	'foxtail millet'	a:	awa
retention of short mid-vowels /e/ an /o/*	'to take'	turi (imperative)	tori (imperative)
	'time'	tukja:	toki
	'to build, to stood'	tati	tate
front close vowel /i/ or back close /u/ instead of the so-called fricative vowel /ɨ/ following the sibilants /s/, /ts/, /dz/	'fortification'	gusɨku	guʨiku
	'rope'	tsɨna	tʨuna
	'road'	mtsɨ	mitʨi
retention of strings Cir/Cur, where C represents any obstruent	'white'	ssu	ʨiru
	'to forget'	baʨʨi	basuri
	'child'	ffa	fura
	'pillow'	maffa	makura
diphthong non-coalescence	'Yaeyama'	ja:ma	jaima
retention of the string /ija/	'young girl, maiden'; 'prostitute' (modern spoken language only)	mja:rabi	mijarabi
retention of front close vowel /i/ after /m/	'road'	mtsɨ	mitʨi
	'tatami mat'	tatam	tatami
non-assimilation of front close vowel /i/ before coronal consonants	'where?'	ndza	idza, ida
	'life'	nnutsɨ ~ nutsɨ	inutsɨ

* All examples of non-raised mid-vowels come from the Karimata song representing the ritual *ta:bi* genre, *Jama-nu fuʨiradzɨ* (Okuhama 2022).

Table 1 pertains almost exclusively to the vowel system or vowel-related reductions of syllable structure. This seems to confirm the hypothesis (§ 3) that the archaicity of linguistic features in the song language may often stem from rhythmical and acoustic concerns. The specifics include the retention of post-sibilant Proto-Japonic *i and *u; the retention of Proto-Ryukyuan mid-vowels *e and *o; pre- and post-nasal *i; strings from before the activation of the Sakishima Flap Assimilation Rule (Jarosz 2018), and strings including Proto-Japonic approximants *w and *j in medial, intervocalic positions. Some of these

changes are very old: centralization of post-sibilant *u (post-alveolar obstruent, to be exact) dates to Proto-Ryukyuan (Thorpe 1983), mid-vowel raising was ongoing at least since Proto-Ryukyuan (Thorpe 1983; Jarosz 2018; 2021), and Flap Assimilation was initiated in Proto-Sakishima. Several Miyako songs reflect thus proto-language stages from centuries before, underpinning the argument about the anciency of the Miyako songs and their language.

Table 1 reflects sound changes which apply to all regional varieties of Miyako-Ryukyuan. In contrast, Table 2 shows characteristics which are archaic only in respect to the more innovative topolects such as Hirara, Irabu/Nakachi, Kurima, and Shiokawa-Tarama. Hirara coalesces proto-language diphthongs *au and *ao to the long vowel /o:/, whereas the more conservative topolects such as Nakasuji-Tarama, Ikema and much of Irabu (Nagahama, Sawada and Kuninaka) retain the Proto-Miyako *au. In the song language, however, the diphthong /au/ is uniformly retained also in pieces of Hirara provenance. Retentions of this kind again might stem from acoustic and rhythmic causes – the dynamics of vocalization are different for falling diphthong nuclei as opposed to a single long vowel nuclei, and articulation-wise as well as auditory-wise, it is more challenging to indicate a short: long vowel contrast in singing than in the spoken register.

Table 2 Non-coalescence of *ao and *au in the Hirara song register

Meaning	Conservative modern topolects (e.g. Sawada, Ikema)	Hirara, song register	Hirara, spoken register
'blue'	au	au	o:
'fan'	audzɨ	audzɨ	o:gzɨ
'stick'	bau	bau	bo:
'gate'	dzau	dzau	dzo:
'good'	dzau	dzau	dzo:
'as if'	jau	jau	jo:
'what?'	nau	nau	no:
'pole, rod'	sau	sau	so:

4.2 Morphology

Archaic morphological features of the Miyako literary language mostly apply to morphology of verbs: their inflectional paradigm (§§ 4.2.1., 4.2.2.), syntactically motivated contrasts (§ 4.2.3.), clausal markers or word-formation patterns. As with phonological retentions shown in Table 2, it is frequently the case that while several of these features still exist residually in some conservative modern Miyako topolects, they have fallen out of use in most other topolects except for the literary language.

4.2.1 Proto-Sakishima Suffix *-o

As analyzed in Jarosz (2020b; 2024), Proto-Sakishima had two variant negative suffixes of the vowel verbs: the general Japonic *-e- < Proto-Japonic *əj-a-, *aj-a-, and the Sakishima-specific *-o- < Proto-Japonic *u-a-. Although the latter is rather broadly reflected in Yaeyama, including among others Shika, Taketomi, Kohama, and Hateruma, it is only scarcely attested in modern Miyako, with distribution limited to the southern outskirts of the Core Miyako area, such as Kurima, Gusukube, and Bora.

The song language from outside these *o-retentive regions, however, provides philological evidence that the *-o- negative used to be indeed more widespread and attested throughout all of the Miyako area. Examples include Tarama, with the form *mijudana* (*mi-judana* see-NCVB; spoken *mi:dana*) in the song *Kamnataduru-nu e:gu* (Nevskiy 1978, 134), and Shimajiri *mju:n* (*mj-u:n* EXP-NEG.NPST, spoken *mi:n*) in the song *Koigusuku*/*Kuiguɕiku* recorded by Tajima (Moromi et al. 2008, 226).

4.2.2 Both Proto-Sakishima Passive-Potential Markers

Jarosz (2023; 2024) demonstrates that Proto-Sakishima had two passive-potential markers inherited from Proto-Japonic. One was PJ *-(r)araj-, cf. Old Japanese *-(r)are-*, which initially was likely a specialized passive marker. The other was *-(r)aje-, cf. Old Japanese *-(r)aje-*, likewise an initially specialized inchoative and potential marker. Although individual modern South Ryukyuan topolects such as Kurima-Miyako, Tarama-Miyako and Hateruma-West Yaeyama have reflexes of both, in most Ryukyuan languages only one of the two is retained, and it serves there to encode all passive-potential-related functions.

In the Miyako topolects, the dominant attested type is the descendant of *-(r)aje-.[7] Oral literature, however, proves that also reflexes of *-(r)araj- > *-(r)ari* used to be more widespread in the language; a couple of generations earlier both markers may have been in a common use.

The following examples illustrate such a literary use of *-(r)araj- in present-day topolects which exclusively employ *-(r)aje- > *-(r)ai*. (4-1.) and (4-2.) represent the potential use of *-(r)ari* in a Hirara narrative *Akarjazzagama-nu panasɨ* (as found in Nevskiy 1992); (4-3.) shows the potential use of *-(r)ari* in a Sawada-Irabu proverb (also as

7 This fact in itself is noteworthy, as most modern Japonic topolects, both Mainland and Ryukyuan, tend to retain only the reflex of *-(r)araj-. This includes Miyako's closest relative, the Macro-Yaeyama languages (Jarosz 2023).

recorded in Nevskiy 1992), and (4-4.) illustrates the passive use of *-(r)ari* in a *taugani* song from Ikema (as recorded by Tajima, Moromi et al. 2008, 251).

4-1. sɨni-middz-u jarj-a:mai-gami-du amɕi-rari-z
death-water-ACC COP-CNC-LIM-FOC bathe-POT-CNCL.NPST
'At least I could bathe them in the death water.' (Hirara)

4-2. ningin-nu puri-munu-ja tsɨgzɨnnutsɨ mavv-ari-gumata-nu
people-NOM stupid-thing-ITM longetivy protect-POT-ASM-GEN
'Humans are so stupid! They could have preserved their longevity.' (Hirara)

4-3. tin-nkai-ja pass-a kakir-ari-n
sky-ALL-TOP bridge-TOP cast-POT-NEG.NPST
'You cannot cast a bridge over to the sky.' (Sawada-Irabu)

4-4. jai kisa-uja-nn-a ujagi-rari
thin kisa-lord-DAT-TOP stop-PSV
'We were interrupted by the thin lord Kisa…' (Ikema)

4.2.3 Distinction Between Conclusive and Attributive Forms

It is broadly reported (e.g. Hirayama 1983; Nakama 1992; Uemura 1997) that Miyako-Ryukyuan has lost the original Proto-Japonic distinction between the conclusive and attributive (adnominal) verb forms. In a few modern Miyako topolects such as Kurima, the syntactic distinction between these two positions is reflected morphologically in the vowel verb classes through the conclusive form, marked with *-z* (cf. *psɨtu-nu-du uri-z* person-NOM-FOC descend-CNCL.NPST 'a person descends') contrasting with the bare stem form used adnominally (*uri psɨtu* descend.NPST person 'a person who descends'). This distinction, however, is an apparent Miyako innovation which applies only to a minority of topolects in a limited number of verb classes (cf. Jarosz 2024), while there are virtually no traces[8] of the original Proto-Japonic attributive marker *-o contrasted with conclusive *-u (cf. Hattori 1976; Jarosz 2019).

The situation is once again different in the song language. In several instances of a predicate used adnominally, whether as a noun modifier or as a substantive, the predicate-final vowel reflects

8 There are several exceptions, e.g. the irregular verb 'to do', which in Kurima has both *asi* (reflecting Proto-Japonic conclusive *su) and *asu* (reflecting Proto-Japonic attributive *so) attested in the attributive position (Jarosz 2024).

Proto-Sakishima < Proto-Japonic *o, which in general corresponds in modern Miyako to the close back vowel /u/ with the preceding consonant unchanged, as opposed to conclusive verb forms which, depending on the region and the verb subtype, may end in a final consonant, the vowel /i/ or a diphthong.

The attestations include predominantly vowel verbs found in songs from Tarama, as introduced in Table 3 together with their modern spoken equivalents which combine the conclusive and attributive functions and descend from the Proto-Japonic conclusive (Jarosz 2019).

Table 3 Attestations of a morphologically distinct attributive form in Tarama

Attestation	Song language phrase	Analysis	Modern spoken form
Sjo:gatsɨ-nu e:gu (Nevskiy 1978, 120)	sɨdi-ru sjaku	be.revived-ATR level 'until I am revived'	sɨdil
Sjo:gatsɨ-nu e:gu (Nevskiy 1978, 120)	mujui-ru(/mui-ru) sjaku	burn-ATR level 'until they burn'	muil
Sjo:gatsɨ-nu e:gu (Nevskiy 1978, 120)	mui-ta-ru-ga	bud-PST-ATR-NOM 'that which has budded'	muital
Sjo:gatsɨ-nu e:gu (Nevskiy 1978, 120)	hai-ta-jur-u-ga	grow-PST-PROG-ATR-NOM 'that which has grown'	haital
Bunagama-ga e:gu (Nevskiy 1978, 120)	mi-ru psɨtu	see-ATR person 'the person who sees'	mi:l psɨtu

The Tarama evidence is expanded by examples from other regions, such as Ikema. In an Ikema *taugani* recorded by Nevskiy (1978, 155), the form *umujuru* 'beloved' (*umu-jur-u* think-PROG-ATR) appears in relative clauses *umujuru bunarja* 'beloved woman', *umujuru kanaɕa* 'beloved dear one', and it contrasts with the expected conclusive form *umui-ui* → *umujui/umuju:*.

4.2.4 The Conditional Marker

Modern Miyako topolects display a variety of regional realizations of the conditional marker hosted by verbs in conditional or sequential (temporal) clauses. The attested forms include *-tika:*, *-tiga:*, *-tɕika:*, and *-kka:*, and based on their comparison, they are hypothesized to reflect the Proto-Miyako marker *-tikara(:) (Celik 2020, 220). The only known Miyako variety that retains the form identical to Proto-Miyako is Tarama, alongside the more innovative allomorph *-tika:* (Celik 2020, 220).

In the literary language, this most archaic proto-form is attested abundantly regardless of the region. As noted by Celik (2020, 224), *-tikara* already appears in the song *Yoshima-no oya hashi-age-no*

ayago 四島の親橋積上げのあやご (The Song of Piling Up the Bridge for the Lord of the Four Islands, Japanized title) recorded in the chronicle *Yōsei Kyūki* (§ 2). Some other examples, including the variant palatalized form *-tɕikara*, come from the epic song from Karimata, *Jusima-nu ɕu:* (The Lord of the Four Islands) (Nevskiy 1978, 142; 4-5.), and an untitled *taugani* song from Sawada-Irabu, recorded by Nevskiy in his lexicographic manuscripts (Nevskiy 1992; 4-6.)

4-5.	an	ja-tɕikara	un	ja-tɕikara	Ju-sima-nu-ɕu:
	like.that	COP-CND	like.this	COP-CND	four-island-GEN-lord

'If that is the case, if this is the case, you Lord of the Four Islands' (Karimata)

4-6.	ni:	oi-tja-m	ja-tigara
	root	grow-PST-RLS	COP-CND

'After the roots have grown…' (Sawada)

4.2.5 Formation of Verbalized Adjectives

There are two general verbalization patterns of adjectival roots in Ryukyuan: the *-sa*-pattern, derived from the nominalized adjectival form (*-sa) followed by the stative verb *ari, and the *-ku*-pattern, derived from the adverbial form (*-ku) followed by *ari. The *-sa*-pattern is more pervasive, while the *-ku*-pattern seems limited to Miyako, Hateruma-West Yaeyama, and individual North Ryukyuan topolects in the Amami and Kikai islands (Uemura 1997, 349).

The only modern Miyako topolect which displays the *-sa*-pattern is Tarama. Considering that Tarama is an outlier island separated from the rest of the Miyako-speaking area by a 60-kilometer-long stretch of the sea, as well as its long history of contact with the speakers of Ishigaki-based Yaeyama languages (cf. Makino 1983), it would be reasonable to infer that the Tarama pattern was contact-induced (borrowed from Ishigaki), and Proto-Miyako never had any adjectival derivation strategy other than the *-ku*-pattern. The song language, however, again provides evidence that the *-sa*-pattern was in fact also in use in Miyako outside Tarama. The evidence comes from the already cited songs from the northern region of the Miyako main island, with the pertinent forms applying to the desiderative adjectival root *pusɨ* 'wanting, yearning': Karimata's *Jusima-nu ɕu:* (Nevskiy 1978, 142; 4-7.) and Shimajiri's *Kuigusiku* (4-8., 4-9.), this time in the version recorded by Nevskiy.[9]

9 From an unpublished manuscript stored in the Tenri University Library with the registry number 088イ2-A23 <B700> (Nevskiĭ 192-).

4-7.	nau	pu-ɕa-du	ba-ga
	what	wanting-NMN-FOC	1SG-GEN
	Jaima-n	mja-taz-ga-ju	
	Yaeyama-DAT	come.HON-PST-WHITR-ITM	

'What did you want so that you came to my Yaeyama islands, what are you after?' (Karimata)

4-8.	nau	pu-ɕa-ga	mmjaː-tɕaz-ga
	what	wanting-NMN-WHFOC	come.HON-PST-WHITR

'What did you want so that you came here, what are you after?' (Shimajiri)

4-9.	vva	bata	daks-ɨ	bata	puɕ-an-du
	2SG	stomach	embrace-NPST	stomach	want-VRB.NPST-ITM

'Your belly, the belly you hold, it is what I am after.' (Shimajiri)

(4-7.) and (4-8.) actually show a nominalized form *puɕa* functioning as adjectival predicate in a construction with the honorific motion verb *mmjaːz*, altogether constituting a purposive meaning 'to come wanting to' = 'to come in order to'. While not precisely a verbalized form, by participating in a predicative expression *puɕa* appears to reflect an intermediate stage in the development of the full-fledged verbalized use of the *-sa*-pattern.[10] In contrast, (4-9.) shows a prototypical, tense-marked verbalized adjective *puɕan*.

This evidence allows to postulate that Proto-Miyako did in fact have both *-sa-* and *-ku*-verbalizing strategies, much like modern northern Amami and Kikai topolects, which are reported to synchronically use both *-sa-* and *-ku*-patterns (Uemura 1997, 349, 453).

5 Conclusions

The present chapter argues that a language does not necessarily have to be an established "literary" language to warrant philological research. Despite their overall status of unwritten languages, there is a need to revisit Ryukyuan languages as legitimate subjects of philological study. As illustrated throughout this chapter, the oldest preserved written records of Miyako date back to early eighteenth century, which is a significantly long time span allowing for relevant scrutiny into the synchrony and diachrony of the language. A systematic method needs to be, however, developed in order to accurately decipher Miyako texts written in the *ryūkyūgana*, which apart

10 This purposive construction 'to come wanting to' itself appears to be obsolete; no reports of its existence in modern spoken language has been obtained as of yet.

from the epic songs in the *Yōsei Kyūki* include the late nineteenth- and early twentieth-century documentation by Tajima, Kiyomura and several others (§ 5).

The language of Miyako oral literature merits in-depth study as a research topic in its own right, enabling a study of its stylistic conventions and registers with its specialized lexicon, figurative expressions, formal components such as the *hayashi*, and morphological as well as articulatory characteristics differing from the spoken register (§ 3).

At the same time, the language of the Miyako oral traditions displays numerous archaisms and archaic stylizations, the examples of which introduced in § 4 are a mere tip of the iceberg. A diachronic and comparative analysis of these data provides insights into the history of the language that would otherwise be impossible.

It may be superfluous to observe that the reverberations of such a study are not confined to the Miyako language alone. Even the small sampling of data presented in this chapter supplies strong evidence of the proposed developments of the Ryukyuan and/or Japonic language area (such as the presence of two passive-potential markers of different origins, § 4.2.1, or the erstwhile pervasiveness of the *o-negative marker which corroborates it not being an independent southern Miyako innovation, § 4.2.2), and helps establish facts for which the modern spoken language does not provide enough support (such as the preservation of conclusive and attributive forms of the PJ provenance, § 4.2.3, or the reconstruction of a *-sa*-pattern verbalizer in Proto-Miyako, § 4.2.5).

Some of the archaic features of the Miyako song language phonology reflect proto-language stages like Proto-Sakishima and Proto-Ryukyuan which precede Proto-Miyako by century-long margins (§ 4.1), emphasizing the articulatory conservativeness of the song language as well as the significantly ancient roots of the Miyako songs preserved to date. Also a comparative study of the oral literature-specific vocabulary (§ 3) can reveal new insights about the historical language contact (Miyako-Okinawan or Miyako-(Okinawan-)Japanese) or the reconstructible lexicon of various levels of proto-languages, assuming that the oral literature language preserves some vocabulary that has fallen out of use in the spoken register.

All in all, discoveries from a linguistic study of Miyako oral traditions are expected to affect the perspectives on not just Proto-Miyako, but also the ancestor languages: Proto-Sakishima, Proto-Ryukyuan, possibly even Proto-Kyushu-Ryukyuan and, ultimately, Proto-Japonic.

Abbreviations

1	first person
2	second person
ACC	accusative
ALL	allative
ASM	assumptive
ATR	attributive
CNC	concessive
CNCL	conclusive
CND	conditional
COP	copula
CVB	converb
DAT	dative
EXP	experiential
FOC	focus
GEN	genitive
HON	honorific
HRS	hearsay
IMP	imperative
ITM	interactional marker
LIM	limitative
NCVB	negative converb
NEG	negative
NMN	nominalizer
NOM	nominative
NPST	non-past
POT	potential
PROG	progressive
PST	past
PSV	passive
QUOT	quotative
RFX	reflexive
RLS	realis
RSL	resultative
SG	singular
TOP	topic
VOL	volitional
VRB	verbalizer
WHFOC	wh-focus
WHITR	wh-interrogative
YNITR	-no interrogative

Bibliography

Arazato Hiroshi 新里博 (2003). *Miyako kogen ongi. Jōdai wago-no kaseki: Miyako hōgen- no tebiki. Furoku Miyako hōgen gaisetsu* 宮古古諺音義 上代倭語の化石: 宮古方言の手引き 付録宮古方言概説 (Forms and Meanings of Old Miyako Proverbs. A Guide to the Miyako Dialect, the Fossil of Old Japanese, with an Outline on the Miyako Dialects Appended). Tōkyō: Shibuya Shogen Daigaku Jimukyoku.

Celik Kenan セリック・ケナン (2020). *Minami Ryūkyū Miyakogo-shi* 南琉球宮古語史 (History of Miyako, a South Ryukyuan Language) [PhD Dissertation]. Kyoto: Kyoto University.

Fukuda Atsushi 福田晃 et al. (1991). *Nantō mukashibanashi sōsho 7-8, Gusukube-chō-no mukashibanashi: Okinawa-ken Miyako-gun Gusukube-chō jō/ge* 南島昔話叢書７　城辺町の昔話: 沖縄県宮古郡城辺町　上・下 (Collection of Ryukyuan Folk Tales, Vols 7-8, Folk Tales from Gusukube in Miyako District, Okinawa Prefecture, Parts 1 and 2). Kyoto: Dōmeisha.

Hattori Shirō 服部四郎 (1976). "Ryūkyū hōgen-to hondo hōgen 琉球方言と本土方言 (Ryukyuan Dialects and Mainland Dialects)". Hattori Shirō; Uwano Zendō 上野善道補注 (eds), *Nihon sogo-no saiken* 日本祖語の再建 (Reconstructing Proto-Japanese). Tōkyō: Iwanami Shoten, 45-82.

Hattori Shirō (1978-79). "Nihon sogo-ni tsuite 日本祖語について (About Proto-Japanese)". Hattori Shirō; Uwano Zendō 上野善道補注 (eds), *Nihon sogo-no saiken* 日本祖語の再建 (Reconstructing Proto-Japanese). Tōkyō: Iwanami Shoten, 87-402.

Hattori Shirō (1979a). "Ryūkyū gogen jiten-no kōsō 琉球語源辞典の構想(Designing an Etymological Dictionary of Ryukyuan)". Hattori Shirō; Uwano Zendō 上野善道補注 (eds), *Nihon sogo-no saiken* 日本祖語の再建 (Reconstructing Proto-Japanese). Tōkyō: Iwanami Shoten, 403-38.

Hattori Shirō (1979b). "On'in hōsoku-no reigai 音韻法則の例外 (Exceptions in Phonological Rules)". Hattori Shirō; Uwano Zendō 上野善道補注 (eds), *Nihon sogo-no saiken* 日本祖語の再建 (Reconstructing Proto-Japanese). Tōkyō: Iwanami Shoten, 439-72.

Hirayama Teruo 平山輝男 (author-editor) (1983). *Ryūkyū Miyako shotō hōgen kisogoi-no sōgōteki kenkyū* 琉球宮古諸島方言基礎語彙の総合的研究 (A General Study of the Basic Vocabulary of the Dialect of the Miyako Islands in the Ryukyus). Tōkyō: Ōfūsha.

Hokama Shuzen 外間守善 (1971). *Okinawa-no gengoshi* 沖縄の言語史 (Linguistic History of Okinawa). Tōkyō: Hōsei Daigaku Shuppankyoku.

Hokama Shuzen; Shinzato Kōshō 新里幸昭 (1972). *Miyakojima-no kamiuta* 宮古島の神歌 (Sacred Songs of the Miyako Island). Tōkyō: San'ichi Shobō.

Hokama Shuzen; Shinzato Kōshō (1978). *Nantō kayō taisei 3 Miyakojimahen* 南島歌謡大成 3 宮古篇 (Great Collection of Ryukyuan Songs. Vol. 3, Miyako Island). Tōkyō: Kadokawa Shoten.

Hokama Shuzen (ed.); Okinawa Kogo Daijiten Henshū Iinkai (coord.) 沖縄古語大辞典編集委員会 (1995). *Okinawa kogo daijiten* 沖縄古語大辞典 (A Great Dictionary of Old Okinawan). Tōkyō: Kadokawa Shoten.

Gillan, M. (2015). "Ryukyuan Languages in Ryukyuan Music". Heinrich, P.; Miyara, S.; Shimoji, M. (eds), *Handbook of the Ryukyuan Languages: History, Structure, and Use*. Berlin, Boston: De Gruyter Mouton, 685-702. https://doi.org/10.1515/9781614511151.685.

Inamura Kempu 稲村賢敷 (1962). "Miyako-ni okeru shika-no hattatsu-ni tsuite 宮古における史歌の発達について (Development of Historical Songs in Miyako)". Hokama Shuzen 外間守善 (ed.), *Okinawa bunka ronsō daiyonkan, bungaku/geinōhen* 沖縄文化論叢　第四巻　文学・芸能 (Papers on Okinawan Culture. Vol. 4, Literature and Performing Arts). Tōkyō: Heibonsha, 413-37.

Inamura K. (1977). *Miyakojima Kyūki oyobi shika shūkai* 宮古島旧記並史歌集解 (Old Chronicles of the Miyako Island with a Collection and Commentary of Historical Songs). Tōkyō: Shigensha.

Jarosz, A. (2015). *Nikolay Nevskiy's Miyakoan Dictionary: Reconstruction from the Manuscript and Its Ethnolinguistic Analysis* [PhD Dissertation]. Poznan: Adam Mickiewicz University.

Jarosz, A. (2018). "Reflexes of Proto-Ryukyuan *i and *u in Miyakoan as a Chain Shift". *Lingua Posnaniensis*, LX(2), 99-121. https://doi.org/10.2478/linpo-2018-0014

Jarosz, A. (2019). "Fine Details of a Larger Picture: Proto-Sakishima *au and *ao Diphthongs". *Folia Linguistica Historica*, 40(2), 401-37. https://doi.org/10.1515/flih-2019-0016.

Jarosz, A. (2020a). "Taishō jidai-ni okeru miyakogo-no minwa, Nikorai A. Nefusukii kiroku-no *I:nu panasɨ*/Ei-ni tsuite 大正時代における宮古語の民話―ニコライ・A・ネフスキー記録のI:nu panasɨ「エイについて」― ('The Flatfish Story', a Taishō Era Folk Narrative in the Miyako Language as Recorded by Nikolay A. Nevskiy)". *Ryūkyū-no Hōgen* 琉球の方言, 44, 151-71.

Jarosz, A. (2020b). "Proto-Japonic Verbal Morphology: Insights from the South Ryukyuan Negative Forms of Vowel Verbs". *International Journal of Eurasian Linguistics*, 2, 210-28. https://doi.org/10.1163/25898833-12340032.

Jarosz, A. (2021). "Old Japanese Post-Nasal Non-Back Close Vowels in a Comparative Perspective". *International Journal of Eurasian Linguistics*, 3, 50-82. https://doi.org/10.1163/25898833-12340041.

Jarosz, A. (2022). "Nefusukii saishū-no kōshō shiryō-o tōshite mieru mja:kufutsɨ-no sugata ネフスキー採集の口承資料を通して見えるみゃーくふつの姿 (Miyako Language Seen Through the Oral Literature Materials Collected by Nevskiy)". Miyagawa Kōji 宮川耕次 et al., *Ninupabusɨ. Nikorai A. Nefusukiii tanjō 130-nen/raitō 100-nen kinenbunshū* 子ぬ方星　ニコライ・A・ネフスキー誕生130年・来島100年記念文集 (Northern Star: A Collection of Texts to Commemorate the 130th Anniversary of Nikolay Nevskiy's Birth and the 100th Anniversary of His First Visit in the Miyako Islands). Miyakojima City: Hana View Shuppan, 59-75.

Jarosz, A. (2023). "The Syncretism of Passive and Potential Marking in Japonic Seen Through Modern South Ryukyuan Languages". Dal Corso, E.; Kim, S. (eds), *Endangered Languages of East Asia. Language Obsolescence in China, Japan, Siberia, and Taiwan*. Leiden: Brill, 193-232. https://doi.org/10.1163/9789004523944_008.

Jarosz, A. (2024). *Kurima-Miyako. Descriptive Grammar with a Diachronic Perspective*. Leiden: Brill. Languages of Asia.

Karimata Yoshimasa. 狩俣吉正 (2011). *Okinawa-Miyakojima Karimata minzokushi* 沖縄・宮古島狩俣民族誌 (Folklore Texts from Karimata on the Miyako Island, Okinawa). Naha: Media Ekusupuresu.

Katō Kyūzō 加藤九祚 (2011). *Kampon Ten-no hebi. Nikorai Nefusukii-no shōgai* 完本天の蛇　ニコライ・ネフスキーの生涯 (Snake Out of Heaven, a Biography of Nikolay Nevskiy, Complete Edition). Tōkyō: Kawade Shobō.

Kiyomura Kōnin 慶世村恒任 [1927] (2008). *Shimpan Miyako shiden* 新版 宮古史伝 (History and Folklore of Miyako, New Edition). Tōkyō: Fuzambō Intānashonaru.

Linguapax Asia (2019). *Pride, Love and Community – Linguapax Asia Video Project*. https://youtu.be/sSeLL0J1_pw.

Makino Kiyoshi 牧野清 (1983). "Taramada 多良間田 (The Tarama Field)". Okinawa Daihyakka Jiten Kankō Jimukyoku 沖縄大百科事典刊行事務局編 (ed.), *Okinawa Daihyakka Jiten chūkan, ke–to* 沖縄大百科事典中巻、ケ～ト (Okinawa Encyclopedia, vol. 2). Naha: Okinawa Taimususha, 732.

Miyake, M.H. (2013). *Old Japanese: A Phonetic Reconstruction*. 2nd edition. London: Routledge. https://doi.org/10.4324/9780203510728.

Moromi Nana 諸見菜々 et al. (2008). "Honkoku Tajima Risaburō Miyakojima-no Uta (Ryūkyū Daigaku Fuzoku Toshokan Iha Fuyū Bunko shozō) 翻刻田島利三郎『宮古島の歌』(琉球大学付属図書館伊波普猷文庫所蔵) (Transcript of Risaburō Tajima's Miyakojima-no Uta, Stored in the Fuyū Iha Archive of the Ryukyu University Library)". Tamaki Masami 玉城政美, *Okinawa-ken Miyako shotō-ni okeru girei kayō-no shūshū/kenkyū-to dētabēsuka* 沖縄県宮古諸島における儀礼歌謡の収集・研究とデータベース化 (Collection, Research and Archiving of Ritual Songs of the Miyako Islands in Okinawa Prefecture). Nishihara: Ryūkyū Daigaku Hōbungakubu, 171-280.

Nakama Mitsunari 名嘉真三成 (1992). *Ryūkyū hōgen-no kosō* 琉球方言の古層 (Substrata of Ryukyuan Dialects). Tōkyō: Daiichi Shobō.

Nevskiy, N.A.; Невский, Н.А. (1978). *Фольклор островов Мияко. Fol'klor ostrovov Miyako* (Folklore of the Miyako Islands). Москва: Издательство "Наука".

Nevskiy, N.A.; Невский, Н.А. [1921-28] (1992). *Матерьялы для изучения говора Островов Мияко: словарь. Materyaly dlya izucheniya govora Ostrovov Miyako: slovar'.* (Materials for the Study of the Language of the Miyako Islands: A Dictionary). Photocopy of the manuscript stored in the Institute of Oriental Manuscripts, Russian Academy of Science, Saint Petersburg. Tōkyō: Waseda University Library.

Nevskiy, N. Невский, Н. (1996). *A collection of articles related to Nevskiy's Miyakoan research.* Gromkovskaya, L.L. Громковская, Л.Л. (ed.), *Николай Невский Переводы, исследования, материалы к биографии. Петербургское востоковедение. На steklakh vechnosti. Nikolay Nevskiy. Perevody, issledovaniya, materialy k biografii* (The windows of eternity. Nikolay Nevskiy. Translations, studies, biographical sources). *St. Petersburg Journal of Oriental Studies*, 8, 239-560.

Nevskiy, N.A. (1998).*Miyako-no fōkuroa* 宮古のフォークロア (Miyako Folklore). Edited by L. Gromkovskaya, リヂア・グロムコフスカヤ. Translated by Karimata Shigehisa et al. Tōkyō: Sunagawa Shobō. The Japanese edition of Nevskiy 1978.

Nevskiy, N.; Oka Masao 岡正雄 (eds) (1971). *Tsuki-to fushi* 月と不死 (The Moon and Immortality). Tōkyō: Heibonsha.

Ogawa, S. (2015). "Orthography Development". Heinrich, P.; Miyara, S.; Shimoji, M. (eds), *Handbook of the Ryukyuan Languages: History, Structure, and Use*. Berlin; Boston: De Gruyter Mouton, 575-89. https://doi.org/10.1515/9781614511151.575.

Okinawa-ken Bunka Kankō Supōtsu-bu 沖縄県文化観光スポーツ部 (2022). *Okinawa-ken-ni okeru "shimakutuba"-no hyōki-ni tsuite* 沖縄県における「しまくとぅば」の表記について (Writing Conventions of the Local Varieties in Okinawa Prefecture). Naha: Okinawa Prefectural Government.

Okuhama Sachiko 奥濱幸子 (2022). "*Miyako Hōgen Nōto*-no saikōsei jiten (A. Yaroshu)-o katsuyō shite ta:bi futsɨ *Yama-no fushiradzu*-o kaidoku suru" 『宮古方言ノート』の再構成辞典 (A.ヤロシュ) を活用して高べ口「ヤマノフシラヅ」を解読する (Decipherment of a Song from the Genre *ta:bi futsɨ*, Jama-nu fuɕiradzɨ, Using the Reconstructed Version of *Miyako Hōgen Nōto* by A. Jarosz)". Miyagawa Kōji 宮川耕次 et al., *Ninupabusɨ. Nikorai A. Nefusukiii tanjō 130-nen/raitō 100-nen kinenbunshū* 子ぬ方星　ニコライ・A・ネフスキー誕生130年・来島100年記念文集 (Northern Star: A Collection of Texts to Commemorate the 130th Anniversary of Nikolay Nevskiy's Birth and the 100th Anniversary of His First Visit in the Miyako Islands). Miyakojima City: Hana View Shuppan, 169-83.

Serafim, L.; Shinzato, R. (2021). *The Language of the Old-Okinawan Omoro Sōshi. Reference Grammar With Textual Selections*. Leiden; Boston: Brill. https://doi.org/10.1163/9789004414686.

Shinzato Kōshō 新里幸昭 (2005). *Miyako kayō-no kenkyū* 宮古歌謡の研究 (Research on the Miyako Folk Songs). Minami Uebaru: Bunshin Insatsu.

Sunagawa Gensei 砂川玄正 (1983). "Miyakojima Kyūki 宮古島旧記 (Old Chronicles of the Miyako Island)". Okinawa Daihyakka Jiten Kankō Jimukyoku 沖縄大百科事典刊行事務局編 (ed.), *Okinawa Daihyakka Jiten gekan, na–n* 沖縄大百科事典下巻、ナ〜ン (Okinawa Encyclopedia. Vol. 3). Naha: Okinawa Taimususha, 591-2.

Takubo, Y. (2021). "Morphophonemics of Ikema Miyakoan". Kupchik J.; de la Fuente, J.A.A.; Miyake, M.H. (eds), *Studies in Asian Historical Linguistics, Philology and Beyond. Festschrift Presented to Alexander V. Vovin in Honor of His 60th Birthday*. Leiden; Boston: Brill, 65-79. https://doi.org/10.1163/9789004448568_007.

Tartarini, S. しるびあたるたりーに (2009). "Koryūkyū-ni okeru moji-no dōnyū/shiyō-ni tsuite 古琉球における文字の導入・使用について (On the Introduction of Writing and Its Use in the Ryukyu Islands)". *Ōbirin Ronshū* 桜美林論集, 36, 37-49.

Thorpe, M.L. (1983). *Ryukyuan Language History* [PhD Dissertation]. Los Angeles: University of Southern California.

Uemura Yukio 上村幸雄 (1997). "Ryūkyū rettō-no gengo. Sōsetsu 琉球列島の言語　総説 (The Language of the Ryukyu Archipelago. General Remarks)". Kamei Takashi 亀井孝; Kōno Rokurō 河野六郎; Chino Ei'ichi 千野栄一 (eds), *Gengogaku daijiten serekushon Nihon Rettō-no gengo* 言語学大辞典セレクション　日本列島の言語 (Selection of Great Dictionaries of Linguistics: Languages of the Japanese Archipelago). Tōkyō: Sanseidō, 311-53.

Vovin, A. (2020). *A Descriptive and Comparative Grammar of Western Old Japanese. Revised, Updated and Enlarged Second Edition*, vols. 1-2. Leiden: Brill.

Yoshimura Gentoku 吉村玄得 (1974). *Okinawa Miyako kotowaza zenshū* 沖縄宮古ことわざ全集 (Complete Collection of Miyako Proverbs). Urasoe: Taigā Insatsu.

Philological and Linguistic Analysis Working Together
Exploring Language Diversity in Japan
edited by Paolo Calvetti

Adopting a Philological Approach Toward *Chishi*

Étienne Baudel
CRLAO - École des hautes études en sciences sociales, France

Abstract Hachijō, a small endangered Japonic variety, is fortunate enough to have been documented since the Edo period, and to have been abundantly described and commented ever since. As a matter of fact, Hachijō linguistic material can be found in about two dozen documents from between 1746 and 1858, which provide valuable lexical, grammatical and phonological insights into early Hachijō. This article proposes a preliminary study of those sources, with a summary of the existing literature about them, and an introduction to the philological questions they raise.

Keywords Japanese. Philology. Dialectology. Historical Linguistics. Palaeography. Codicology.

Edizioni Ca'Foscari

Ca' Foscari Japanese Studies 24
Linguistics and Language Education 4
e-ISSN 2610-9395 | ISSN 2610-8976
ISBN [ebook] 978-88-6969-845-3 | ISBN [print] 978-88-6969-846-0

Peer review | Open access
Submitted 2023-09-22 | Accepted 2024-03-04 | Published 2024-08-28

DOI 10.30687/978-88-6969-845-3/002

1 Introduction

1.1 What is Hachijō?

Hachijō is a small Japonic variety,[1] traditionally spoken on three islands which are located in the south of the Izu 伊豆 archipelago, roughly 300 km south of Tōkyō (namely, from north to south: Hachijō-jima 八丈島, Kojima – or Hachijō-kojima – (八丈) 小島 and Aogashima 青ヶ島). While Hachijō is, in English and in Japanese alike, named after the most populated of those three islands, its native speakers usually simply call it *Shima-kotoba* 島言葉, lit. 'island words', or, more rarely (and mostly only among younger speakers), *Shima-ben* 島弁, lit. 'island dialect'.

There are eight recorded local varieties (or topolects) of Hachijō, each of them corresponding to a different village, and having some lexical and phonological peculiarities. Five of them are located on Hachijō-jima (counter-clockwise: *Mitsune* 三根, *Ōkagō* 大賀郷, *Kashitate* 樫立, *Nakanogō* 中之郷, *Sueyoshi* 末吉), two on Kojima (*Toriuchi* 鳥打 in the north and *Utsuki* 宇津木 in the south), and only one on Aogashima. Regarding the pre-modern sources of Hachijō, it is likely to assume that all attested varieties in that period are from Hachijō island, more specifically from the villages of Ōkagō, Mitsune and Sueyoshi.

The position of Hachijō within the Japonic family is still a matter of debate, but it is often considered to be a descendant of Eastern Old Japanese, either alongside other Eastern Japanese varieties (De Boer 2020, 52), or (most commonly), as its sole descendant (see for instance Kupchik 2011, 7). However, other specialists also consider that there are also several shared innovations between Central Old Japanese and Hachijō, which makes its classification unclear (see, for instance, Hirako, Pellard 2013, 52); and others even estimate that the majority of the linguistic data might actually speak against that traditional classification (Baudel, forthcoming).

1 It should be noted that until recently, Hachijō was almost universally considered to be a dialect of Japanese, rather than an independent language (and subsequently called *Hachijō-hōgen* 八丈方言, 'Hachijō dialect'). This view still appears to be dominant among Hachijō native speakers. Then, in the early 2000s, Japanese sociolinguists started treating Hachijō as a separate language, preferring the term *Hachijō-go* 八丈語 in Japanese (see, for instance Hashimoto, Long 2003; Long 2004; Asahi et al. 2004). This terminology gradually became dominant among linguists, both inside and outside Japan. However, it must be noted that, to the best of my knowledge, no objective criteria was really brought up to justify the use of one term rather than the other, since the putative absence of mutual intelligibility between Hachijō and the dialect of Tōkyō (mentioned, for instance, by Tsuzuku 1955, 37; or more recently, by Iannucci 2019, 100-4) was, in fact, never assessed scientifically. For that reason, I prefer to avoid using both 'language' and 'dialect', which are rather connoted terms, and prefer to use the broader term 'variety' instead.

Finally, it is unknown how many speakers of Hachijō are currently left, but while the number of passive speakers might very well be of a few thousand, the number of fluent speakers of Hachijō is more likely to be somewhere between a few hundred and a thousand people,[2] most of them being elderly. Because of this, Hachijō was included in UNESCO's *Atlas of the World's Languages in Danger* (Moseley 2009), alongside seven other languages of Japan, and granted the status of 'definitely endangered'.

1.2 The Early Sources of Hachijō

Although the south Izu islands were, for centuries, difficult to access from the mainland because of the strong Kuroshio 黒潮 stream, they played an important role in the court religion of ancient Japan, and were therefore already subject to frequent imperial petitions in the Heian period (794-1185) (Alaszewska 2018). However, very few sources about those islands remain from that period, and, to the best of my knowledge, the islands are only mentioned in shrine inventories and in recordings of fortune-teller visits to the imperial court.[3] In Hachijō-jima on the other hand, some annals were preserved (called *Hachijō-jima nendaiki* 八丈島年代記 'Annals of Hachijō island') covering the years 1335 to 1635. However, they barely contain any local word either, since they only focus on recording natural disasters, famines, and land management. Thus, we can probably safely assume that there is no substantial source for Hachijō that would correspond to the period of Old Japanese (roughly sixth to eighth century) or to that of Middle Japanese (roughly ninth to sixteenth century), and that all of its attestations correspond to the Early Modern Japanese period (seventeenth-nineteenth century), at the earliest.

In 1600 the contact of the South Izu islands with the mainland was dramatically increased, when the islands were placed under direct military rule in order to be used as a banishment territory. Thus, the

2 As pointed out by Iannucci (2019, 9), the number of 8,000 speakers provided by the 2009 version of the Unesco's *Atlas of the World's Languages in Danger* (which can still be found on the *ELP* page 'Hachijō') is necessarily largely overestimated, since it exceeds the whole population of Hachijō-jima, where the vast majority of remaining speakers reside. Thus, while more than 20 years ago, Kaneda (2001, 45) estimated that there might be at most 3,000 thousand speakers left, in 2019, Iannucci (2019, 9) estimated that the number of "truly fluent speakers" was more likely "in the low hundreds". For a more specific discussion on the phenomenon of 'language erosion' in the South Izu, and the precise unfolding of the language shift to Tōkyō Japanese in those islands, cf. Baudel, forthcoming.

3 The only notable exception is *The Tale of Hōgen* (*Hōgen Monogatari* 保元物語) (c. 1320), in which Minamoto no Tametomo is said to have ended his life in the South Izu islands. Of course, no real description of the islands or of their dialect appears in this fictitious work, but the image it conveys does reveal the peculiar place that those islands had in the Japanese culture (cf. Mollard 2021, 38-9).

occasional petition of officials to the South Izu became a yearly visit and, during roughly 260 years, over 1,800 convicts, mostly of relatively high social status,[4] were sent to the South Izu islands. This policy made the South Izu islands appear as a remote and ominous land in the capital, and reactivated older myths and legends about them[5] In parallel, the Edo period (1603-1868) saw a boom in scientific and cultural productions in mainland Japan, and especially the birth of a new genre called 'geographical descriptions' (*chishi* 地誌). Namely, many documents on remote territories were written and published, often including richly illustrated geographical, zoological or botanical descriptions, and sometimes local legends or words, which form invaluable sources for the dialects and minority languages of Japan. This trend applied especially to South Izu islands, and several dozens of geographical descriptions were copied or published about those three islands between 1746 and 1868, most being either commissioned from shogunate officials, or written by convicts. A comprehensive inventory of those sources, and a review the existing literature about them will follow this introduction.

1.3 Methodology and Aim of This Study

This article is a preliminary research on the pre-modern sources for Hachijō, namely all documents written before the Meiji Restoration (1868). Therefore, it can be considered an attempt to adopt a philological approach to the study of a minority language, which appears to be quite uncommon in the field of Japanese studies. Thus, this article is based on a review of the existing literature, on a comparison of contemporary editions with available manuscripts, on a critical compilation of all the attested data, and on a comparison of that data with contemporary language sources. In order to do so, a first section will present a comprehensive inventory of those documents and of the existing literature about them, and a second will present some of the insights that those sources can provide about Hachijō, and some of the questions that remain open for further research.

4 For a more detailed presentation of the exiles to the South Izu, cf. Kasai, Yoshida 1964.

5 For example, see the description of Bakin's work by Mollard (2021, 28-33, 38-9).

2 Sources, Editions and Compilation

2.1 Comprehensive Inventory

As stated earlier, there are many pre-modern documents about the South Izu, presenting a great variety of form and content (including annals, chronicles, diaries, petition reports, geographic descriptions, maps, calendars, etc.). However, at this stage of my research, I found that only about twenty of those documents appear to contain substantial[6] linguistic data, namely:[7]

1. 1746 *Hachijō-jima-ki* 八丈島記 'Records of Hachijō island', written by Ike Norimitsu 池則満.
2. 1747 *Hachijō-jima doryaku-shi* 八丈島土畧誌 'Outline document of land properties of Hachijō island', by Furukawa Koshōken 古川古松軒.
3. 1781 *Izu kaitō fudoki* 伊豆海島風土記 'Chronicles of the ocean islands of Izu', by Satō Yukinobu 佐藤行信.
4. 1791 *Nanpō kaitō-shi* 南方海島志 'Documents from the southern sea islands', by Akiyama Funan 秋山富南.
5. 1796 *Shichitō nikki* 七島日記 'Diary from the seven islands', by Kodera Ōsai 小寺応斎.
6. 1796 *Kaitō zatsuwa* 廻島雑話 'Short word about an island trip', by Ōta Hikosuke 太田彦助.
7. 1797 *Hachijō hikki* 八丈島筆記 'Notes about Hachijō island', by Furukawa Koshōken 古川古松軒.
8. 1800 *Izu shisho* 豆州志稿 'Manuscript of Izu', by Akiyama Funan 秋山富南.
9. 1801 *Izu shichitō fudo sairan* 伊豆七島風土細覧 'Detailed observation of the Chronicles of the seven Izu island', by Mishima Masahide 三島正英.
10. 1802 *En'ō kōgo* 園翁交語 'Conversations of old gardeners', by Takahashi Yoichi 高橋與一.
11. undated (c. 1810) *Hachijō-jima runin no hanashi* 八丈島流人之咄 'The history of Hachijō island exiles', by Matsuda Heiemon 松田兵右衛門.
12. undated and anonymous (c.1810) *Hachijō-jima monogatari* 八丈島物語 'Tales from Hachijō island'.
13. 1811 *Hachijō-shi* 八丈誌 'Hachijō documents', by Ōhara Masanori 大原正矩.

6 That is, more than just a few isolated words.

7 It must be noted that most of those works have numerous alternative titles and versions, and their authors alternative pen names. Thus, the names and titles provided here are the ones most commonly found.

14. 1811 *Hachijō kiriko-ori* 八丈裁衣織 'The weaving of Hachijō cut clothes', by Hattori Yoshitaka 服部義高.
15. c. 1811 *Ichiwa ichigen* 一話一言 'Small texts with small comments', by Ōta Nanpo 大田南畝.
16. 1814 anonymous *Hachijō-jima daigai-chō* 八丈島大概帳 'Large account of Hachijō island'.
17. 1839 *Asahi gyakutō-ki* 朝日逆島記 'Asahi's records of topsy-turvy islands', by Sawara no Kisaburō 佐原喜三郎.
18. undated (c. 1840) anonymous *Sangashima keihō dai-hikan* 三ヶ島刑法大秘鑑 'The Large secret account on the penal code of three islands'.
19. 1848 *Yatake no nezamegusa* 八丈の寝覚草 'The waking up of Hachijō grass', by Kakusō Kisan 鶴窓帰山.
20. 1858 *Hachijō jikki* 八丈實記 'The true account of Hachijō', by Kond Tomizo 近藤富蔵.

Those works mostly belong to the genre of 'geographical descriptions' (*chishi*), and subsequently include a lot of botanical and animal lexicon, often coming with rich illustrations. More than half of those documents also include one or several (in the case of Kondō 1858) wordlists of the local variety. In total, not only isolated words, but also, several sentences and two full texts are attested: namely, a letter and a short dialogue, both of them being provided by the 1848 *Yatake no nezamegusa*.

Overall, this can be said to be quite a lot of sources for a minority language of Japan at that time. In addition, the fact that about 15 of those works were written within the same 30 years (1781-1811) shows that there seems to have been some kind of trend in Edo around that time, in the interest for that remote territory.

Regarding the origin and target audience of those works, we can say that they were mostly written, copied and sold in Edo, and can safely be said to have been made for the elite people in the capital, rather than for the local audience of the Izu islands. It can also be noted that they were all written by educated male authors, who were either:

- shogunate officials (Satō, Kodera, Hattori, Ōta Hikosuke),
- convicts who were banished to the island (Mishima, Matsuda, Ōhara, Sawara, Kakusō, Kondō),
- scientists and learned men (Akiyama, Furukawa, Ōta Nanpo).

Only one of those authors appears to have been an islander: Takahashi Yoichi, who was a silk trader form the village of Mitsune.

Finally, it must be noted that the manuscripts of most of those documents have been digitalised and made available online (the only exceptions being Ōta Hikosuke, Mishima, Sawara and Kondō). Several of them also have been edited in modern versions, either in

compilations of *chishi* about the Izu islands (most notably Takakura 1968 or Kanayama 1976), or as separate books. However, several of those works remain without any modern edition to this day, namely Ōhara, Hattori, the *Hachijō-jima runin no hanashi*, the *Hachijō-jima monogatari*, and the *Hachijō-jima daigai-chō*.

More importantly, as far as I know, these sources have never been properly compiled together or systematically criticised, especially regarding their codex variants, possible copy mistakes, and information sources. These questions would require further investigation.

2.2 Existing Literature

In general, studies about the Edo sources for Japanese dialects and varieties can be said to be surprisingly rare, especially when compared with the amount of synchronic approaches: in the case of Hachijō, out of more than 200 existing references about the language, only a handful include pre-modern sources to their consideration (such as Tamura 1928; Motoyama 1934 or Yoshimachi 1951), and the only ones that do only focus on one or two pre-modern documents. The only exception in this regard is the comprehensive study published by NINJAL in 1950, which I will now briefly comment.

2.2.1 The 1950 NINJAL Study

In 1949, the newly founded National Institute for Japanese Language and Linguistics (NINJAL) dedicated its first ever full-scale study to Hachijō. This choice is explained in the introduction of this work (3) to have been motivated by four elements:

1. the South Izu were, at that time, undergoing language shift from the local variety to the standard language from Tōkyō (which is the phenomenon that NINJAL originally wanted to document)
2. Hachijō is different enough from the dialect of Tōkyō to make that language shift easy to observe
3. the classification of Hachijō is unclear, which made its description important
4. Hachijō is attested in sources from the Edo period, which makes it possible to document its recent evolutions

In other words (as stated in point 4), the existence of pre-modern sources is not coincidental with Hachijō being the first ever language studied by NINJAL, but, quite in the opposite, those early attestations were an element that motivated its description. Thus, in total, 25 pre-modern sources for Hachijō are listed and described in this study (272-86), and the lexical content of 20 of those 25 sources is

included in the provided lexicon (321-413), for a total of almost 1,400 entries. However, it must be noted that among those 25 references, some appear to be alternative titles of the same work rather than different documents, while others barely contain any linguistic content. For those reasons, some of those references were not included above in our list of the 20 major pre-modern sources.

More generally, this study is also based on more than 40 modern sources (listed at pages 303-18) on the language, as well as on a 10-day field study conducted by 10 NINJAL members in Hachijō-jima. The methods, circumstances and people involved in this investigation are extremely thoroughly detailed, and several important results are provided about the differentiated use of Hachijō and Tōkyō Japanese, the inner diversity of Hachijō, its phonological, grammatical and lexical peculiarities, notable cultural features of its local community, etc. For all these reasons, this study can be considered one of the most important milestones in the study of Hachijō, alongside its first English description (Dickins, Satow 1878), its first Japanese description (Hoshina 1900), or, more recently, its comprehensive description by Kaneda (2001).

Finally, the question of the continuity between the pre-modern attestations and the contemporary language is also assessed in the study. Namely, a quantitative estimation was conducted comparing the wordlist provided by Ōta (c. 1811) with data gathered during the Institute's fieldwork. The study concludes (221) that in total, 78% of the Hachijō words provided in Ōta's wordlist were still in use in 1949, that 29% were altered in some ways, and that only the remaining 7% had completely fallen out of use, although there were strong differences between generations and villages regarding the preservation of lexicon (222). I will develop further about this estimation in the following paragraph, mostly in order to explain why I disagree with these numbers (cf. *infra*).

2.2.2 Limits of the 1950 NINJAL Study

Given the fact that this question is not at all the focus of this study, the 1950 NINJAL study is not exempt of some weaknesses in its consideration of Hachijō pre-modern sources. Most notably, the data compiled from those sources is 'sorted', but not 'filtered', which means that every different spelling is shown in a separate entry, without any cross-reference or comparison between those entries. This is unfortunate, since allography is particularly strong in the Edo period (especially when writing down a non-standardised oral language), and can give some valuable insights on the pronunciation of some phonemes at that time. For instance, one given lexeme such as *shicchou* (seventh son) (cognate of Tōkyō Japanese given name *Shichirou* 七郎) can be

found in no less than 8 different entries in the study's index: ヒッチョー <*hicchō*>, ヒツテウ <*hitsuteu*>, ジッテウ <*jitteu*> (with probably erroneous *dakuten*), シッチャウ <*shicchau*>, シッチョー <*shicchō*>, シッチョウ <*shicchou*>, シッテョウ <*shitteyou*> (with an erroneous テ for チ) and シッテフ <*shittefu*>. In total, about a third of the 4,700 entries provided in the study's index can be considered to be spelling variants, rather than different lexemes.

In addition, as shown with the example of *shicchou*, probable copy mistakes (whether in the Hachijō headwords or in the provided translation in classical Japanese) are also quite common in texts from that period, and, while they are valuable to establish codical relations, presenting them as separate entries without any commentary can be misleading. For instance, several entries such as *chifuchi*, *chifuno*, *chizuna* are all presented as meaning 'dandelion' (in Japanese *tanpopo* 蒲公英) in pre-modern sources of Hachijō, such as the *Oki no Kojima-ki* (a variant name of the *Izu kaitō fudoki*). However, those forms are unattested in contemporary sources of Hachijō (such as Kobayashi 1984 or Asanuma 1999), which instead have only one etymologically transparent form: *chibuna* (analysable as 'latex plant': *chibu* + *na*). Given the fact that no regular sound change or lexical derivation could explain those three forms in Hachijō, the most reasonable explanation is probably that they are miscopies from <*chibuna*> (or, rather, from <*chifuna*>; that is, *chibuna* written without *dakuten*), caused by the resemblance between the 3 *katakana* <ナ>, <チ> and <ノ>. Such miscopies appear to be especially common in some pre-modern sources, such as the *Oki no Kojima-ki* and the *Izu shichitō fudo sairan* (which could possibly indicate that they were copied by someone who did not speak Hachijō). It is also more frequent for some words than others: for instance, the Hachijō word *hikkan* (Japanese *hikan* 被官, 'servant') appears with several aberrant spellings, such as ヒヅリツン <*hizurishin*> or ヒヅリソン <*hizurison*>, in addition to the expected ヒツクワン <*hitsukuwan*>. This might possibly indicate that those words were rather infrequent or difficult to analyse, compared with other lexical items.

The blindness of the 1950 NINJAL study towards allography and misspellings also has a strong influence on its quantitative estimation of how much old Hachijō remained in use in contemporary times. For instance, while the study concludes that a word from Ōta's wordlist like トウフ <*toufu*> (grotto), was not used any more in 1949, I consider that this form is more likely a miscopy from the widely attested Hachijō word トウラ *toura* ~ *doura* (with <ラ> miscopied as <フ>). The same thing is true for instance of ニロノッテ <*nironotte*> (together), which is probably miscopied (with <モ> misread as <ニ>) of モロノッテ *moronotte* ('together', which is attested, for instance in Kobayashi 1984); for フングミ <*fungumi*> (loincloth), probably miscopied from the widely attested word フンドシ *fundoshi* ('loincloth'

with <ドシ> miscopied as <グミ>), or of the expression フッケフッケ <*fukke-fukke*> (light rain), which I believe, is miscopied from the verbal form フッテフッテ *futte-futte* (falling lightly) (with <テ> misread as <ケ>). Besides, in many other cases, the NINJAL study also considers words that are attested in Ōta's wordlist to have fallen out of use merely because of a missing *dakuten*, while those words are safely attested in many contemporary sources. This can be seen for instance in *ogoru* ('be noisy'; written as <*wokoru*>), *mabaru* ('to watch'; written as <*maharu*>), *tabi* ('menstruation hut'; written as <*tahi*>), and in several other words.[8] This mistake is quite surprising, given the fact that some of those words are extremely common in Hachijō, and given the fact that the use of the *dakuten* is known not to have been systematic in the Edo period.

Finally, and perhaps quite surprisingly, the 1950 NINJAL study also reports that several Hachijō words from this wordlist have fallen out of use, while those are (sometimes abundantly) attested in other contemporary sources. This is the case, for instance, of the words *nusutama* ('thief'; attested, for instance, in Asanuma 1999), *itari* (leisure house), *kubona* (spider) (both found in Isozaki 1977), *nosu* (to eat a lot) (Ogawa 1958), *tomi-sagari* (good rain) (Kobayashi 1984), or *kūrou* (sixth daughter) (attested, for instance, in Isozaki 1977; Kobayashi 1984; Yamada 2010). Thus, we have to assume that either the sample of 20 people chosen to conduct the fieldwork study was too small to represent the whole of the speakers' community, or that there was some kind of flaw in the data collection.

In any case, because of all the aforementioned elements, I think we can safely consider that the proportion of Hachijō words that were preserved between Ōta's wordlist and the contemporary language is more likely to be higher than what the NINJAL study suggests. According to my personal recount, I consider that out of 207 entries in Ōta's wordlist, only 9 clearly attested words (or 4% of the total) appear to have fallen out of use before 1949, while 9 entries remain difficult to analyse. In other words, a total of 189 words (or 92% of the wordlist) appear to have been partly or completely preserved between 1820 and 1949, which shows an even greater continuity between pre-modern and contemporary Hachijō that the study suggested.

However, given the fact that the author of this wordlist, Ōta Nanpo, was neither native from the islands, neither even traveled there

8 Namely: *tega* (hoe) (written as <*teka*>), *tego* (third daughter) (written as <*teko*>), *jīrou* (fifth daughter) (written as <*chiirou*>), *kedouzu* (sewer) (written as <*kedousu*>), *yobi* (belt) (written as <*yohi*>), *donza* (rag) (written as <*tonsa*>), *yagi* (fishing pole) (written as <*yaki*>), *jougi* (bowl) (written as <*shiyaugi*>), *kendon* (gargling bowl) (written as <*kenton*>), and *daisan* (modern Hachijō *de:shan* ~ *dya:shan*) (3 pm) (written as <*taisan*>).

in person, given the fact that the source(s) he used are unknown, and, more importantly, given the many miscopies we saw in his wordlist, I think that this wordlist should probably not be regarded as the most representative pre-modern source of Hachijō. Instead, I think it would be most valuable to compare modern linguistic data to compiled data from all pre-modern sources, in order to get a clearer idea of the lexical, phonological and grammatical retentions and innovations that occurred during the last centuries. Thus, we can now, in a final section, introduce to some preliminary findings in this topic, and to prospective research questions that will need to be investigated about the pre-modern sources of Hachijō.

3 Preliminary Findings and Prospective Research Topics

3.1 Lexicon

As stated earlier, the first task of this research on the pre-modern sources of Hachijō was to compile all attested lexemes into a comprehensive table, to sort those lexemes by removing double counts, and, finally, to compare them with the more than 8,000 Hachijō lexemes gathered from contemporary sources. In total, when adding up lexical content from all 20 aforementioned pre-modern sources, the pre-modern Hachijō lexemes add up to roughly 1,000 items. This number shows once again how well attested this variety was in the Edo period, when compared with other minority languages of Japan. Besides, these lexemes can be noted to include a lot of highly iconic Hachijō words (such as *nyoko* 'first daughter', *menarabe* 'girl', *ojari yare* 'welcome', etc.); but also many words that are unattested in contemporary sources, most of which being of obscure composition, or occurring only once.

Due to the difficulty of analysing a lot of those forms, it is hard to get a precise account of the percentage of pre-modern lemmas that were preserved in the contemporary language. According to my current counting, it can, at this stage, only be roughly estimated that roughly 70% of pre-modern words can safely be considered to have been fully or partly[9] preserved in the contemporary language, and that less than 10% of pre-modern Hachijō words can probably safely be considered lost.[10] In other words, I consider that, at this stage,

9 As a matter of fact, some roots were preserved, but the composition of the independent lemma was altered, for instance *kona* (silkworm) and *kasuru* (to forget) are both well attested in pre-modern Hachijō. However, in the contemporary language, they are obligatorily, respectively, suffixed (*kona-sama*), and prefixed (*hik-kasuru*).

10 An example of word that can safely be considered lost, is *takadara* (bamboo basket'. As a matter of fact, this words respects two different criteria, namely:

roughly 20% of pre-modern Hachijō lexemes are not well attested or analysable enough to be considered safe attestations. As I hope to have shown with the example of Ōta's wordlist, it is likely that a sound part of those 20% obscure terms are actually well attested lexemes that were simply misread or miscopied in the available manuscripts. Thus, providing an analysis for those roughly 190 forms will be an important part of the future research on that topic, and will require a lot of philological work.

In any case, it can already be noted that the attested pre-modern Hachijō lexemes are extremely diverse, and cover all parts of speech, all lexical *strata* (native, Sino-Japanese, onomatopoeic), as well as various semantic fields (such as flora and fauna, kinship lexicon, body parts, society and culture, etc.). In this perspective, it would also be interesting to study which parts of lexicon were best preserved in the contemporary language, and which ones were replaced more quickly. For instance, it would be interesting to see whether the proportion of different *strata* in the pre-modern sources matches the one of the contemporary language, since it would be plausible to assume that the pervasive influence of Tōkyō Japanese on Hachijō caused an increase in the use of Sino-Japanese in the contemporary language.

3.2 Phonology

Perhaps more importantly, the pre-modern sources of Hachijō can also provide some valuable insights on its historical phonology. However, in this perspective, we need to be careful about the classical Japanese spelling, which might in some cases be misleading. For instance, while looking at some attested forms like *kuwan* (coffin) or *hikkuwan* (servant), we might think that pre-modern Hachijō preserved Sino-Japanese labio-velars /kw/ and /gw/ before /a/. However, another hypothesis is to interpret this spelling as being imitated from the classical Japanese cognates *kuwan* 棺 and *hikuwan* 被官. In the absence of any evidence in favour of one or the other of these hypotheses, this question cannot be answered yet. Similarly, since we observe some variations between palatal /s/ and /h/ in initial position, as in *shicchou* (seventh son) (most sources)/*hicchou* (Ōta, c. 1811), we might think that, in pre-modern Hachijō, like in the modern-day language, there was no clear opposition between those two phonemes in this phonological context. However, since this /sy/ ~ /hy/ merging

– It is safely attested in pre-modern Hachijō (with at least five different occurrences, and a rather transparent composition), meaning that it is unlikely to have been miscopied from something else.
– It is, to the best of my knowledge, absent from all contemporary sources (except when those are quoting pre-modern documents).

is also a well known feature of the Edo/Tōkyō dialect,[11] it is not clear whether this variation is due to the object language, to the language of the transcriber, or to a convergence of both. These questions require further investigation.

In other cases however, as stated before, the important allography found in pre-modern sources can provide some valuable insights on the phonology of pre-modern Hachijō. Namely, based on the available data, we can assume that pre-modern Hachijō had:

- systematic palatalisation of /s/ before /e/ in initial position, hence the writing of the word *shei* (to know) (< *shoi, adnominal *shoke*) as <*sei*>
- occasional medial gemination of occlusive /k/, hence a variation between <*yoke*> (good) (Takahashi 1802; Kondō 1858) and <*yokke*>,[12] or between *nekoi* (small) (Kodera 1796; Kondō 1858) and *nekkoi* (most other sources)
- sporadic palatalisation of medial /s/, hence <*daisan*> (Hattori 1811; Ōta c. 1811; Kondō 1858) ~ <*daishan*> (Takahashi 1802; Kondō 1858)

All of these features are also well documented in the contemporary language.

In this perspective, another interesting variation is the initial *r-* ~ *d-*, which can be seen in the name 'sixth son', which is written as *rokurou* (in most sources) and *dokurou* (in Akiyama 1791; Kodera 1796). This variation is interesting to observe, not only because it exists in the contemporary language and is consciously (usually negatively) perceived by the speakers, but also because *rokurou* is the only word presenting this alternation in pre-modern sources. As a matter of fact, other words that start with /r/ in standard Japanese are 'only' attested with initial /d/ in Hachijō sources in the Edo period (like *dzunin* (exile) – Japanese *runin* 流人, or *djinki* (jealousy) – Japanese *rinki* 悋気). Thus, in my opinion, we can assume that the initial /r/ had undergone full fortition into /d/ in the stage of pre-modern Hachijō, before being shifted back to /r/ later by influence of the standard language, when the social stigma made such features of local speech appear negatively.

Similarly, a couple of sound changes can be safely said to have occurred before the Edo period in Hachijō, since they are already reflected in pre-modern sources, for instance:

- the dropping of several medial consonants, such as /s/, or /m/, e.g.: <*aitaba*> (angelica plant) (modern *e:taba* / *ya:taba*

11 Cf. for instance Martin 1987, 1.

12 Satō 1781; Akiyama 1791; Mishima 1801; Hattori 1811.

MJ < *asitaba* – Japanese *ashitaba*); <*akai*> (mallotus tree) (modern *akke:* / *akkya:* MJ < *akame* – Japanese *akame*)

- the merging of several vowel hiatuses into /ei/ or /e:/, as showed for instance in <*tei*> 'hand.ACC' (< te wo), <*kimei*> (guts) (<*kimo-wi), or <*tenegehe*> (towel) (< tenegui)[13]
- the gemination of voiced consonants after moraic /n/ in some topolects (which is a well-known characteristic of 'uphill' Hachijō), for instance in <*atsude*> /*adde*/ (why) in Kakusō 1858 (modern-day *ande* ~ *adde*).[14]

On the other hand, it is interesting to observe that some of the most widespread sound changes of Hachijō are not yet attested in pre-modern sources, such as:

- the differentiated treatment of /ei/ to either /ei/ or /i:/, for instance in *peiru* (to get wet) (most varieties) / *pi:ru* (Kashitate, Nakanogō, Sueyoshi). Pre-modern sources usually[15] all have <*ei*> (written as <ei> or <ehi>) in all occurrences
- the raising of /o/ to /u/ near labial[16] (for instance, pre-modern sources have *yowai*, while contemporary Hachijō is *yuwea*, *yuwakya*)
- the sporadic dropping of medial /r/ (pre-modern *anmari*, modern *anmai* '(not) quite')
- the merging of /z/ and /j/: pre-modern *zanmai*; modern *zanme:* ~ *janme:* (forgiveness')
- etc.

Therefore, it seems likely to assume that all of these innovations occurred later in the history of Hachijō. However, given the fact that the data is quite shallow in some cases (especially regarding some

13 There are, however, a few exceptions to this treatment, since we can observe spellings such as <*kogoheru*> (to be cold) (modern *kogeiru*) or <*nekkohi*> (small) (modern *nekkei*, *nekkokya*). This could either indicate that <ei> was still occurring in free variation with the original vowel hiatuses; but it is also possible to assume that those minority spellings are secondary, since they could be either etymological (in the case of *kogeiru*), or influenced by the rest of the pattern (in the case of *nekkei*, *nekk**o**ke*). Finally, it is also possible that this shift was still ongoing at that time (cf. *infra*).

14 This feature is only clearly attested in this source, perhaps due to 'uphill' topolects being overall less attested than the other topolects. There might also be one occurrence in Kondō 1858 *yotsupari* (urine), if we assume that this form is altered from **yobbari* (modern Hachijō *yonbari* ~ *yobbari*). However, it is equally possible that <ツ> is simply a miscopy from <ン>.

15 There might actually be one occurrence of /i:/ in Ōta's wordlist: <*hiiru*> (moth): all other pre-modern sources have <*heiru*> for the same word. However, I consider one occurrence too shallow an evidence to ascertain that this sound change was already starting at that time.

16 Again, there might be one occurrence of this sound change in Ōta's wordlist: *marubu* (to die), while all other pre-modern sources have *marobu*.

topolects), it is difficult to be certain at this stage, and much more research will be needed in order to grasp the details of Hachijō's historical phonology.

Finally, and perhaps most interestingly, a couple of sound changes might actually be directly documented in the pre-modern sources; namely:

- the differentiated treatment of /awa/ in different varieties (namely /o:/ in most varieties, /a:/ in Sueyoshi and /oa/ in Kashitate and Nakanogō, hence for instance *no:* ~ *na:* ~ *noa* 'rope'). For instance, the word for 'mother' *hawa (Japanese *haha*) appears as *hou* (Takahashi 1802), *hoa* (Hattori 1811) and *ha:* (Ōta 1820; Kondō 1858). Since this change is attested in some words, but not in other within the same sources, we can assume that the older and the more recent pronunciation were still, at that time, in more or less free variation
- the merging of /ai/, /ae/ and /ayu/ to either /e:/ or /ya:/,[17] depending on the topolect (as in *ne:* ~ *nya:* 'seedling'). In pre-modern sources, <*ai*>, <*ahi*>, <*ae*>, <*ahe*> are by far the most common spellings for that /e:/ ~ /ya:/ phoneme, and appear to be in rather free variation.[18] However, <*ei*> also occurs sporadically: for instance, *akke:* ~ *akkya:* (heel) occurs as <*akkahi*> in Ōta, but as <*akkei*> in Kondō (1858); similarly *ke:daruku* ~ *kya:daruku* (being dull) appears in Kondō as *keidaruku*.[19] This seems to show that there was at that time some degree of free variation between older and newer pronunciations, at least in some topolects; perhaps not unlike in modern-day colloquial Japanese (*urusai* ~ *uruse:* 'noisy')[20]
- (possibly) the lowering of /i/ to /e/ after labial: *hebira* (clothes) (most sources) ~ *hebera* (Hond 1810), *tsugumi* (kneecap) (Satō 1781) ~ *tsugume* (most sources)

However, the data remains very shallow, and since it is difficult to know for sure which Hachijō topolect is attested in each text, it is almost impossible to have certainties at this point. Besides, due to the scarcity

17 Examples of <*yaa*> spellings are not entirely clear in pre-modern sources, possibly because 'uphill' topolects are less attested. More research is needed on that topic.

18 As a matter of fact, <*ae*> and <*ahe*> also occur in cases in which they are not etymological, such as *mahe* (cocoon) in Ōta's wordlist (modern *me:* ~ *mya:* < *mayu) or *nahe* 'not.FIN' (modern *ne:* ~ *nya:* < *nasi).

19 There are several possible occurrences of <ei> < *ai in other sources than Kondō (including older ones). However, the etymology of those words is less certain, which makes them less solid evidence.

20 This free variation is perhaps best attested in Dickins and Satow's (1878) account on Hachijō, in which the verb *mairu* (to take), occurs as both <*meeru*> and <*mairu*> in *rōmaji*.

of sources, there are, in many cases, no occurrences at all which could indicate whether a sound change had occurred or not in premodern Hachijō. For instance, none of the words which exhibit the merging of /m/ and /n/ in palatal environment (such as *michiru* ~ *nichiru* 'to get full') are attested in pre-modern sources.[21] Thus, it is impossible to determine whether this change took place in pre-modern times or not. In other cases, it is impossible to tell whether a sound change is attested or not, because of the classical spelling: for instance, the modern Hachijō word for 'coffin' is *gan* which appears in old sources as <*kuwan*>, and it is unclear whether the initial consonant was voiced after the Edo period, or whether the old sources simply exhibit a missing *dakuten*.

More generally, in many cases there are too few occurrences to ascertain whether a sound change occurred, or whether the few possible occurrences are miscopies. For instance, the raising of /ou/ to /u:/ is well documented in some contemporary topolects of Hachijō and could be attested in the pre-modern language in the word *magou ni* (really), since a form <*magu: ni*> is also found in Kondō (1858). However, since we are dealing with a *hapax*, it is also equally reasonable to assume that this occurrence is a miscopy of <ゴ> to <グ>, and further evidence will have to be brought up to decide between those two possibilities.

In fact, even when we are dealing with a widely attested word, the allography can be difficult to analyse. For instance, the word for 'ninth son' exhibits several different spellings:

- <*kurau*> (Satō 1781)
- <*kuurou*> (Furukawa 1797)
- <*kuteu*> (Akiyama 1791; Kodera 1796)
- <*kutteu*> (Kakusō 1848), <*kutsuchiyau*> (Kondō 1858)

However, it is difficult to make sense of this diversity. While <*kuteu*> appears to be the best attested form, and fits the modern *kuchou* (Nagakubo 1937; Isozaki 1977), then it is unclear how to interpret the geminate in Kakusō and Kondō. This could either indicate that pre-modern Hachijō did not have any opposition between plain and geminated /ch/ in medial position; or it could instead be a secondary spelling, influenced by the forms *shicchou* (< *shichirou*) and *hacchou* (< *hachirou*). Finally, in that case, we could assume that <*kurau*> was miscopied from <*kuteu*> (with <テ> misread as <ラ>),

21 In fact, there might be two examples of neutralization of the *m* ~ *n* opposition in palatal environment in Ōta's wordlist. Namely, we could assume that the form <*iden*> 'oh well' is miscopied for **ideni*, which could be a variant of the well attested form *idemi* (which also occurs in the same wordlist). Similarly, we can assume that the form <*mishihitowomofu*> (which is translated as 'ugly') is to be read **mikei* to omou 'I think that [it] is ugly', with **mikei* being a variant of the well attested form *nikei* 'ugly' (SJ *nikui*). However, given the high number of miscopies in that text, these two examples are probably not clear enough to be considered proof of this phenomenon.

and that <*kuurou*> exhibits either a spelling influenced by the standard *kurou* 九郎, or a miscopy of <ッテ> as <ウロ>. On the other hand, it is also possible to assume the contrary, that is, that *kurou* (attested in Satō with classical spelling <*au*> for /ou/) is the genuine old form (as expected through the etymology), and that Furukawa <*kuurou*> exhibits the same form with an erroneous lengthening. In that case, we could assume that the refection of *kurou* to *ku(c)chou* on the model of the preceding forms *shicchou* and *hacchou* (which also extended to the following *jicchou*) was rather recent, so that both form were still coexisting at the end of the eighteenth century. Again, more research is needed in order to decide between those two possibilities.

3.3 Grammar

Finally, it must be noted that although Hachijō pre-modern sources mostly contain isolated words rather than complete sentences, they still exhibit a lot of grammatical elements that are considered characteristic for that variety, such as:

- the adjective adnominal form *-ke* (first attested in 1781)[22] and its final form *-kya* (c. 1811)
- the verb adnominal *-o* (1848)[23]
- the adjectival negative final form *-nnaka* (1801) and its adnominal counterpart *-nnoa* ~ *-nno:* ~ *-nna:* (1797)
- the copula *dara* (1802) and its adnominal form *doa* ~ *-do:* ~ *-da:* (1848)
- the past tense ending *-ara* (1848), and its adnominal *-oa* ~ *-o:* ~ *-a:* (1811)
- the reduplicated past *-arara*, and its adnominal *-aroa* ~ *-aro:* ~ *-ara:* (1848)
- the conjectural morpheme *nou* (1811)
- the comparative/volitional morpheme *-gon* ~ *-gan* (1802)
- the diminutive/derogative/nominalising morpheme *-me* (c. 1811)
- the intensive verbal prefix *hiQ-* (1796)
- etc.

However, several less typical morphemes also abundantly occur in the same sources, such as a final adjective morpheme *-shi*, a past tense morpheme *-ta*, an adnominal copula *na*, a negative auxiliary *-nu*, etc. In some cases, it seems likely that this allomorphy is due to an influence of classical Japanese (for instance, in the case of adjective

22 The date of the first attestations are based on current research results, but might be revised as new sources are discovered and treated.

23 Many grammatical morphemes are first attested at this date, since Kakusō's book contains the first two long texts in Hachijō.

adnominal *-ki* instead of *-ke*). However, in other cases, it can be difficult to decide whereas a form was influenced by other varieties of Japanese (and especially by classical Japanese or by the dialect of Edo), or whether is an inherited form, which gradually fell out of use in the recent history of the language. All those questions would need further investigation that would by far exceed the scope of this article.

4 Conclusion

Thanks to their unique position within Japan, the South Izu islands were fortunate enough to be abundantly described during the Edo period in many documents. However, while those pre-modern sources for Hachijō are remarkably numerous, diverse and invaluable, they remain critically understudied to this day. Several of those documents are still in need of a modern edition, and a comprehensive study of all those sources would also prove most useful, in order to analyse obscure forms. The only existing attempt of a compilation of those documents (NINJAL 1950) is now quite outdated, and not exempt of shortcomings.

Given the facts that those works document more than 100 years of Hachijō's linguistic history, they might exhibit some ongoing language changes and archaisms that could help shed light on important questions about Hachijō, such as its classification within Japonic languages. However, much more research is needed at this stage, in order to get a clearer picture.

Bibliography

Akiyama Funan 秋山富南 (1791). *Nanpō kaitō-shi* 南方海島志 (Documents from the Southern Sea Islands) [manuscript]. Provided by the National Institute of Japanese Literature. 3 different versions. https://kotenseki.nijl.ac.jp/biblio/200017656/viewer?ln=en.

Akiyama Funan (1791). *Nanpō kaitō-shi.* Hagiwara Masao 萩原正夫 (1901). *Izu shichitō-shi* 伊豆七島志 (Documents from the Seven Izu Islands). Self-published.

Akiyama Funan (1800). *Izu shisho* (Manuscript of Izu) [manuscript]. National Institute of Japanese Literature. https://kokusho.nijl.ac.jp/biblio/100334767/1?ln=en.

Akiyama Funan [1800] (1929). *Izu shisho* 豆州志稿. Shizuoka: Shizuoka kyōdo kenkyū kai.

Alaszewska, J. (2018). "Taming the Plague Demons Border Islanders and the Ritual Defense of Japan". Rambelli, F., *The Sea and the Sacred in Japan: Aspects of Maritime Religion*. London: Bloomsbury Publishing, 23-37.

Asahi Yoshiyuki 朝日祥之 et al. (2004). "Okinawa-ken ni denshō sareteiru Hachijō-go — Minami-daitō-jima umare no Asanuma-en-san no chūkan hōkoku" 沖縄県に伝承されている八丈語—南大東島生まれの浅沼圓さんの中間報告 (The Transmission of the Hachijō Language in the Okinawa Prefecture – Interim Report of Mr. Asanuma En, Born on Minami-daitō Island). *Chiiki gengo* 地域言語, 15, 1-14.

Asanuma Ryōji 浅沼良次 (1999). *Hachijō-jima no hōgen jiten* 八丈島の方言辞典 (Dialect Dictionary of the Dialect of Hachijō Island). Tōkyō: Asahi Shimbun Publishing Service.

Baudel, É. (forthcoming). *Shima-kotoba: A Synchronic and Diachronic Study of Hachijō* [PhD Dissertation]. Paris: EHESS.

De Boer, E. (2020). "The Classification of the Japonic Languages". Robbeets, M.; Savelyev, A. (eds), *The Oxford Guide to the Transeurasian Languages*. Oxford: Oxford University Press, 40-58.

Dickins, F.V.; Satow, E. (1878). "Notes of a Visit to Hachijô in 1878". *Transactions of the Asiatic Society of Japan*, 6, 435-77.

Furukawa Koshōken 古川古松軒 (1747). *Hachijō doryaku-shi* 八丈島土畧誌 (Outline document of land properties of Hachijō island) [manuscript]. Made available by the Tōkyō National Museum. https://webarchives.tnm.jp/dlib/detail/4785.

Furukawa Koshōken (1797). *Hachijō hikki* 八丈島筆記 (Notes about Hachijō Island) [manuscript]. Tōkyō National Museum. https://webarchives.tnm.jp/dlib/detail/4785.

Hachijō-jima daigai-chō 八丈島大概帳 (Large Account of Hachijō Island) [manuscript]. [1814] (1839). Provided by the University of Waseda. https://archive.wul.waseda.ac.jp/kosho/i14/i14_a0556/i14_a0556.pdf.

Hattori Yoshitaka 服部義高 (1811). *Hachijō kiriko-ori* 八丈裁衣織 (The Weaving of Hachijō Cut Clothes) [manuscript]. Several versions are available online, for instance one provided by the Edo Tōkyō Museum. https://www.edohakuarchives.jp/detail-508.html.

Hashimoto Naoyuki 橋本直幸; Long, D. (2003). *Oya-hōgen to ko-hōgen – Ogasawara-shotō to Minami-daitōjima ni okeru Hachijō-go* 親方言と子方言—小笠原諸島と南大東島における八丈語 (Mother Dialect and Daughter Dialect – The Hachijō Language in the Ogasawara Islands and Minamidaitō

Island). Presentation given at the 105th Conference of the Variation Research Group (変異理論研究会).

Hirako Tatsuya 平子達也; Pellard, T. (2013). "Hachijō-go no furusa to atarashisa" 八丈語の古さと新しさ (The Archaisms and Innovations of the Hachijō Language). Kibe Nobuko 木部 暢子, *Shōmetsu kiki hōgen no chōsa hozon no tame no sōgō-teki kenkyū – Hachijō hōgen chōsa hōkokusho* 消滅危機方言の調査・保存のための総合的研究 ー 八丈方言調査報告書. (General Study for Research and Conservation of Endangered Dialects in Japan – Research Report on Hachijō Dialects). Tachikawa: NINJAL, 47-67.

Hirayama Teruo 平山輝男 et al. (eds) (1992-94). *Gendai Nihongo hōgen daijiten*. 現代日本語方言大辞典 (Large Dialect Dictionary of Contemporary Japanese). Tōkyō: Meiji shoin.

Honda Toshiaki 本多利明 (c. 1810). *Meii kokuwa* 明異国話 (Famous stories from foreign places) [manuscript].

Honda Toshiaki (c. 1810). *Hachijō-jima monogatari* 八丈島物語 (Tales from Hachijō Island) [undated] [manuscript].

Hoshina Kōichi 保科孝一 (1900). "Hachijō-jima hōgen" 八丈島方言 (The Dialect of Hachijō Island) (five articles). *Gengogaku zasshi* 言語学雑誌, 1(2), 39-50; 1(3), 33-40; 1(4), 57-67; 1(7), 22-47; 1(10), 59-66.

Iannucci, D. (2019). *The Hachijō Language of Japan: Phonology and Historical Development* [PhD Dissertation]. Mānoa: University of Hawai'i.

Ike Norimitsu 池則満 (1796). *Hachijō-jima-ki* 八丈島記 (Records of Hachijō Island) [manuscript]. Made available online by the University of Tsukuba. https://kokusho.nijl.ac.jp/biblio/100334739/1?ln=ja.

Isozaki Otohiko 磯崎乙彦 (1977). *Hachijō kaiko* 八丈回顧 (Remembering Hachijō). Self-published.

Kakusō Kizan 鶴窓帰山 (1848). *Yatake no nezamegusa* 八丈の寝覚草 (The waking up of Hachijō grass) [manuscript]. Provided by the National Diet Library. https://dl.ndl.go.jp/pid/2537093/1/2.

Kakusō Kizan [1848] (1984). *Yatake no nezamegusa*. Edited by Nakada Norio 中田祝夫. 1984. Tōkyō: Benseisha Bunko.

Kanayama Shōkō 金山正好 (ed.) (1976). *Izu shotō junken kiroku-shū* 伊豆諸島巡見記録集 (Collection of Inspection Records in the Izu Islands). Tōkyō: Ryokuchi-sha.

Kaneda Akihiro 金田章宏 (2001). *Hachijō hōgen dōshi kiso kenkyū* 八丈方言動詞の基礎研究 (Basic Study of the Verbs in the Hachijō dialect). Tōkyō: Kasama shoin.

Kaneda Akihiro (2012). "Hachijō hōgen ni okeru arata na henka to jōdaigo" 八丈方言における新たな変化と上代語 (New Changes and Old Japanese in Hachijō Dialect). *Gengo kenkyū* 言語研究, 142, 119-42.

Kasai Shigeo 葛西重雄; Yoshida Kanzō 吉田貫三 (1964). *Hachijō-jima runin meimei-den* 八丈島流人銘々伝 (The Individual Bibliography of Hachijō island exiles). Tōkyō: Daiichi shobō.

Kobayashi Iichi 小林亥一 (1984). *Aogashima no kotoba* 青ケ島の言葉 (The Words of Aogashima). Aogashima: Aogashima Village Office.

Kodera Ōsai 小寺応斎 (1796). *Shichitō nikki* 七島日記 (Diary From the Seven Islands) [manuscript]. Waseda University: https://www.wul.waseda.ac.jp/kotenseki/html/ru04/ru04_03948/index.html.

Kodera Ōsai [1796] (1976). "Shichitō nikki". Kanayama, Shōkō, *Izu shotō junken kiroku-shū.* Tōkyō: Ryokuchi-sha, 5-116.

Kondō Tomizō 近藤富蔵 [1858] (1964-76). *Hachijō Jikki* 八丈實記 (The true account of Hachijō). 7 vols. Edited by the Hachijō Jikki Kankō kai 八丈実記刊行会. Tōkyō: Ryokuchi-sha.

Kupchik, J. (2011). *A Grammar of the Eastern Old Japanese Dialects* [PhD Dissertation]. Mānoa: University of Hawai'i.

Long, D. (2004). "Shima no gengo no futatsu no kao – Sesshoku to koritsu no haza made" 島の言語の二つの顔–接触と孤立のはざまで (The Two Faces of Insular Languages – Between Contact and Isolation). *Gekkan gengo*, 33(1), 42-50.

Martin, S.E. (1987). *The Japanese Language Through Time*. New Haven: Yale University Press.

Matsuda Heiemon 松田兵右衛門 (c. 1810). *Hachijō-jima runin no hanashi* 八丈島流人之話 (The story of the Hachijō island exiles). Honda Toshiaki, *Meii kokuwa* [manuscript].

Mishima Masahide 三島正英 [1801] (1968). *Izu Nanashima fudo sairan* 伊豆七島風土細覧 (Detailed Observation of the Chronicles of the Seven Izu Island). Takakura, Shin'ichirō (ed.) 高倉新一郎, *Nihon shomin seikatsu shiryō shūsei* 日本庶民生活史料集成. Vol. 1, *Nantō - Tanken, kikō, chishi* 南島, 探検・紀行・地誌 (Southern Islands, Explorations, travelogues, topographies). Tōkyō: San'ichi shobō, 644-61.

Mollard, N. (2021). "Kyokutei Bakin's Philological Research and the Writing of Historical Narrative: Study of the Izu Islands in the *Marvelous Story of the Drawn-Bow Moon*". *Studies in Japanese Literature and Culture*, 4, 19-44.

Moseley, C. (ed.) (2009). *Atlas of the World's Languages in Danger*. Paris: UNESCO.

Motoyama Keisen 本山桂川 (1934). "Hachijō-jima hōgen-shō" 八丈島方言抄 (Excerpt of Hachijō island dialect). *Kaitō minzoku shi, Izu shotō-hen* 海島民俗誌伊豆諸島篇 (Magazine of Insular Folklore - The Izu Islands). Tōkyō: Issei-sha. 163-82.

Nagakubo Mitsuru 永久保満 (1937). *Shumi no tōkyō-fu Hachijō-jima shi* 趣味の東京府八丈島誌 (Selected Island Notes from Hachijō Island). Tōkyō: Minami Nihon shinbun-sha.

NINJAL (国立国語研究所) (1950). *Hachijō-jima no gengo-chōsa* 八丈島の言語調査 (Linguistic Study of Hachijō island). Tōkyō: Shūei Shuppan.

Ogawa Takeshi 小川 武 (1958). "Hachijō hōgen no ruikei to keitō" 八丈島方言の類型と系統 (Typology and Phylogeny of the Hachijō Dialect). *Shin-chiri* 新地理, 7(2), 132-5.

Ōhara Masanori 大原正矩 (1811). *Hachijō-shi* 八丈誌 (Hachijō Documents) [manuscript]. Several versions available online, for instance provided by the National Diet Library. https://dl.ndl.go.jp/info:ndljp/pid/2609642.

Ōta Hikosuke 太田彦助 [1796](1976). *Kaitō zatsuwa* 廻島雑話 (Short word about an island trip). Kanayama Shōkō 金山正好 (ed), *Izu shotō junken kiroku-shū* 伊豆諸島巡見記録集 (Collection of Inspection Records in the Izu Islands). Tōkyō: Ryokuchi-sha, 117-69.

Ōta Nanpo 大田南畝 (c. 1811). *Ichiwa ichigen* 一話一言 (Small texts with small comments) [manuscript]. Provided by the National Institute of Japanese Literature. https://kokusho.nijl.ac.jp/biblio/100183452/1?ln=en.

Ōta Nanpo [c. 1811] (1996). *Ichiwa ichigen*. Edited by Nihon Zuihitsu Taisei 日本随筆大成. Tōkyō: Yoshikawa kōbunkan.

Satō Yukinobu 佐藤行信 (1781). *Izu kaitō fudoki* 伊豆海島風土記 (Chronicles of the Ocean Islands of Izu) [manuscript]. Available on the website of the Fuji library. https://library.fujishi.jp/c1/bib/fuji_11696026.pdf.

Satō Yukinobu. [1781] (1929). *Izu kaitō fudoki.* Edited by Shizuoka Kyōdo kenkyū-kai 静岡郷土研究会. Shizuoka: Shizuoka kyōdo kenkyū kai.

Satō Yukinobu [1781](1974). *Izu kaitō fudoki* 伊豆海島風土記. Edited by Higuchi Hideo 樋口秀雄. Tōkyō: Ryokuchi-sha.

Sawara no Kisaburō 佐原喜三郎 [1839] (1968). *Asahi gyakutō-ki* 朝日逆島記 (Asahi's Records of Topsy-turvy Islands). Takakura, Shin'ichirō 高倉 新一郎 (ed.), *Nihon shomin seikatsu shiryō shūsei* 日本庶民生活史料集成. Vol. 1, *Tanken, kikō, chishi* 探検・紀行・地誌. Tōkyō: San'ichi shobō, 701-8.

Takahashi Yoichi 高橋與市, also written 高橋 与一. [1802](2012). *En'ō kōgo* 園翁交語 (Conversation of Old Gardeners). Hachijō-jima no komonjo o yomu kai 八丈島の古文書を読む会, *Hachijō-jima no Komonjoshū* 八丈島の古文書集, vol. 1. Hachijō: Nankai taimusu, 1-169.

Takakura Shin'ichirō 高倉新一郎 (ed.) (1968). *Tanken kikō chishi* 探検・紀行・地誌 (Explorations, Travelogues, Topographies). Vol. 1, *Nantō*, *Nihon shomin seikatsu shiryō shūsei* 南島,日本庶民生活史料集成 (Southern Islands, Collection of Japanese Common People's Life History Materials). Tōkyō: San'ichi shobō, 644-841.

Tamura Eitarō 田村栄太郎 (1928). "Hachijō-jima hōgen" 八丈島方言 (The Dialect of Hachijō Island). *Kokugo Kokubun no Kenkyū* 国語国文の研究, 20, 127-35.

Tsuzuku Tsuneo 都竹 通年雄 (1955). *Hachijō-jima hōgen* 八丈島方言 (The Dialect of Hachijō Island). Kindaichi Haruhiko 金田一春彦 (ed.), *Nihongo no shujusō*, *Kōza nihongo daisankan* 日本語の種々相, 講座日本語 第三巻 (Various Aspects of the Japanese Language, Lectures on the Japanese Language, Vol. 3). Tōkyō: Ōtsuki shoten, 37-8.

Yamada Heiemon 山田平右工門 (2010). *Kieteiku shima-kotoba: Hachijō-go no keishō to sonzoku wo negatte* 消えていく島言葉:八丈語の継承と存続を願って (A Disappearing Insular Language: Hoping for the Inheritance and Survival of Hachijō Language). Tōkyō: Ikuhō-sha.

Yoshimachi Yoshio 吉町義雄 (1951). "En'ō Kōgo to Hachijō Jikki 'no shima kotoba'" 「園翁交語」と「八丈実記」の島言葉 (The Island Words of the En'ō Kōgo and the *Hachijō Chronicle*). *Bungaku Kenkyū*, 42, 127-54.

Philological and Linguistic Analysis Working Together
Exploring Language Diversity in Japan
edited by Paolo Calvetti

Elements of Sakhalin Ainu Phonetics, Phonology, and Morphosyntax in Bronisław Piłsudski's Corpus of Ainu Folklore

Elia Dal Corso
Università Ca' Foscari Venezia, Italy

Abstract In this paper the author focuses on a number of linguistic features of East Sakhalin Ainu dialects. The language data for this study comes from a collection of traditional Ainu folklore elicited in 1903-04 and published as *Materials for the Study of the Ainu Language and Folklore* by Bronisław Piłsudski in 1912. The observations on phonetics, phonology, and morphosyntax will be discussed in light of more recent data coming from West Sakhalin Ainu dialects collected in the 1960s-1970s. This will allow for a brief consideration of the dialectal differences occurring between the two dialect subgroups as well as of the possible path of language development that took place during the 60-70 years that separate the two sources. The aim of the author is to highlight how a thorough descriptive work on old sources of an indigenous, minority language like Ainu can give to East Asian historical linguistics and to linguistic typology more generally.

Keywords Ainu language. Ainu folklore. Bronisław Piłsudski. West Sakhalin Ainu dialects. Historical linguistics.

This essay is partially based on: Dal Corso, E. (2022). "A Comparative Analysis of an East Sakhalin Ainu Folktale Collected by Bronisław Piłsudski". *Annali di Ca' Foscari. Serie orientale*, 58(1), 907-46. http://doi.org/10.30687/AnnOr/2385-3042/2022/01/030.

Edizioni Ca'Foscari

Ca' Foscari Japanese Studies 24
Linguistics and Language Education 4
e-ISSN 2610-9395 | ISSN 2610-8976
ISBN [ebook] 978-88-6969-845-3 | ISBN [print] 978-88-6969-846-0

Peer review | Open access
Submitted 2023-09-19 | Accepted 2024-01-11 | Published 2024-08-28

DOI 10.30687/978-88-6969-845-3/003

1 Introduction

This paper gives an overview of a number of grammatical features characteristic of the eastern dialect subgroup of Sakhalin Ainu, an isolate autochthonous of Russia. The main reference for this study is the corpus of 27 folktales of the Sakhalin Ainu oral tradition collected by the Polish ethnographer and linguist Bronisław Piłsudski in 1903-04 during the final years of his second stay on Sakhalin, which were published in 1912 as *Materials for the Study of the Ainu Language and Folklore* (abbreviated as *Materials…* henceforth). As I will reiterate below, the *Materials…* are a priceless resource for Ainu studies because they represent one of the few documentation works ever carried out on the Sakhalin variety of Ainu, which went extinct at the end of the twentieth century. Furthermore, the 27 folktales that constitute the corpus were elicited from first-language native speakers who, for the most part, were completely monolingual. Therefore, the texts in Piłsudski's corpus illustrate a stage of Ainu when the language had little to no influence from the prestige languages of the colonizers (i.e. Japanese and Russian). The *Materials…* are today easily accessible and a detailed philological account of their earliest redactions which eventually led to their publication is given in Majewicz (1998a), but the folktales therein still await to be fully analyzed linguistically. While providing a detailed analysis of even one specific feature of the Sakhalin Ainu language depicted in Piłsudski's *Materials…* falls far beyond the scope of this paper, it is still possible to hint to the valuable grammatical information contained in the texts by focusing on a number of language behaviors that appear striking when compared against more recent data and descriptions available for western dialects of Sakhalin Ainu (Murasaki 1976; 1979; Dal Corso 2021a). Therefore, by taking into brief consideration the phonetics, phonology, and morphosyntax of East Sakhalin Ainu, I highlight the importance of Piłsudski's documentation work for the diachronic and synchronic study of Sakhalin Ainu and of the Ainu language as a whole. The main aim of this paper is to underline the contribution that a thorough descriptive work on old sources of an indigenous, minority language like Ainu can contribute to East Asian historical linguistics and to linguistic typology more generally.

The paper is organized as follows. Section 2 provides some generalities and typological information on Sakhalin Ainu, while Section 3 introduces Piłsudski's *Materials…* and the West Sakhalin Ainu sources to which I refer for comparison. Section 4 is dedicated to the discussion of peculiar features of East Sakhalin Ainu pertaining to phonetics (phonemic inventory and phonotactics, Section 4.1), phonology (phonological processes, Section 4.2), and morphosyntax (verbal transitive agreement and the verbal prefixes *e-*, *ko-*, and *yay-*, Sections 4.3 and 4.4). The peculiarities concerning these areas of the

East Sakhalin grammar will be briefly compared with language data coming from West Sakhalin Ainu in order to highlight possible dialectal differences or traces of a diachronic development of the language. Section 5 concludes.

2 The Sakhalin Ainu Language

The Ainu language shows great dialectal variation. The language has three main varieties that are defined geographically and correspond to the language once spoken on Hokkaidō,[1] Sakhalin, and the Northern Kuril Islands. Of these three varieties only Hokkaidō Ainu survives today in some of its southern and central-eastern dialects which are all regarded as critically endangered. Both the Sakhalin and Kuril varieties are extinct.[2] On the basis of variations in the lexicon, we can distinguish the dialect groups illustrated in the map in Figure 1 (Hattori, Chiri 1960; Asai 1974). Figure 2 shows a map with the location of the Ainu villages along the Sakhalin east coast where Piłsudski collected the texts that I examine in this study. The settlements of Rayciska and Maoka also appear in the map – these were the native villages of the two consultants who provided the language data for West Sakhalin Ainu collected by Murasaki (1976), which constitute the basis for the only language descriptions available for Sakhalin Ainu (Murasaki 1979; Dal Corso 2021a).

The canonical word order in intransitive and transitive clauses of Sakhalin Ainu is SV and AOV respectively. Sakhalin Ainu is a polysynthetic, agglutinating language, it is strongly head-marking and right-headed with a rich but largely non-productive morphology. There is no grammatical agreement of gender while number may be non-obligatorily distinguished on nouns and verbs by dedicated morphology

1 The Hokkaidō variety extended also outside of the main island to Matomai, Shikotan, Iturup, and Kunashir – the four southernmost islands of the Kuril chain where a most probably distinct dialect of the variety was spoken. It is reported already in XVIII century travelogs, among which Krasheninnikov (1755) that the language spoken on these islands was hardly comprehensible to the Ainu living on Paramushir and Onekotan – two of the northernmost islands on the opposite end of the chain. On the basis of linguistic evidence coming from a Kuril Ainu glossary later collected by captain Vasily Mikhaylovich Golovnin in 1811, Bugaeva and Satō (2021) argue that Northern Kuril Ainu and Southern Kuril Ainu were in fact two distinct dialects of a separate Kuril Ainu variety.

2 Although its last native speaker died in 1994 (Murasaki 2001, 2), Sakhalin Ainu is still presently spoken and it is the target of the tireless revitalization efforts of today's Ainu community in Japan, with one of the main activists involved being Kitahara Jirōta Mokottunas (see Section 3). In today's revival movement, the Sakhalin Ainu language is also referred to as *Enciw'itah* 'language of the Enciw', a glossonym that makes use of the historical Sakhalin Ainu autonym *enciw*, meaning 'human being'. This distinguishes Sakhalin Ainu from Hokkaidō Ainu, which is referred to by its speakers as *Aynu itak* 'Ainu language' or *Akor itak* 'our language'.

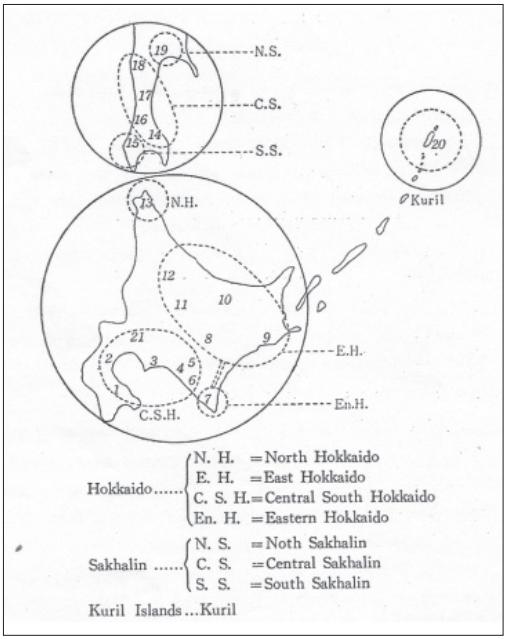

Figure 1 Ainu dialects (Asai 1974, 100)

Figure 2 Ainu settlements on the Sakhalin east coast

or via morphosyntactic processes (e.g. verb stem reduplication). Verbal morphology shows a number of morphemes dedicated to encoding applicative, causative, antipassive and other valency-changing strategies. There is no formal marking for tense, but the language displays synthetic and analytic constructions to mark mood, aspect, and evidentiality of the predicate (Chiri [1942] 1973; Murasaki 1979; Dal Corso 2021a).

3 Piłsudski's *Materials for the Study of the Ainu Language and Folklore* (1912) and West Sakhalin Ainu Sources

Bronisław Piotr Piłsudski (Zułów (present day Zalavas) 1866 – Paris 1918) was a Polish ethnographer and linguist. After being sentenced to 15 years of penal labor on Sakhalin in 1887 for his alleged participation in a failed assassination attack against the Tsar Alexander III, he came in touch with the Ainu population of the island. Between 1896 and 1905 Piłsudski carried out extensive research on the language and traditions of various Ainu communities that inhabited the east coast of Sakhalin. His inquisitive nature and the fact that he

could mediate with Russian people on their behalf helped Piłsudski befriend the Ainu and, with time, he gained enough knowledge of the Sakhalin Ainu language to be able to discuss its lexicon and structure in a critical way.

Piłsudski elicited language data from the Ainu of Sakhalin in a period of less than ten years before the outburst of the Russo-Japanese war of 1904-05, an event that irremediably affected all future possibilities for Russian citizens to conduct research on the Ainu communities living on Sakhalin and, more importantly, the very lifestyle of those communities and the vitality of the language they spoke. After Japan's victory over Russia, the territories of Sakhalin island south of the 50th parallel were ceded to the Japanese who re-named them 'Karafuto'.[3] In the course of World War One, the imperialist government of Tōkyō intensified the progressive japanization of cultural and ethnic minorities, among which the Ainu, that had been started during the Meiji period (1868-1912). One act of Japan's assimilation policies was to discourage the Ainu from using their native language and to force them to take on a Japanese lifestyle. After World War Two and Japan's defeat in the conflict, Allied Forces decided for Karafuto to be given back to Russia, together with other territories facing the Okhotsk sea over which Japan had taken control. Since by that time the Ainu living on Karafuto had been granted Japanese citizenship (primarily to further legitimate Japan's presence on the island), the transfer of the territories to Russia resulted in the forced relocation of a large number of Sakhalin Ainu to Hokkaidō or elsewhere in Japan. Only a small part of them decided to live in Russia instead, where they were assimilated culturally and linguistically.

Relocation played a decisive role in the steady decline of Ainu traditions and linguistic vitality that had been set in motion by Japan's assimilation policies. The few Sakhalin Ainu who still practiced a traditional Ainu lifestyle could not continue to do so in a territory that was geographically so different from their native Sakhalin. Having their lifestyle become unsustainable, they had to leave it behind in favor of the Japanese way of living. Continuous prejudice and oppression towards them through the decades following the war also pushed the Ainu to abandon their language, which, in the case of Sakhalin Ainu, ceased to be actively passed on to new generations already during the first half of the twentieth century. Despite the critical situation, documentation of Sakhalin Ainu was still possible from the 1940s until the 1980s thanks to the last surviving native speakers of the language. The most notable outcomes of these documentation efforts are the collections of folklore texts and vocabulary recorded by Wada Bunjirō

3 Still today it is common in Japanese academia to refer to the Sakhalin variety of the Ainu language as 樺太アイヌ語 *Karafuto ainugo* (Karafuto Ainu language).

from a speaker of the Usoro dialect,[4] the corpus of folklore texts elicited from Asai Take who was also a speaker of the Rayciska dialect (Murasaki 2001), and the corpus of folklore texts and conversations elicited from Fujiyama Haru and Ōta Yuk, speakers of the Rayciska and Maoka dialects respectively (originally published in Murasaki 1976 and re-edited in Dal Corso 2021a). This latter corpus specifically contains the language data that will serve as a comparison for the observations on East Sakhalin Ainu in Section 4. All these works document different dialects of the Sakhalin West coast and, as such, represent a priceless resource on this Sakhalin Ainu dialect group. It was after the death of the last native speaker, Asai Take from Odasu (Otasuh 顎, now Parusnoe), in 1994 that Sakhalin Ainu was declared extinct.

Piłsudski recorded the language of different Ainu communities on the Sakhalin east coast at a time when Ainu was still used as the everyday means of communication by native (and in most cases monolingual) speakers. Indeed we must account for possible imprecisions in the texts that have reached us through Piłsudski's transcription, mainly due to the elicitation method he employed (see below). Nonetheless, the collected data depict a stage in the history of East Sakhalin Ainu when the language had little to no sustained contact with a main or prestige language such as Russian or Japanese and bilingualism was almost absent – a scenario that has rarely re-presented itself in following documentation efforts for any Ainu dialect or variety. Furthermore, given that all other substantial documentation work on the Sakhalin variety has been carried out on West Sakhalin Ainu, as mentioned above, the materials gathered by Piłsudski constitute our only considerable linguistic resource on Eastern dialects, which exhibit grammatical characteristics unfound in Western dialects. Therefore, the value of Piłsudski's work for the study of the Ainu language as a whole cannot be overstated.

Piłsudski (1912) contains twenty-seven *ucaskuma* (folktales) collected in the Ainu settlements of Tarayka, Tunayci, Ay, and Hunup. The first two texts in the collection are presented with a line-by-line English translation on the side and a literary translation also in English at the end. All remaining *ucaskuma* have no interlinear translation but are followed only by a literary translation in English. After the translation of each text, Piłsudski provides a large number of notes on syntax, semantics, lexicon, and the Ainu culture especially for those passages that would be hardly intelligible to the unacquainted European reader. Ainu speakers are briefly introduced at the beginning of the note section the first time they appear as the language informant of

4 Kitahara 2013; 2014; 2016; 2017; 2019.

a text. All texts in the collection were elicited through dictation[5] and, as Piłsudski (1912, IX) himself asserts, eliciting folktales from his informants with this method was the most efficient way for him to get more and more familiar with the language in the first place. Part of the total 350 texts collected on Sakhalin (most of which still unpublished today) was also recorded on phonographic wax cylinders, the majority of which unfortunately have deteriorated to the point that the audio track can no longer be reproduced.[6] Although Piłsudski was indubitably a careful transcriber and could very well discern a fluent speaker from a less skilled one, the method of eliciting through dictation has a number of (at least partially) inevitable pitfalls. Dictation is a slow process which in most cases forces an informant to interrupt in mid-narration or to repeat some passages in order to allow the transcriber to correctly report what is being said. Piłsudski (1912, 102) himself notices that this can very easily compromise the quality of narrative style and, ultimately, that of the linguistic data obtained. When conducting whatever kind of linguistic analysis on Piłsudski's corpora it is, therefore, important to consider that this method of elicitation may have influenced the quality of the data we analyze and, despite Piłsudski's meticulous editing work, we must account for possible incongruences, like unexpected syntax and other cases of 'broken' Ainu.

4 Grammatical Features of East Sakhalin Ainu Dialects

4.1 Phonemic Inventory and Phonotactics

Although there are no audio recordings available for the *Materials…*, we luckily can rely on the accurate close-phonetic transcription of Piłsudski and on his detailed explanation of the orthographic conventions employed in the redaction of the corpus. The *Materials…* thus become a valuable resource for investigation on phonetics and phonology, albeit with all due reservations on Piłsudski's listening skills (see my comment on the glottal stop below).

The phonemic inventory of East Sakhalin Ainu counts five vowel phonemes /a/, /ɛ/, /ɔ/, /i/, /u/, which Piłsudski represents in orthography as *a*, *e*, *o*, *i*, *u*, and eleven consonant phonemes /p/, /t/, /k/, /t͡ɕ/, /m/, /n/, /s/, /v/, /h/, /ɾ/, /j/ represented in the orthography by *p*, *t*, *k*,

5 He used the same collection method during his stay with the Nivkh of Northern Sakhalin.

6 For further discussion on Piłsudski's recordings on wax cylinders and the most recent laser technology that allowed to restore the audio from them see, among others, Murasaki 2013.

ć, *m*, *n*, *s*, *v*, *h*, *r*, *j*. Absent from the East Sakhalin Ainu inventory is the glottal stop /ʔ/, which is provisionally included in the phonemic inventory of West Sakhalin Ainu as a phoneme in free variation (Dal Corso 2021a, 23-4). Since the works of Hattori and Tamura in the 1950s, the glottal stop is marked with an apostrophe in Latin-script based transcriptions of Ainu (thus for instance *'aynu* to represent /ʔajnu/) but, for the sake of readability, it has become conventional not to mark it in texts. If in recent publications not writing glottal stops in non-ambiguous cases is only a matter of conventions, Piłsudski (1912) does not mention this sound at all. Given that the glottal stop seems to be a phonetic feature common to all Hokkaidō dialects and also to West Sakhalin dialects, it is at least odd that none of the East Sakhalin dialects surveyed by Piłsudski is said to have it. A possibility is that Piłsudski's 'Slav ear', as he himself defines it, is at fault in this instance and that glottal stops, though present, are never reported simply because they were not heard. That is, it is impossible to ascertain the presence of the glottal stop in East Sakhalin Ainu.

4.1.1 Vowels

In East Sakhalin Ainu, *e* and *o* are said to represent the two open-mid vowels /ɛ/ and /ɔ/ (Piłsudski 1912, 1). The quality of mid vowels of East Sakhalin Ainu contrasts sharply with that of the West Sakhalin *e* and *o*, which only correspond to the close-mid sounds /e/ and /o/. Piłsudski reports that, in unaccented syllables, the pronunciation of *e* and *o* tends to be 'narrowed' while that of *i* and *u* tends to be 'more open', with *e* and *i*, on the one hand, and *o* and *u*, on the other, becoming hardly distinguishable in many instances. From this explanation, I infer that *e* and *o* might have as possible allophones the close-mid [e] and [o] respectively, while *i* and *u* might have [ɪ] and [ʊ] as their respective allophones. This change in vocalic height is almost never attested in the folktales elicited from speakers of the Hunup and Tarayka dialects, but it appears quite commonly in the Ay and Tunayci dialects. In such instances, lowered *i* and *u* are transliterated as *e* and *o* while raised *e* and *o* are left unaltered in orthography, with no suprasegmental indicating the change. In contrast, [ɨ] deriving from rising and centralization of /e/ is overtly signaled in orthography, as *y*. The allophone [ɨ] is found in the environment nd_# and, more precisely, this realization concerns the *e* in the causative suffix *-te* (where voicing of /t/ into [t̬] is triggered by the preceding nasal) – e.g. *ćivénd***y** [t͡ɕivɛnt̬ɨ] < *ciwente* 'was destroyed'. The realization as [ɨ] is inferred from Piłsudski's (1912, 2) description of the sound as "akin to the Russian ы".

4.1.2 Plosives

Plosives /p/, /t/, and /k/ are always voiceless word-initially – e.g. ***p**oro* [pɔɾɔ] < *poro* 'to be big'. Voicing of plosives in intervocalic position is most common for all three phonemes, in contrast to what is observed in West Sakhalin Ainu where voicing is never attested for /p/ and remains a rare phenomenon even for /t/ and /k/ (Dal Corso 2021a, 26).[7] Voicing of plosives is also attested after the approximant /j/ in East Sakhalin Ainu. As Piłsudski (1912, 4) notes, the pronunciation of plosives in the two environments V_V and $C_{approx.}$_V is not exactly that of their voiced counterparts [b], [d], and [g], but rather it "wavers between the [voiceless] and [voiced] group". From this, I infer the allophones [p̬], [t̬], and [k̬]. Free variation of voiceless and voiced plosives is a common trait of the Ainu language since, as far as we know, in all its varieties and dialects the voicing opposition is not distinctive. Voicing, normally not rendered in modern transliterations, is evidenced in Piłsudski's corpus via the use of *b*, *d*, and *g* in orthography – e.g. *pájg**g**ara* [pajk̬aɾa] < *paykara* 'spring'. Voicing of plosives is otherwise possible, but not systematic, in the environment C_{nasal}_, which is not attested in West Sakhalin Ainu – e.g. *ćivén**d**y* [t͡ɕivɛnt̬ɨ] < *ciwente* 'was destroyed'.

Palatalization, marked in orthography by a diacritic ´, of /p/ and /k/ (but not /t/) and of the voiced allophones [p̬] and [k̬] is also reported (Piłsudski 1912, 4-5) – e.g. *emújk**é*** [ɛmujkʲɛ] < *emuyke* 'all', *tam**bé*** [tamp̬ʲɛ] < *tanpe* 'this thing'; palatalization is systematic for /k/ in the environment _ɛ. Finally, palatalization seems to show a correlation with voicing in the environment C_{nasal}_ɛ – compare *ta**mbé*** [tamp̬ʲɛ] < *tanpe* 'this thing' and *á**mpe**ne* [ampɛnɛ] 'really'. Forms such as the adverb *ampene* or the topic marker *neampe*, where palatalization of the plosive is missing, diachronically contain the same dependent noun *-pe* 'thing' found in *tanpe* and therefore the lack of palatalization could be accounted for by lexicalization. Palatalization is not attested in West Sakhalin Ainu.

Plosives are not allowed word-finally. In West Sakhalin Ainu, plosives in this position are systematically debuccalized. In East Sakhalin Ainu, these sounds are normally never noted in the transcription – thus e.g. *pate* [patɛ] < *pate* (for intended *pateh*, underlying form *patek**)[8] 'only', cf. West Sakhalin Ainu *pate**h*** [pateh], same meaning.

7 However, it is interesting how Piłsudski (1912, 4) asserts that voicing of plosive is much more common in Western Sakhalin, quite contrarily to what is reported in Dal Corso. It might be that contact with Japanese was among the causes of the absence of this trait of plosives that stands out in the more recent West Sakhalin Ainu data collected by Murasaki.

8 Here I employ an asterisk to mark the underlying phonological form of words. This symbol should not be taken as indicating a reconstructed form as it is conventional in historical linguistics.

This orthographic peculiarity is most likely to be ascribed to the fact that debuccalized plosives (or plosives with no audible release, as they are reported to be in some Sakhalin dialects (Tangiku 2022, 331)) are difficult to perceive and, therefore, are easily left out in the transcription. The underlying word-final plosive surfaces if the word is followed by a bound morpheme beginning with a vowel. Though more rarely than elision, debuccalization may be present word-finally also in East Sakhalin dialects. This happens between words in the environments $V_{1}\_V_{1}$ and $\_C_{plosive}$ – e.g. *na**x** án* [nax an] < *nah an* 'to be so', *ita**x** ki* [itax ki] < *itah ki* 'to make a speech', where *nah* and *itah* have underlying forms *nak** and *itak**. In both cases, debuccalized plosives are realized as [x]. In particular, debuccalization in the case of *nax án* may be indicative of the presence of a glottal stop (i.e. *'an*), but all other analogous cases should be taken into account to determine how this phoneme is distributed in East Sakhalin Ainu. Besides being common at morpheme boundary, debuccalization is possible also word-internally but only in the environment $\_C_{plosive}$ – e.g. *jú**f**́ke* [juɸkʲe] < *yuhke* 'to be strong', where [x] is further realized as [ɸ] under account of the preceding /u/ (cf. *yupu* 'to fasten', which shares with *yuhke* the same root *yup* featuring /p/ in a V_V environment).

4.1.3 Nasals

East Sakhalin Ainu dialects have the two nasals /n/ and /m/. While /m/ does not undergo any specific phonotactic change, Piłsudski (1912, 5) reports for /n/ the allophones [ŋ], rarely encountered, and [ɲ], which is found in the environments _# and _ɛ – e.g. *etun**ń**e* [ɛtunɲɛ] < *etunne* 'to do not want'. This allophone is rendered in orthography as *ń*. Orthography does not make it clear whether in a word like *etunne* the palatalization of [ɲ] is transferred also on the /n/ preceding it, thus giving [ɲː].

4.1.4 Fricatives

The phoneme /s/ is represented as *s*. This phoneme has an allophone which, from Piłsudski's (1912, 5) description, corresponds to the alveolo-palatal fricative [ɕ], also present in West Sakhalin Ainu, and rendered by Piłsudski as *ś*. The allophone [ɕ] occurs systematically before and after /i/ word-initially, word-finally, and word-internally – e.g. ***ś**ine* [ɕinɛ] < *sine* 'one', *eoći**ś*** [ɛɔt͡ɕiɕ] < *eocis* 'to feel resentful for', *i**ś**o* [iɕɔ] < *iso* 'bear'. It is also present in the environments $i\_C_{plosive}$ and _n – e.g. *hú**ś**ko* < *husko* [huɕkɔ] 'to be old', which is a regular change

in West Sakhalin Ainu (Dal Corso 2021a, 27). Word-initially, provided that it does not precede /i/, [ɕ] can appear in free variation with [s] – e.g. ***ś****uj* [ɕuj] < *su*y 'again', but cf. ***s****uj* [suj] < *su*y, same meaning.

The labiodental fricative /v/, that Piłsudski (1912, 6) generally reports for all the East Sakhalin Ainu dialects he documented, is not present in West Sakhalin Ainu. From the discussion of semivowels (i.e. approximants, § 4.1.7) in Piłsudski (1912, 3) and comparing the examples provided therein with West Sakhalin cognates, I infer that rather than being "only an alternative form of *u*", /w/ could in fact be the underlying phoneme of which [v] and possibly [u] are allophones. Even so, the allophone [v] would still represent a peculiarity of East Sakhalin dialects in contrast with all other varieties and dialects of Ainu. The phonemic status of /w/ or /v/ would need to be corroborated by considering a larger number of examples and /v/ is included in the East Sakhalin Ainu phonemic inventory only provisionally.

The grapheme *h* represents the voiceless glottal fricative /h/. Realization of /h/ as [ɸ] (*f* in orthography) in the environments _u and u_, which is systematic in West Sakhalin Ainu (Dal Corso 2021a, 27), is not found always in East Sakhalin Ainu – e.g. ***h****úxkara* [huxkaɾa] < *huhkara* 'small forest', but ***f****ura* [ɸuɾa] < *hura* 'smell'. Piłsudski (1912, 6) mentions voicing of intervocalic /h/ as [ɦ], but does not specify where it appears. The voiceless velar fricative [x], besides being found at morpheme boundary, occurs in a coda position within words that contain the sequence $C_{plosive}$+$C_{plosive/nasal}$ either synchronically or diachronically. That is, [x] derives from debuccalization and should primarily been regarded as an allophone of the plosive sounds that feature this process – e.g. *hò****x****ḱe* [hɔxkʲɛ] < *hohke* 'to lie down' (cf. Southern Hokkaidō Ainu *hotke*, same meaning). [x] also appears word-finally, as discussed in § 4.1.2, and before [t͡ɕ] within the collective verbal suffix *-hci*. It should be noted that [x] appears word-initially in the corpus only in certain words having either *o* or *u* in the first syllable – e.g. ***x****ośíbi* [xɔɕipi] < *hosipi* 'to return', *xunana* [xunana] <*hunana* 'seeĸ' (cf. again ***h****òx̌ḱe* above). It is still not clear whether this is a piece of evidence for the synchronic or diachronic presence of a fricative consonant distinct from /h/. If this were the case, the phonemic status of [x] in East Sakhalin Ainu would have to be reconsidered.

4.1.5 Affricates

East Sakhalin Ainu has one affricate rendered as *ć*. Piłsudski (1912, 7) says that "the Ainus have no sound of *c* as […] the German *z*" nor a "pure *č*" as in the English *chalk*, and that this sound is "something akin to the sound of the Polish *ć*". This means that the sound represented by *ć* is in its turn different from /t͡s/ and /t͡ʃ/, and possibly close to /t͡ɕ/ which underlies the Polish grapheme *ć*. Furthermore, in

other texts redacted by Piłsudski using the Cyrillic alphabet, the East Sakhalin Ainu affricate is rendered as *ц*, and the occurrence of this grapheme followed by a soft vowel suggests palatalization or retraction. Therefore, I assume that East Sakhalin Ainu has the voiceless retroflex affricate /t͡ʂ/, yet noting that Piłsudski acknowledges great variation among dialects and even among single individuals. The allophone [d͡ʐ] (*ź* in orthography) appears only after nasals – e.g. *únźi* [und͡ʐi] < *unci* 'fire'. I reconstruct this pronunciation given that this is "a sonant variation of *ć*" (Piłsudski 1912, 7).

4.1.6 Liquids

East Sakhalin Ainu has one liquid consonant. With reference to the guide to Sakhalin Ainu phonetics in Dobrotvorskij (1875), Piłsudski (1912, 8-9) lists a variety of possible pronunciations associated with the grapheme *r*. By the end of his excursus, Piłsudski is elusive with regard to the underlying phoneme that *r* subsumes. Exclusively from his endorsing of professor Abbé Rousselot's observation that "*r* in between vowels has its normal sound" I infer that the phoneme associated to *r* is the alveolar tap /ɾ/, if one understands this "normal sound" as the most common one in both Hokkaidō and Sakhalin Ainu.

The only allophones of /ɾ/ that Piłsudski discusses are found at the beginning of words and after the nasal /n/. In these environments, /ɾ/ is articulated simultaneously with the plosive sound /t/ which is however "farther up [on the palate] than the usual *t*". From this, one can infer the presence of either the affricate [t͡ʈɽ] or a single retroflex consonant [ʈ] or [ɽ] (Alonso de la Fuente 2022, 157-9). After the nasal /n/, this affricate may be voiced, thus probably being realized as [d͡ɖɽ]. [t͡ʈɽ] is reported to have a lot of variation from speaker to speaker, with the plosive or trill trait of the affricate being in turns more prominent than the other. Word-final *r* is usually followed by an echoed vowel identical to the last vowel of the root, which is a common trait of most Ainu dialects – e.g. *kisara* [kisaɾa] < *kisara* (underlying form *kisar**) 'ear'. Like plosives (§ 4.1.2), the realization of /ɾ/ word-finally in those rare cases when it is not followed by a copy of the vowel preceding it is often not noted in the orthography – e.g. *máxneku* [maxnɛku] < *mahneku* 'young woman' (for intended *mahnekuh*, underlying form *mahnekur**). In such instances, *r* is most likely debuccalized as it happens in West Sakhalin Ainu (Dal Corso 2021a, 25-6).

4.1.7 Approximants

Provided the doubts on the phonemic status of /w/ (§ 4.1.4), the only approximant of East Sakhalin Ainu is /j/, represented in orthography as *j*.

4.2 Phonological Processes

Piłsudski's orthography is insightful also with regard to phonological processes. Most of the processes attested in the *Materials…* are also encountered in West Sakhalin Ainu sources and, as such, they seem to be distinctive of the Sakhalin Ainu variety in general. Among the phonological processes common to both coasts of Sakhalin we find, for instance, the assimilation of /ɾ/ to [n] preceding /n/, the dissimilation of /i/ to [j] following a vowel, or the insertion of [i] between a word ending in /n/ followed by the clause linker *ike* 'and' whose initial [i] is then dissimilated to [j] (for a list of all phonological processes found in West Sakhalin Ainu see Dal Corso 2021a, 29-40). In this subsection, I concentrate only on those phonological processes suggested by Piłsudski's transliteration of East Sakhalin Ainu that are unattested in West Sakhalin Ainu.

The alternation outlined in (a) provides an example. Orthography in this case suggests the realization of the bilabial nasal /m/ as [ŋ] before the plosive /k/ (i.e. assimilation). However, the change does not mirrored completely in orthography with the expected grapheme *ŋ*, in spite of the fact that Piłsudski does use this grapheme [ŋ]. This sound is reported to be "very rarely met with" (Piłsudski 1912, 5) and the alternation might be ascribed to the fact that /m/ in coda position are often pronounced as [n] in careless, fast speech (Chiri [1942] 1973, 472).

a. /m/ > |n| / _k – e.g. *sá**n**ḱeta* [sankʲɛta] < *sa**m**_-ke_ta* (by-PTV-in) 'by sth.'.

Dissimilation is present in Piłsudski's corpus mostly as regressive dissimilation. Of these instances, the dissimilation of a nasal /n/ or /m/ before the fricative /s/ is outstanding in that it appears far more commonly in East Sakhalin than in West Sakhalin (where it is in fact encountered only once, Dal Corso 2021a, 31)

b. m/ > |j| / _s – e.g. *tú**j** śáta* [tujɕata] < *tu**m**_-sa-ta* (center-?-in) 'in the center of'.

Equally peculiar is the kind of elision that concerns *h* within the collective *-hci* when the suffix is attached to monosyllabic open-syllable verbs (c). This process is, however, not regular and there are a good number of counterexamples in the corpus. This alternation is never

attested in West Sakhalin sources. Elision of *h* may also happen in the possessive suffix *-(i)hi* (d).[9]

c. [x] > Ø | #CV_ – e.g. *nùćí* [nuʧi] < *nu-**[h]**ci* (hear-COLL) '[they] heard'.

d. /h/ > Ø | #C_{nasal}_ – e.g. *kotánu* [kɔtanu] < *kotan-**[h]**u* (3/village-POSS) '[their] village'.

4.3 Transitive Agreement Paradigm

In the Ainu language, person agreement except for third persons is marked by affixes on the verb. In reference grammars of Sakhalin and Hokkaidō dialects, person agreement marking is said to be obligatory, but the data in the *Materials…* contrasts with this general rule, as we seldom find verb forms unmarked for person agreement when they should be. The conditions that make person agreement superfluous are not yet understood and they will not be addressed here. Rather, in this section I focus on the formal encoding of participant interactions that we observe in the transitive agreement paradigm of East Sakhalin Ainu, which shows yet again interesting points of divergence with West Sakhalin Ainu.

One felicitous generality that can safely be made for all varieties and dialects of Ainu is that the language exhibits a disuniform morphological alignment. For West Sakhalin Ainu, Dal Corso (2021a, 83-9) lists the following forms. First person singular displays nominative-accusative alignment (*ku-* for S/A and *en-* for O); second person singular and plural feature direct alignment (2S *e-* and 2P *eci-* for all S/A/O functions). The same direct alignment can be recognized for third persons (i.e. non-speech act participants) that are always zero-marked.[10] West Sakhalin Ainu has one more set of person agreement forms: *-an*, *an-*, and *i-* (for S, A, and O respectively) exhibiting tripartite morphological alignment, which indicate, under specific semantico-pragmatic circumstances, a mismatch between the referenced

9 The underlying form of the possessive suffix is assumed to contain the vowel *i*, since this vowel appears in the most varied phonological environments (Dal Corso 2021a, 32). Furthermore, the first vowel that appears in the realization of the suffix as *-VhV*, on consonant-final nouns, is treated as a case of optional insertion to solve the *Ch* cluster. This is assumed in light of the fact that forms such as *kotanuhu* and *kotanhu* 'one's village' are found in free distribution in the corpus.

10 The suffix *-(a)hci* may be optionally added to express plurality or collectivity of a third person subject or object. Contrary to how it is discussed in Dal Corso (2021a, 88-9), who follows Murasaki's (1979, 49) interpretation of this suffix, *-(a)hci* is not a person agreement marker and it should rather be considered a marker of number, as suggested in Sakaguchi (2018).

participant and the agreement form used and which, for this reason, Dal Corso (2023) discusses as markers of participant referentiality mismatch (PRM). All these affixes are proper of East Sakhalin Ainu too.

A first important difference with West Sakhalin Ainu lies in the presence of a separate set of agreement affixes for the exclusive first person plural, which in East Sakhalin Ainu are *ci-* for A, *-as* for S, and *in-* for O, unattested in West Sakhalin Ainu (Dal Corso 2021a, 84).[11] Like participant-referentiality-mismatch affixes, this set too exhibits tripartite morphological alignment. These markers are cognate forms of the Southern and Central-Eastern Hokkaidō *ci-/-as/un-*,[12] with the O form *in-* representing the only point of divergence with these latter dialect groups (Sato 1985). The prefix *ci-* is otherwise present in West Sakhalin, but only as a resultative marker, a function also attested in Piłsudski's corpus synchronically to the person agreement one.

The transitive agreement paradigm presented in Table 1 illustrates the affix combinations employed in a transitive clause – that is, when it is necessary to reference both participants covering the A and O functions. When a third person interacts with a first or second person (i.e. a speech act participant, SAP, which may also be cross-referenced on the verb via participant-referentiality-mismatch agreement), the only overt prefix on the verb is the one that references the SAP object (if the third person is the subject) or the SAP subject (if the third person is the object) (1)-(2). Otherwise, if both A and O are third persons, the verb appears unmarked for person (3).

(1) *Kamúi utara inránu.*
Kamuy-utara ***in****-ranu[p].*
god-COLL **1P.O.EXCL-3P.A/**love
'The gods love me.' (Piłsudski 1912, 86)

(2) *Ećítom óxkajo tarap ekorō!*
Ecitom ohkayo tarap **e-***koro-[h]o?*
Ecitom young.man strap **2S.A-3S.O/**have-PK
'Young man of Ecitom, have you got a strap?' (Piłsudski 1912, 114)

(3) *Nea niśpa emuś stómuśite.*
Nea nispa emus sitomusite.
that noble.man sword **3S.A/3S.O/**put.on.side
'That noble man tied a sword to his belt.' (Piłsudski 1912, 77)

11 With the exception of one isolate instance of *ci-*.

12 See e.g. Tamura 2000; Bugaeva 2012; Tamura 2010; Takahashi 2018.

When both referents of A and O are speech act participants, a combination of the relative A and O prefix forms available for each person are attested. One example is the combination *e-i-* that expresses a second person singular subject (*e-*) acting upon a first person singular object marked by participant-referentiality-mismatch agreement (*i-*).

(4) *Eirájki.*
E-i-*ray-ki.*
2S.A-PRM.O-die-CAUS
'You kill me.' (Piłsudski 1912, 125)

It is, however, more common to find affix combinations or portmanteau forms that cannot be felicitously separated into the two distinct A and O agreement forms otherwise employed for the persons involved. One example of this is the prefix *eci-* used to mark a first person singular subject acting upon a second person plural object (in example (5) a beneficiary indirect object), where there is no trace of the expected first person singular subject prefix *ku-*. The same portmanteau prefix is present in Southern Hokkaidō Ainu. Another example is the combination of affixes *in-…-an* used to mark a second person plural subject acting upon a first person plural object (6),[13] where together with the expected O form *in-* there appears *-an* which is never used to cross-reference a second person plural subject on transitive verbs in combination with a third person object nor on intransitive verbs.

(5) *Ećikóndehé.*
Eci*-kor_-te-he.*
1S.A>2S.IO-3S.O/have-CAUS-PK
'I gave it to you.' (Piłsudski 1912, 77)

(6) *Inránupan.*
In*-ranup-****an****.*
2P.A>1P.O.EXCL-love-**2P.A>1P.O.EXCL**
'You love me.' (Piłsudski 1912, 86)

In Table 1, the forms within the red square are those used for speech-act-participant interactions. Shaded areas indicate reflexive or reciprocal person relations that are marked with dedicated suffixes. Forms in bold are attested in East Sakhalin Ainu but not in West Sakhalin Ainu, while forms within square brackets are encountered

13 In the example the exclusive first person plural is rendered in translation as 'me'. The use of first person plural forms as well as participant-referentiality-mismatch agreement to cross-reference the speaker is common in folklore.

in West Sakhalin Ainu but not in Piłsudski's corpus. All other forms are attested in both dialects or in neither one of them (N/A). Finally, abbreviations in small caps indicate in which dialect(s) each form is attested (TA for Tarayka, HU for Hunup, AY for Ay, and TU for Tunayci).

Table 1 Transitive agreement paradigm of Sakhalin Ainu

	1s.o	1p.o	2s.o	2p.o	prm.o	3s/p.o
1s.a			*eci-* TA, AY	***eci-*** TA *eci-…-yan* TA		KU- TU, HU
1p.a			***eci-*** HU	N/A		***ci-*** TA, HU
2s.a	*[en-]*	N/A			***i-*** TU ***e-i-*** HU, TA, TU	***e-*** HU, TA, TU
2p.a	*[en-…-yan]*	in-…-(y)an ta			*[i-…-yan]* *[eci-i-]*	*eci-* TA, TU
prm.a			*an-e-* AY, HU, TU	*[eci-…-yan]*	***an-i-*** HU, TU	***an-*** AY, HU, TA, TU
3s/p.a	*[en-]*	***in-*** HU, TA, TU	*e-* HU, TA, TU	*eci-* TA, TU	*i-* TU, HU	-

Let us consider the forms that are peculiar to East Sakhalin Ainu, highlighted in bold in Table 1. The first person plural exclusive O form *in-*, besides of course being found to mark the object with a third person subject, also appears in combination with *-(y)an* to mark a 2P.A>1P.O.EXCL relation. This affix combination is not attested in West Sakhalin Ainu since in this dialect the set of first person exclusive affixes is not attested. However, this morphological layout is in line with other combinations within the dimension of SAP-relations: see, for instance, *en-…-yan* for 2P.A>1S.O or *i-…-yan* for 2P.A>PRM.O. As noted in Dal Corso (2021a, 86), in these cases of combined affixation the prefix always encodes the object while *-(y)an* always refers to the subject. Furthermore, one of the referenced participants, may it be a subject or an object, is invariably a second person plural. *In-…-(y)an* attested in East Sakhalin Ainu is no exception to this. I can, therefore, provisionally conclude that combined affixation for SAP interactions involving a second person plural was a common feature of West Sakhalin Ainu and at least the northern dialect of Tarayka on the east coast. The lack of attestations of this type of agreement in the other East Sakhalin dialects prevents me from saying that it was in fact a common feature of the Sakhalin Ainu variety in general.

Other similarities can be observed between East Sakhalin Ainu and Hokkaidō Ainu dialects. One example is the portmanteau prefix *eci-*, employed to encode 1P.A.EXCL>2S.O and 1S.A>2P.O interactions. In West Sakhalin Ainu, *eci-* may also encode a 1S.A>2S.O interaction. Given the formal equivalence, in these cases the correct interaction intended is necessarily retrieved from context. This formal unity of agreement in first-second person interactions is also a property of southern Hokkaidō Ainu dialects (Tamura 2000, 59; Bugaeva 2012, 474). Another similarity with the southern Hokkaidō Ainu agreement paradigm is the combination *a-i-* for the PRM.A>PRM.O interaction. This combination is not attested in West Sakhalin Ainu, though the conditions for its use are seldom present – in such cases only the subject marker *an-* is found, while the object marker *i-* is suppressed (see Dal Corso 2021a, 87; 2023 specifically for the relevance of this discrepancy between eastern and western dialects for the study of the impersonal passive construction).

(7) *'Ekasi-'oro-wa […]* ***'an**-e-ko-_itak_-kara*.
old.man-place-ABL/ELA **PRM.A>PRM.O-**APPL-APPL-3S.O/speak-make
'I was instructed about it by [my] old man.'
(Murasaki 1976, 94; Dal Corso 2021a, 389)

In (7), from the western dialect of Rayciska, the prefix *an-* cross-references the oblique agent *ekasiorowa* in the impersonal passive construction. However, the object, which in the rest of this folktale is consistently marked with PRM agreement, does not appear marked with the expected prefix *i-*.

4.4 Valency-Changing Strategies

Sakhalin Ainu showcases a number of verbal affixes that mark valency changes for the predicate. Among these we find, for instance, the valency-decreasing reciprocal *u-* and antipassive *i-* or the valency-increasing causative suffixes *-te/-re/-ke/-ki/-ka/-yara*. In this section, I concentrate on three prefixes: the 'applicatives' *e-* and *ko-*, and the reflexive *yay-*.[14] Specifically, the applicative *ko-* and the *reflexive yay-*, together with their primary valency-changing function, less frequently also exhibit what appears to be a valency-preserving function. As for the applicative *e-*, it appears on the verb *a* 'sit' that follows a verb in a seeming aspectual function. The valency-preserving

14 Sakhalin Ainu has one more applicative prefix (*o-*) and one more reflexive prefix (*si-*), which will not be discussed here.

and aspect-related functions of these prefixes is rarely encountered in the West Sakhalin Ainu corpus, but is far more common in the *Materials*…. Although this uneven distribution in the two sources may be indicative of a dialectal difference between eastern and western dialects, it may also be informative of language development or better, in this particular case, of language erosion. The scant attestations of this type of verbal morphology in the West Sakhalin data can mean that this secondary function of *e-*, *ko-* and *yay-* fell out of use with time probably as a consequence of the contact of Ainu with Japanese which intensified after the relocation of native speakers to Hokkaidō (§ 3). Indeed this case of erosion would be in line with the loss of productivity of verbal morphology documented for Ainu in general.

4.4.1 Applicatives *e-* and *ko-*

Through applicativization an oblique is promoted to argument (either a direct or indirect object) and included in the base verb's subcategorization (Kroeger 2004). Applicativization results in the transitivization of intransitive verbs and in the ditransitivization of (mono-) transitive verbs, it does not affect the promoted argument's thematic role, and it functions at discourse level as a focusing strategy for an otherwise peripheral element. The three applicative forms of Sakhalin Ainu are not interchangeable and their choice depends on the thematic role of the promoted argument. Example (9) shows the applicative version of example (8), where the oblique denoting instrument is promoted to indirect object of the verb *tuhseka* 'to kick' – the promotion is marked by the applicative *e-*.

(8) *Suma kema ani tuhse-ka.*
rock foot 3S.O/with 3S.A/3S.O/jump-CAUS
'(S)he kicked the rock with [his/her] foot.'

(9) *Suma kema e-tuhse-ka.*
rock foot **APPL-**3S.A/3S.O/3S.IO/jump-CAUS
'(S)he kicked the rock with [his/her] foot.' (Murasaki 1979, 32)

Dal Corso (2021b) proposes a unitary explanation for both the applicative and non-applicative functions of the prefixes *e-* and *ko-*, under Hopper and Thompson's (1980) theory of semantic transitivity and defines *e-* and *ko-* as morphological markers of high transitivity. However, upon a closer look at the data, Dal Corso (2021b) reports this non-applicative use of the two prefixes as too common a feature of *e-* and *ko-* than it actually is. Therefore, in an example like (10), the

prefix *e-* is better recognized as a phonological variant of the deictic *he-* 'head', which indicates that the motion event takes place up towards the goal of the action (i.e. the house). Most of the tokens featuring an alleged non-applicative *e-* discussed in Dal Corso (2021b) can be explained this way.

(10) *Ćíše oxta eśirep án.*
Cise *or_-ta* ***(h)e-****sirepa-an.*
house place-LOC/LAT **DEI.HEAD-**arrive-PRM.S
'I arrived up to the house.' (Piłsudski 1912, 160)

Albeit this revision to Dal Corso (2021b), some of the examples taken into account therein still cannot be easily explained as featuring the deictic *he-*. In those instances, we see *e-* prefixed to the verb *a* 'sit', which fulfills an aspectual function (11)-(13).

(11) *Tá maxneku níśkoro oxta etáras ea.*
Ta *mahneku[h]* *niskoro-or_-ta* *etaras* ***e-****a.*
that young.woman cloud-place-LOC/LAT 3S.S/stand **e-**sit.PC/IPFV(?)
'Right there that young woman was standing on the clouds.' (Piłsudski 1912, 90)

(12) *Suḱe ea kánne ahupan.*
Suke ***e-****a* *kanne* *ahup-an.*
3S.S/cook **e-**sit.PC/IPFV(?) ADV enter.PC-PRM.S
'I entered as [my wife] was presently cooking.' (Piłsudski 1912, 134)

(13) An-kućaha orova pā numa eá-kusu án.
An-kuca-ha-oro-wa *paa numa* ***e-****a* *kusu an.*
PRM.PSR-hut-POSS-place-ABL/ELA smoke 3S.S/rise **e-**sit.PC/IPFV(?) PROG.PC
'And there, smoke was rising from my hut.' (Piłsudski 1912, 134)

Here *ea* conveys a presentative meaning – that is, it brings attention to a specific participant in the event (which in the examples above is the subject of the notional verb). The syntactic layout where *ea* follows another verb makes it more likely for *e-* to be an applicative than the deictic *he-;* under such reading *e-* would introduce a direct object of the verb *a* (i.e. the full zero-nominalized clause containing the notional verb) possibly with the semantic role of aim/purpose. At any rate, this use of *e-* highlights an exceptional functional use of the verb *a* 'sit'. This verb is used as an aspectual auxiliary to mark actional perfect in Hokkaidō Ainu dialects (Bugaeva 2012), but it is never attested with this function in West Sakhalin Ainu. Another possible analysis is one that sees the segment *a* of the form *ea* as the verb *an* 'exist', whose final nasal has been elided. In this case, the syntax of the V+*ea* construction would be less problematic since *an* following a verb is commonly found as an auxiliary marking imperfective

aspect (Dal Corso 2022). One piece of evidence to support this last interpretation is the fact that the base predicate appearing with *ea* is either a state (11) or an activity (12)-(13) – that is, an atelic predicate. The imperfective *an* derives a continuous aspect reading for the predicate when this latter is an atelic predicate (Dal Corso 2022, 70-4). The co-occurrence of *ea* with progressive aspect *kusu an*, as in (13), is also non-trivial. It further supports the hypothesis that *ea* may in fact be an alloform of *ean* since progressivity is in line with the continuous aspect reading.

One more observation concerns the applicative *ko-*. A proper applicative function for this prefix is difficult to retrieve in three instances in Piłsudski (1912), where *ko-* is attached to a zero-valency verb indicating natural conditions, like *sistono* 'dawn' (14).

(14) *Mokoro poka ankojákuś, réuśi án. Kośíštonó tonosḱe kooman-kane…*
Mokoro *poka* *an-koyakus* *reusi-an*
sleep.NMLZ at.least PRM.A-3S.O/not.be.able pass.the.night-PRM.S
I could not even sleep, but I spent the night [there].
ko-*sir-tono* *to-noske* *ko-oman* *kane…*
ko-appearence-be.bright day-middle APPL-3S.A/3S.O/go.PC ADV
'Having [the day] gotten to midday when it was plainly bright…'
(Piłsudski 1912, 146)

An applicative function (i.e. that of adding a goal direct object) can be recognized for *ko-* prefixed to *oman* 'go'; here the subject of the verb can be interpreted as an implied 'the day'. However, the same cannot be said for *ko-* prefixed to *sistono* 'dawn' since it is difficult to retrieve an antecedent within the context that has a semantic role compatible with those encoded by *ko-*. Therefore, *ko-* in these cases seems to show the high transitivity function discussed by Dal Corso (2021b). The instances of this use of *ko-* are very scant and I cannot provide any further insights at the moment.

4.4.2 Reflexive *yay-*

If the valency-preserving function of the applicative *e-* and *ko-* has already been investigated, that of the reflexive prefix *yay-* remains to be addressed. The prefix *yay-* attaches to transitive (usually actional) base verbs and indicates that the action denoted by the verb falls back on the agent-subject, which thus corresponds also to the patient-object. As such, reflexivization marked by *yay-* is a valency-decreasing strategy that results in the intransitivization of the base verb. The change in valency may become clear from the locus of affixation of person agreement on originally mono-transitive verbs like *reske* 'raise' – see the participant-referentiality-mismatch agreement

suffix *-an* reserved for the S argument of intransitives in (16), which is the reflexive version of (15).

(15) *Mahpoo-ho-hcin* *reske-hci.*
3.PSR/daughter-POSS-COLL 3P.A/3S.O/raise-COLL
'They raised their daughter.'
(Murasaki 1976, 28; Dal Corso 2021a, 281)

(16) ***Yay-****reske-'an.*
REFL-raise-PRM.S
'I raised myself (= I grew up alone).'
(Murasaki 2010, 72; Dal Corso 2021a, 415)

Sometimes the locus of affixation of *yay-* is on a noun that has been incorporated to the base verb. In these cases, the reflexive *yay-* cross-references the possessor of the incorporated noun and is co-indexed with the subject of the base verb (the "quasi-possessive reflexive" of Bugaeva, Kobayashi 2022, 526). Example (17) provides an illustration. Here the base transitive *koro* 'have' incorporates the whole nominal constituent *yaycise* 'own house'. This means that, although reflexivization is normally a valency-decreasing strategy, when *yay-* gets incorporated together with the possessee noun it does not affect the base verb's valency, which is decreased only once by the process of incorporation.

(17) *Sinene-h-ponne* *yay-cise-koro.*
alone-?-ADV? 3S.S/**REFL.PSR-**house-have
'He had his own house.' (Piłsudski 1912, 238)

This use of *yay-* is limited to incorporated inalienable nouns (most commonly *cise* 'house', *kotan* 'village', *cara* 'mouth', *ram* 'soul' in Piłsudski's corpus) and it never appears to mark a third person possessor co-indexed with the verb's subject on unbound possessed noun forms.[15] Third person possessors, in line with the formal unmarkedness of third persons witnessed in verb person agreement (§ 4.3), are always zero-marked, and the eventual co-indexicality with the subject is retrieved only through contextual information – that is, the form **yaymahpoho* for 'his own daughter' is never encountered. Thus the possible double interpretation out of context for *mahpoho* in (14).

15 Indeed prefixation of reflexive *yay-* and *si-* as well as of the reciprocal *u-* is possible on locative nouns (e.g. *u-sam* RECP-next.to 'next to each other', which Bugaeva and Kobayashi (2022, 525) call the quasi-possessive reciprocal type).

(18) *Nea henke nea mahpo-ho cip_ ‘or_-ta*
this old.man this **3.PSR**/daughter-POSS boat place-LOC/LAT
‘aa-re.
3S.A/3S.O/sit.PC-CAUS
‘The old man had his (somebody else’s)/his own daughter sit in the boat.’
(Murasaki 1976, 31; Dal Corso 2021a, 288)

Even more strikingly, *yay-* seldom appears prefixed to base intransitives whose subject is a third person, like *ahkas* ‘walk’ in (19) or *a* ‘sit’ in (20). In these instances as well, the reflexive does not affect the verb’s valency, but rather seems to convey a dislocative or andative meaning and stress that the event takes place away from the contextual center of attention.

(19) *Kotánḱeś-un nispa jaj áxkaś jaxka…*
*Kotankes un nispa **yay-**ahkas yahka…*
Kotankes 3S.A/3S.O/live.in noble.man **REFL-**3S.S/walk though
‘A noble man who lived in Kotankes strolled by himself [away from the village], but…’ (Piłsudski 1912, 76)

(20) *Pon kapári-óxkajo maćí turá kotán-hu ónne xośíbi kusú, poró ciś áni irúra-utará iván-níu irúra, móromaxpo jajá ajn(u) énko ćipó.*
Pon kapari ohkayo mat_-i tura
be.small plaice young.man 3.PSR/wife-POSS 3S.O/together
kotan-hu-or_-ne hosipi kusu poro cip_ ani
3.PSR/village-POSS-place-ALL 3P.S/return CAU.FIN be.big boat 3S.O/with
i-rura utara iwan-n-iw i-rura
ANTIP-3P.S/transport people six-EP-CLN.ANI ANTIP-3P.S/transport
moromahpo yay-a aynu enko cip-o.
young.woman **REFL-**3S.S/sit.PC man half 3P.S/boatget.in
‘The young male plaice returned to his village together with his wife, so six rowers with a big boat rowed, the young woman sat by herself [on one side, and] half of the men led the boat.’ (Majewicz 1998b, 287)

What I have discussed about verbal morphology in this section shows how the language data in the *Materials…* presents descriptivists working on Ainu with important issues that have to do not only with the language’s structure but also with the semantic that influence morphosyntax. Being semantics, together with pragmatics, an area of the Sakhalin Ainu grammar that has rarely been investigated to date, the data in Piłsudski’s corpus proves to be a precious resource.

5 Conclusion

In this paper, I focused on a number of peculiar grammatical features attested in the East Sakhalin dialects of Ainu documented by the Polish ethnographer and linguist Bronisław Piłsudski. Among the texts collected by Piłsudski, twenty-seven tales of traditional folklore constitute the corpus published as *Materials for the Study of the Ainu Language and Folklore*, which was the main reference for this study. The areas of East Sakhalin Ainu that I surveyed in this paper were the phonemic inventory and phonotactic rules, some phonological alternations, and verbal morphology. In particular, with regard to verbal morphology, I addressed the transitive agreement paradigm and the valency-preserving functions of the applicative prefixes *e-* and *ko-* and of the reflexive prefix *yay-*. These areas of the East Sakhalin Ainu grammar show interesting points of divergence with the neighbouring West Sakhalin Ainu dialects, thus providing important insights and inputs for future research about dialectal differences within the Sakhalin Ainu variety and about the possible historical development of the language. The main aim of this paper was to underline the importance of analysing old sources on Ainu, of which Piłsudski's *Materials…* is just an example, for the description and the study of the history of minority, endangered languages of Japan.

Abbreviations

1P	first person plural
1S	first person singular
2P	second person plural
2S	second person singular
3	third person
3P	third person plural
3S	third person singular
A	transitive subject
ABL	ablative
ADV	adverb
ANTIP	antipassive
APPL	applicative
CAU.FIN	causal-final linker
CAUS	causative
CLN.ANI	animate numeral classifier
COLL	collective
DEI	deictic
ELA	elative
EP	epenthetic consonant
EXCL	exclusive
IO	indirect object
IPFV	imperfective
LAT	lative
LOC	locative
O	object
PC	paucal
PK	personal knowledge evidential
POSS	possessive
PRM	participant referentiality mismatch marker
PROG	progressive
PSR	possessor
RECP	reciprocal
REFL	reflexive
S	intransitive subject

Bibliography

Alonso de la Fuente, J.A. (2022). "The Ainu Language Through Time". Bugaeva, A (ed.), *Handbook of the Ainu Language*. Berlin; Boston: De Gruyter Mouton, 147-84.

Asai Tōru 浅井享 (1974). "Hōgen Bunrui 'Ainugo Shohōgen no Kurasutā Bunseki'" 方言分類「アイヌ語諸方言のクラスター分析」(Dialectal Classification 'a Cluster Analysis of the Many Dialects of the Ainu Language'). *Hoppō Bunka Kenkyū Daiichigō* 北方文化研究第一号 (Studies on the Cultures of the North no. 1). Sapporo: Hokkaidō University, 45-136.

Bugaeva, A. (2012). "Southern Hokkaidō Ainu". Tranter, N. (ed.), *The Languages of Japan and Korea*. London: Routledge, 461-99.

Bugaeva, A.; Satō Tomomi 佐藤知己 (2021). "A Kuril Ainu Glossary by Captain V.M. Golovnin (1811)". *International Journal of Eurasian Linguistics*, 3(2), 171-216.

Bugaeva, A.; Kobayashi Miki 小林美紀 (2022). "Verbal Valency". Bugaeva, A. (ed.), *Handbook of the Ainu Language*. Berlin; Boston: De Gruyter Mouton, 515-48.

Chiri Mashiho 知里真志保 [1942] (1973). *Ainugohō kenkyū* アイヌ語法研究 (A Grammatical Sketch of Ainu). Reprinted (1973) in *Chiri Mashiho chosakushū 3* 知 里真志保著作集 3 (Selected Writings of Chiri Mashiho 3). Tōkyō: Heibonsha, 455-586.

Dal Corso, E. (2021a). *The Language and Folklore of West Sakhalin Ainu. A Re-Edition of Kyōko Murasaki's "Karafuto Ainugo" with Translation and Grammatical Notes*. München: Lincom Publishing.

Dal Corso, E. (2021b). "Polysemy and Apparent Polyfunctionality of the Sakhalin Ainu Prefixes e- and ko-: Insights on Applicativization and Aspect". *International Journal of Eurasian Linguistics*, 3(2), 217-44. http://dx.doi.org/10.1163/25898833-00320003.

Dal Corso, E. (2022). "Some Insights on the Clause Linker *-teh* of Sakhalin Ainu Through a Consideration of Aspect". Ragagnin, E.; Khabtagaeva, B. (eds), *Endangered Languages of Northern Asia*. Leiden; Boston: Brill, 57-80. Languages of Asia Series 28.

Dal Corso, E. (2022a). "A Comparative Analysis of an East Sakhalin Ainu Folktale Collected by Bronisław Piłsudski". *Annali di Ca' Foscari. Serie orientale*, 58(1), 907-46. http://doi.org/10.30687/AnnOr/2385-3042/2022/01/030.

Dal Corso, E. (2023). "On the Object Individuation Function of the East Sakhalin Ainu Impersonal Passive". *Folia Linguistica*, 57(3), 563-92.

Dobrotvorskij, M.M. (1875). *Ajnsko-Russkij Slovar' Айнско-Русскій Словарь* (Ainu-Russian Dictionary). Kazan: Kazan University Press.

Hattori Shirō 服部四郎; Chiri Mashiho 知里真志保 (1960). *Ainugo hōgen jiten* アイヌ語方言辞典 (An Ainu Dialect Dictionary). Tōkyō: Iwanami Shoten.

Hopper, P.J.; Thompson, S.A. (1980). "Transitivity in Grammar and Discourse". *Language*, 56, 251-99.

Kitahara Jirōta Mokottunas 北原次郎太モコットゥナ (ed.) (2013). *Wada Bunjirō – Karafuto Ainu Setsuwashū 1* 和田文治郎 – 樺太アイヌ説話集 1 (Wada Bunjirō – Collection of Karafuto Ainu Folklore 1). Sapporo: Aynu Teetawanoankur Kanpinuye Cise.

Kitahara Jirōta Mokottunas (ed.) (2014). *Wada Bunjirō – Karafuto Ainu Setsuwashū 2* 和田文治郎 – 樺太アイヌ説話集 2 (Wada Bunjirō – Collection of Karafuto Ainu Folklore 2). Sapporo: Aynu Teetawanoankur Kanpinuye Cise.

Kitahara Jirōta Mokottunas (ed.) (2016). *Wada Bunjirō – Karafuto Ainu Setsuwashū 3* 和田文治郎 – 樺太アイヌ説話集 3 (Wada Bunjirō – Collection of Karafuto Ainu Folklore 3). Sapporo: Aynu Teetawanoankur Kanpinuye Cise.

Kitahara Jirōta Mokottunas (ed.) (2017). *Wada Bunjirō Karafuto Ainu Gengo-Bunkashi – Shussan, Ikuji, Tsūkagirei* 和田文治郎樺太アイヌ言語・文 化誌 – 出産・育児・通過儀礼 (Wada Bunjirō Writings on the Ainu Language and Culture – Birth, Childcare, Rite of Passage). Sapporo: Aynu Teetawanoankur Kanpinuye Cise.

Kitahara Jirōta Mokottunas (ed.) (2019). *Akita Bunjirō – Karafuto Ainu Setsuwashū 4* 和田文治郎 – 樺太アイヌ説話集 4 (Wada Bunjirō – Collection of Karafuto Ainu Folklore 4). Sapporo: Aynu Teetawanoankur Kanpinuye Cise.

Krasheninnikov, S.P. (1755). *Opisanije zemli Kamchatki: tom" vtoroj* Описаніе земли Камчатки: том второй (Description of the Land of Kamchatka – Second Volume). St. Petersburg: Imperial Academy of Sciences.

Kroeger, P. (2004). *Analyzing Syntax: A Lexical-functional Approach*. Cambridge: Cambridge University Press.

Majewicz, A.F. (ed.) (1998a). *The Collected Works of Bronisław Piłsudski*. Vol. 2, *Materials for the Study of the Ainu Language and Folklore (Cracow 1912)*. Berlin: De Gruyter Mouton.

Majewicz, A.F. (ed.) (1998b). *The Collected Works of Bronisław Piłsudski*. Vol. 3, *Materials for the Study of the Ainu Language and Folklore 2*. Berlin: De Gruyter Mouton.

Murasaki Kyōko 村崎恭子 (1976). *Karafuto Ainugo – Shiryō* カラフトアイヌ語:資料 (Karafuto Ainu. Materials). Tōkyō: Kokusho Kankōkai.

Murasaki Kyōko (1979). *Karafuto Ainugo – Bunpō* カラフトアイヌ語:文法 (Karafuto Ainu. Grammar). Tōkyō: Kokusho Kankōkai.

Murasaki Kyōko (2001). *Karafuto Ainu no Mukashibanashi – Asai Take Kōjutsu* 樺太アイヌの昔話 – 浅井タケ口述 (Folklore of the Ainu of Karafuto. Dictated by Asai Take). Tōkyō: Sōfūkan.

Murasaki Kyōko (2010). *Karafuto ainu no minwa (uchashikuma) – Wenenekaipe monogatari sanpen* 樺太アイヌの民話 (ウチャシクマ) ウェネネカイペ物語 3編 (Sakhalin Ainu Folktales (ucaskuma) – Wenenekaype). Tōkyō: Tōkyō Gaikokugo Daigaku Ajia-Afurika Gengo Bunka Kenkyūjo.

Murasaki Kyōko (2013). "Karafuto Ainugo no Kenkyū ni Okeru B. Piusutsuki Rōkan Saisei no Kōseki" 樺太アイヌ語の研究におけるB.ピウスツキ蠟管再生の功績 (Merits of Sound Reproduction on B. Piłsudski's Wax Phonograph Cylinders for Sakhalin Ainu Study). *Pōrando no ainu kenkyūsha Piusutsuki no shigoto* ポーランドのアイヌ研究者ピウスツキの仕事 (The Work of the Polish Ainu Scholar Piłsudski). Sapporo: Hokkaido University Collection of Scholarly and Academic Papers, 55-60.

Piłsudski, B. (1912). *Materials for the Study of the Ainu Language and Folklore*. Cracow: Imperial Academy of Sciences.

Sakaguchi Ryō 阪口諒 (2018). "Karafuto Ainugo tōnanbu hōgen ni okeru fukusū hyōji -ahsin -ahsi" 樺太アイヌ語東南部方言における複数表示 (Plural Markers in Southeast Sakhalin Ainu: Focusing on -ahsin and -ahsi) [Unpublished Master's Thesis]. Chiba: Chiba University Graduate School of Humanities and Studies on Public Affairs.

Sato Tomomi 佐藤知己 (1985). "The First Person Objective Affix IN- in the East Coast Dialects of Sakhalin Ainu". *Proceedings of the International*

Symposium on B. Piłsudski's Phonographic Records and the Ainu Culture. Sapporo: Hokkaidō University, 157-67.

Takahashi Yasushige 高橋靖以 (2018). アイヌ語十勝方言文法概説 *Ainugo Tokachi hōgen bunpō gaisetsu* (Grammar Sketch of the Tokachi dialect of Ainu). Sapporo: Aynu Teetawanoankur Kampinuye Cise.

Tamura Masashi 田村雅史 (2010). アイヌ語白糠方言記述 *Ainugo Shiranuka hōgen kijutsu* (Description of the Shiranuka dialect of Ainu). Chiba: University of Chiba.

Tamura Suzuko 田村すずこ (2000). *The Ainu Language*. Tōkyō: Sanseido.

Tangiku Itsuji 丹菊逸治 (2022). "Differences Between Karafuto and Hokkaido Ainu Dialects". Bugaeva, A. (ed.) *Handbook of the Ainu Language*. Berlin; Boston: De Gruyter Mouton, 329-62.

Philological and Linguistic Analysis Working together
Exploring Language Diversity in Japan
edited by Paolo Calvetti

Reflexes of Proto-Ryukyuan Mid Vowels in *Haedong Chegukki*

Marc Miyake
Independent scholar

Abstract Maner Thorpe reconstructed only two mid vowels in Proto-Ryukyuan: *e and *o. He also reconstructed nonphonemic aspiration before nonhigh vowels in a daughter of PR, Proto-Amami-Okinawan. Leon Serafim and Shinzato Rumiko build upon his reconstruction, positing intermediate stages between it and their phonetic interpretation of the Old Okinawan of the *Omoro sōshi*. However, an examination of the earlier stage of Okinawan recorded in *Haedong chegukki* reveals that (1) Proto-Ryukyuan had *əj and *a(:)j in addition to *e and (2) aspiration before nonhigh vowels was still in progress as late as the fifteenth century CE.

Keywords Okinawan. Ryukyuan. Haedong chegukki. Vowels. Aspiration.

1 Introduction

Forty years have passed since Thorpe (1983, 31) reconstructed two mid vowels in Proto-Ryukyuan (PR), *e and *o.

PR *e is a continuation of Proto-Japonic (PJ) *e that merged with PJ *əj and PJ *aj.[1] PR *o is a merger of PJ *o and *ə.

1 Although some scholars (e.g. Martin 1987) write -i instead of -j, I prefer to distinguish between vowel sequences and vowel-glide sequences in Japonic reconstructions. *ai could be mistaken as a sequence of two syllables *a and *i that is somehow different from Korean /aj/.

Edizioni Ca'Foscari

Ca' Foscari Japanese Studies 24
Linguistics and Language Education 4
e-ISSN 2610-9395 | ISSN 2610-8976
ISBN [ebook] 978-88-6969-845-3 | ISBN [print] 978-88-6969-846-0

Peer review | Open access
Submitted 2023-09-22 | Accepted 2024-03-12 | Published 2024-08-28

DOI 10.30687/978-88-6969-845-3/004

In the Old Okinawan (OOK) language of the *Omoro sōshi* おもろさうし (Book of *Omoro* Songs) (1531-1623) as reconstructed[2] by Serafim and Shinzato (2021, 27, 34), PR mid *e raised to near-high [ɪ] except after [ʔ] and [j], where it raised even further to true high [i].

One might expect PR *o to have raised to [ʊ], the back counterpart of front near-high [ɪ], in OOK. And indeed Serafim and Shinzato (2021, 29) do reconstruct near-high *ʊ as a reflex of PR mid *o in pre-*Omoro sōshi* OOK. However, in *Omoro sōshi*, *ʊ raised to true high [u], though stops like *k that were aspirated before nonhigh vowels retained their aspiration in OOK (Serafim, Shinzato 2021, 28) (1).

(1) pre-ooĸ *ko [kʰɔ][3] > *kʰʊ > ooĸ [kʰu]

OOK [kʰu] with aspirated [kʰ] contrasted with OOK [ku] with unaspirated [k] from PR *ku.

Phonemic aspiration was absent from PR. Thorpe (1983, 54) reconstructs nonphonemic aspiration of stops before and between nonhigh vowels in Proto-Amami-Okinawan, an intermediate stage between PR and OOK.[4]

Haedong chegukki 東國諸國紀 (Record of Countries across the Sea to the East; HC) contains an appendix dated 1501 recording an early variety of Okinawan that appears to be a late intermediate stage between Proto-Amami-Okinawan and the OOK of *Omoro sōshi*. I will call that variety *Haedong chegukki* Okinawan (HCO).

In this paper, I will demonstrate that the reflexes of PR mid vowels in HCO do not quite fit the Thorpe/Serafim/Shinzato (TSS) reconstruction that I have just outlined. HCO may contain a vowel-glide sequence /əj/ in some instances where the TSS model predicts a reflex of PR *e. There is even one instance of /a(:)j/ instead of a reflex of *e. Moreover, aspiration is often absent in HCO where the TSS model predicts it. I propose a revision of the TSS model, reconstructing *əj and *a(:)j in PR, and regarding aspiration as an innovation in progress in HCO rather than as a *fait accompli* at the Proto-Amami-Okinawan or PR level.

2 I omit asterisks for reconstructed interpretations of written forms such as OOK as recorded in *Omoro sōshi* or what I will call *Haedong chegukki* Okinawan. Asterisks only indicate reconstructions of unwritten forms.

3 The phonetic interpretation of pre-OOK *o as [ɔ] is from Serafim and Shinzato's work (2021, 28).

4 Serafim and Shinzato (2021, 28) also write that "Thorpe (1983, 53-5) says that pRk [Proto-Ryukyuan] had phonetic aspiration", but Thorpe (1983, 54) posits a rule of aspiration for Proto-Amami-Okinawan, not PR.

2 *Haedong chegukki*

Haedong chegukki is a report in Classical Chinese on Japan and the Ryukyu Kingdom submitted to the Korean court in 1472 (Robinson 2012, 1). The original does not survive; the extant version from 1512 has appendices absent from the original.

One of those appendices dated 1501 is a bilingual phrasebook with 169 Ming Mandarin phrases and words followed by their equivalents in *Haedong chegukki* Okinawan (HCO). Ming Mandarin was the spoken language that educated Koreans and Ryukyuans had in common. HCO was only transcribed in the then-new hangul (H) alphabet. There was no attempt to write HCO in *kana* and *kanji*. The first entry in the phrasebook is (2).

(2) 你是那裏的人
lit. 'you be where ATTR person'
우라ᄌᆞ마피츄 (HC1)[5]
H *ura tsʌma pʰitsʰju*
HCO /ʔʊra Ntuma Fitʊ/
lit. 'you where person'
'Where are you from?'

I transcribe H in italicized IPA, whereas I write HCO (i.e. my phonemic interpretation of H) in roman in slashes. I have added word spacing for clarity.

(2) exemplifies some of the difficulties of interpreting H. Although H is an alphabet and is therefore far more versatile than Chinese characters or *kana* which represent syllables rather than segments, it is not IPA. H does not have letters for all of the segments in HCO. (2) contains several instances of H letters that only approximate HCO segments (3-7):

(3) H ㅜ *u*: HCO /ʊ/ (not /u/)

(4) H ᄌᆞ *tsʌ*: HCO /Ntu/, phonetically [dz̩] preceded by vowel nasalization (not /tsʌ/)

(5) H ㅍ *pʰ*: HCO /F/ (not /pʰ/)

(6) H ㅊ *tsʰ*: HCO /t/ [tɕʰ], palatalized after /i/ and aspirated before a nonhigh vowel (not /tsʰ/)

(7) H ㅠ *ju*: HCO /jʊ/ (not /ju/)

5 All HC appendix entries are numbered according to Tanaka (1991). Entry numbers are preceded by 'HC' to distinguish them from example numbers which also appear in parentheses: e.g. (HC1) refers to the first entry in HC, whereas (1) refers to example 1 in this paper.

I will not discuss the reasoning for all of the differences between HCO segments and their H transcriptions here, though I will explain my choice of the non-IPA symbol /F/ later.

The limitation of H that is most relevant for this paper is its paucity of letters for front vowels and for back labial vowels. H only has one front vowel letter (ㅣ *i*) and letters for two back labial vowel letters with or without a preceding palatal glide (ㅜ *u*, ㅠ *yu*, ㅗ *o*, ㅛ *yo*). It lacks letters for [e ɪ ʊ], the vowels I might expect to find in HCO if the TSS model is correct. I write 'might' because without looking further at the data, I cannot guess where HCO is on the continuum between PR and the OOK of the *Omoro sōshi*.

In theory, HCO could still retain PR *e and *o, and in fact Lin (2015, 128) reconstructs the mid vowel phonemes /e/ and /o/ in HCO.

On the other hand, HCO could be closer to the OOK of the *Omoro sōshi*. PR *e and *o could have raised to [ɪ] and [ʊ] or even [u] as in Serafim and Shinzato's (2021) reconstruction.

How can I determine the degree of raising in HCO or even if raising had occurred at all if H only had letters for *i*, (*j*)*u*, and (*j*)*o*? H *i*, the sole front vowel letter, could represent the HCO front vowels [ɪ] and [e], and both H *u* and H *o* could represent HCO [ʊ], a vowel between them in height. But was that actually the case?

Before tackling that question in the following sections, I wish to point out two further types of issues with HC.

The first involves whoever transcribed HCO into H. Was there just one transcriber? If there was more than one transcriber, different transcribers could have used more than one style of transcription. And even if there was only a single transcriber, he could have changed his mind about how to transcribe HCO during the compilation of the appendix.

The second involves whoever spoke HCO for the transcriber(s). Was there just one informant? If there was more than one informant, different informants could have had different dialects or idiolects. And even if there was only a single informant, he could have varied his style of speech during the compilation of the appendix.

In short, H is an imperfect medium for recording HCO, and we do not know enough about the recorders and informants to be certain that we are dealing with a consistent transcription of a single variety of HCO. The data in HC cannot be held to the same standard as modern linguistic fieldwork. Nonetheless, we can only work with what we have. And even though the transcriptions in HC are indeed inconsistent, I will demonstrate below how that inconsistency may in fact be their saving grace.

3 Reflexes of PR *e in *Haedong chegukki*

The TSS model of Okinawan language history predicts that the HCO reflexes of PR *e were front vowels somewhere along a spectrum from mid [e] to high [i].

Without looking at HC, I might expect those reflexes to be transcribed in H as ㅣ *i*, ㅑ *jə*, and/or ㅖ *jəj*. H ㅣ *i* might represent [i] or [ɪ]. Although the other two letters may seem to be odd choices for transcribing front vowels, H ㅖ *jəj* is the dominant transcription of Japanese /e/ in *Irop'a* 伊路波 (The *Iroha* Syllabary), a 1492 Korean textbook of Japanese that was roughly contemporaneous with HC. H ㅖ *jəj* was also the usual transcription of Japanese /e/ in *Ch'ŏphae sinŏ* 捷解新語 (Rapidly Understanding a New Language, 1618; first published 1676). The missionary romanization of Japanese from the same period renders Japanese /e/ as *e*, confirming that H ㅖ *jəj* in *Ch'ŏphae sinŏ* was a Korean approximation of a Japanese mid front vowel. Chinese character transcriptions indicate that the Old Japanese (OJ) phoneme /e/ was also roughly pronounced [e] in the eighth century CE (Miyake 2003, 227, 250). If the pronunciation of Japanese /e/ was [e] in both the eighth and seventeenth centuries, it would be simplest to also interpret H ㅖ *jəj* in *Irop'a* as a transcription of Japanese [e] in the fifteenth century. H *jəj* is not a palatal mid vowel, but its segments share features with Japanese [e]: *j* is palatal and *ə* is mid. H ㅑ *jə*, a less frequent transcription of Japanese /e/ in both *Irop'a* and *Ch'ŏphae sinŏ*, is also a combination of a palatal glide and a mid vowel. Fifteenth century Korean /ə/ may have had a fronted allophone like [ə̟] or even [e] in the vicinity of /j/, though such a phonetic detail is impossible to verify.

Although H has a letter ㅖ that is now pronounced [e] in modern Korean, in the fifteenth century, that letter was pronounced [əj] in Middle Korean, and that letter never appears in HC.

The letters that do appear in HC transcriptions of words with reflexes of PR *e are in Table 1.

Table 1 Transcriptions of reflexes of PR *e in *Haedong chegukki*

Hangul letter(s)	Transcription	Frequency
ㅡㅣ	ɯj	24
ㅣ	i	15
ㅖ	jəj	15
ㅑ	jə	5
ㅜㅣ	uj	1
·ㅣ	ʌj	1
ㅏㅇㅣ	ai	1

The most frequent letter in Table 1 is –ㅣ *ɯj*, which does not transcribe any Japanese vowels in *Irop'a* or *Chŏphae sinŏ*. Its absence from Korean transcriptions of Japanese indicates that it represents a vowel present in HCO that was not present in Japanese **[tab. 1]**.

That vowel could have been the [ɪ] reconstructed by Serafim and Shinzato (2021). [ɪ] is slightly backer than front [i]. To Korean ears, HCO [ɪ] could have sounded like a high vowel between Korean high back ㅡ /ɯ/ and high front ㅣ /i/, whose letters combine to form the composite letter –ㅣ *ɯj*.

The second most frequent letter is ㅣ *i*, which must represent [ɪ] rather than [i] as a reflex of PR *e. (H ㅣ *i* also represents [i] as a reflex of PR *i, but such cases are outside the scope of this paper.) I do not regard the H spelling ㅣ *i* as evidence for [i] as an HCO reflex of PR *e because the reflexes of the PR subordinative converb *-te were transcribed with ㅖ *jəj* (8) and ㅕ *jə* (9) as well as ㅣ *i* (10), indicating that its HCO reflex could not have simply ended in [i].

(8) H 타졔 *t^{h}atsjəj* 'leave-CON-SUB': PR *tat-i-te (HC8)

(9) H 왜쳐 *wajts^{h}jə* 'be-CON-SUB': PR *wor-i-te (HC24)[6]

(10) H 랃디 *ratti*[7] 'become-CON-SUB': PR *nar-i-te (HC23)

The merger of the Okinawan reflexes of PR *e and PR *i (i.e. [ɪ] and [i]) did not occur until "just before the arrival of Westerners in Okinawa around 1800" (Serafim 2008, 87).

The next two most frequent letters are H ㅖ *jəj* and ㅕ *jə*, which both represent Japanese /e/ in *Irop'a* and *Chŏphae sinŏ*. They probably similarly represent an [e]-like vowel in HCO. The fact that they can be used interchangeably with H ㅣ *i* to write the HCO reflex of the suffix *-te in (8) and (9) suggests that they stood for a raised [e̝] nearly as high as [ɪ] if not [ɪ] itself.

All of the evidence so far indicates that the HCO reflex of PR *e was a front vowel between [e] and [i] in height: i.e. [e̝] and/or [ɪ]. The exact height of the vowel may have varied by dialect and/or generation: e.g. older speakers could have pronounced it lower than younger speakers. For phonemic purposes, all that matters is that the vowel was lower than high /i/.

6 The identification of this item is from Lin (2015, 106), though the PR form is mine.

7 The HCO reflexes of PR *n and *r are frequently confused in the H transcription. That is not evidence for a merger of those phonemes in HCO because their Okinawan reflexes /n/ and /r/ are distinct to this day. The compiler(s) of the HC vocabulary must have had difficulty distinguishing between allophones of HCO /n/ and /r/.

I will symbolize that vowel phoneme as /ɪ/. Although HCO /ɪ/ may have been pronounced as low as [e̝], /e̝/ is difficult to type and easily confused with /e/ without a raising diacritic. Moreover, the symbol /ɪ/ reminds the reader that HCO /ɪ/ was higher than PR *e. Writing the HCO vowel as /e/ would give the false impression that PR *e had remained unchanged for centuries.

And I could end the story of PR *e in HCO there if not for several inconvenient facts.

First, there is a strong correlation between PJ *aj and H –ㅢ *ɯj*. Nearly all reflexes of PJ *aj in HC were transcribed with H –ㅢ *ɯj*.

(11) H 아긔 *akɯj* 'to raise': PJ *aNka-j- (HC13)

(12) H 사긔 *sakɯj* 'wine': PJ *sakaj[8] (HC17)

(13) H 아릐 *arɯj* 'that (distal)': PJ *araj (HC21)

(14) H 구릐 *kurɯj* 'this': PJ *kəraj (HC28)

(15) H 아믜 *amɯj* 'rain': PJ *amaj (HC34)

(16) H 칸즤 *k^{h}antsɯj* 'wind': PJ *kaNsaj (HC41)

(17) the second syllable of H 쏘믜디 *stomɯjti* 'morning': PJ *tuto-ma-j-taj (HC43)[9]

(18) H 고믜 *komɯj* 'rice': PJ *kəmaj (HC83)

(19) H 나븨 *napɯj* 'pot': PJ *naNpaj (HC130)

Citations are not exhaustive: e.g. (12) occurs ten times in HC. Each of the three exceptions (20-22) appears only once in HC.

(20) the third syllable of H 쏘믜디 *stomɯjti* 'morning': PJ *tuto-ma-j-taj (HC43)

(21) H 코몌 *k^{h}omjəj* 'rice': PJ *kəmaj (HC80)

(22) H 티 *t^{h}i* 'hand': PJ *taj (HC148)

8 PJ nouns ending in *-aj are often analyzed as *-a-final roots with *-i suffixes, but Frellesvig (2021) has called that assumption into question by proposing that some *-aj nouns in fact originally ended in *-a followed by a consonant that may not have been *-j. Although I believe Frellesvig is correct, I will continue to reconstruct these nouns with *-aj since their original consonantal codas, if any, are not relevant to the question of what H –ㅢ *ɯj* represents in HC. One could regard Frellesvig's *-aC (with *C representing a consonant) reconstructions as early PJ and the conventional *-aj reconstructions as late PJ.

9 The structure of this word is opaque. I follow Martin's (1987, 558) reconstruction but convert his *-i to *-j and delete his intervocalic PJ *-C- which is unattested in any later Japonic language.

(20) may not be an exception if *-ti* is not from PJ *taj. (21) is a variant spelling of (18) which is not an exception. (22) is definitely from PJ *taj, so I cannot potentially dismiss it in the same manner as (20). It may be significant that *ɯj* is not attested after *t* in HC, though I cannot formulate a general constraint against *ɯj* after coronals, since H 릐 *rɯj* (13, 14) and H 즤 *tsɯj* (16) are attested. The lack of H †띄*tɯj* in HC may be accidental.

I will examine two other isolated exceptions (26 and 28) later.

Second, H ㅢ *ɯj* almost never corresponds to PR *e from sources other than *aj. The only exceptions I have found so far are (23)-(25).

(23) H 아랴븨란 *arjapɯjran* 'exist-CONV-POL-NEG': < *ar-i + loan of Middle Japanese /ɸaNper/ + *-an (HC16)

The polite auxiliary in (23) is not a verb inherited from PR but rather a borrowing of a Japanese innovation, an irregular fusion of *pap-i-ni-ar- (Martin 1987, 682).

(24) H 픠루 *pʰɯjru* 'garlic': PR *peru (Thorpe 1983, 290) < PJ *peru (HC98)

(25) H 외븨 *ojpɯj* 'finger': PR *UjUbe (Thorpe 1983, 285) < PJ *əjoNpe (HC151)

I will return to the problems of (23)-(25) later in my discussion of (36)-(38) below.

Given that the compiler(s) of the HCO phrasebook did not know PJ, their ability to use H ㅢ *ɯj* almost exclusively for reflexes of PJ *aj long after it should have monophthongized to PR *e is difficult to explain within the TSS framework. At the end of this section, I will present an alternate framework that will account for the correlation between H ㅢ *ɯj* and PJ *aj.

Third, the HCO reflex of PR *me (Thorpe 1983, 284) is (26), the only instance of H ㅟ *uj* corresponding to PR *e.

(26) H 뮈 *muj* 'eye': PR *me < PJ *maj (HC143)

The TSS framework would lead me to expect H †며 *mjə*, †몌 *mjəj*, or †미 *mi* for HCO [mɐ̯] and/or [mɪ], not H 뮈 *muj*. I am hesitant to explain the H ㅜ *-u-* by proposing a subphonemic [w] between HCO labials and /e/ since this is the only instance of H ㅟ*uj* in a labial plus /e/ syllable. H 뮈 *muj* may simply originate from a scribal error for H †믜 *mɯj*. I will propose an alternate explanation for the anomalous H ㅜ *-u-* later in my discussion of (47).

Fourth, the HCO reflex of PR *ane is (27), the only instance of H ·ㅣ *ʌj* corresponding to PR *e.

(27) H 아ᄀᆞᆯ *arʌj* 'older sister': PR *ane < PJ *ani-me 'older brother-female' (Martin 1987, 382) (HC6)

H ㄹ *r* is a common misperception of HCO /n/. See fn. 6.

The TSS framework would lead me to expect H ᄐᆞ려 *rjə*, ᄐᆞ례 *rjəj*, or ᄐᆞ리 *ri* for HCO [nẹ] and/or [nɪ], not H ᄀᆞᆯ *rʌj*. I will return to this word in my discussion of (54) below.

Fifth, the HCO reflex of PR *ke is (28), the only instance of H ㅏ이 *aj* corresponding to PR *e.

(28) H 카이 *kʰai* 'box': PR *ke < PJ *kaj (Martin 1987, 448) (HC128)

The TSS framework would lead me to expect H ᄐᆞ켜 *kʰjə*, ᄐᆞ켸 *kʰjəj*, or ᄐᆞ키 *kʰi* for HCO [kʰẹ] and/or [kʰɪ], not H 카이 *kʰai*. See (48) below for a different expectation according to another framework.

H ㅏ이 *ai* is reminiscent of the ㅐ *aj* used to transcribe fifteenth century Japanese /fe/ and /re/ as 홰 *faj* and 래 *raj* in *Irop'a*. H ㅏ이 *ai* could have originated from a scribal error for H ᄐᆞ개 *aj*. However, H ㅐ *aj* is used elsewhere in HC for reflexes of earlier *apəj and *api and borrowings of Chinese *-a:j and *-əj, not reflexes of PR *e (29-32).

(29) H 패 *pʰaj* 'bow': Late Middle Chinese 拜 *pà:j[10] (HC63)

(30) H 소내 *sonaj* 'vinegared dish':[11] pre-HCO *so no apəj[12] (hc99)

(31) H 캐 *kʰaj* 'spoon': PJ *kapi (Martin 1987, 433) (HC122)

(32) H 대 *taj* 'table': Early Middle Chinese 臺 *dəj[13] (HC138)

10 Chinese reconstructions in this paper are from Pulleyblank (1991). These loans from Middle Chinese were probably borrowed through Japanese.

11 The HCO gloss is 菜蔬 'vegetables', which seems to be an error since the word is literally 'vinegar GEN mix' and corresponds to modern Okinawan /sune:/ or /su:ne:/ 'vinegared dish'.

12 I only reconstruct this collocation at the pre-HCO level because I do not know if it is attested outside Okinawan. The final element of the collocation would be *ape in the TSS model. However, I reconstruct *əj instead of *e because that element corresponds to OJ /apəj/ < PJ *apaj. I explain my decision to distinguish between *əj from PJ *aj and *e from other PJ sources below.

13 The corresponding Sino-Japanese reading is *dai* < *ndai /Ntai/. The HCO word was probably [ndaj] /Ntaj/. There are two ways to explain the unexpected correspondence between /a/ in the Japonic forms and Early Middle Chinese *ə. The latter may have been pronounced [ʌ] as reconstructed by Starostin (1989). The lower mid vowel [ʌ] may have sounded more like low /a/ than mid /ə/ to Japonic ears. Alternately, the Japonic forms could reflect a Chinese pronunciation like *daj that was transitional between Early Middle Chinese *dəj and Late Middle Chinese *tɦaj.

I interpret H ㅐ *aj* as a transcription of HCO /aj/ except in (30) where it may have represented HCO /aɪ/ with an /ɪ/ from PR *e. It is unclear whether /aj/ and /aɪ/ were distinct in HCO. A difference between HCO /aj/ and /aɪ/ may have been too subtle for the compiler(s) of the HC phrasebook to detect. In any case, H ㅏ이 *ai* transcribes HCO segmental sequences of relatively recent origin postdating the monophthongization of PJ *ai to PR *e, whereas H ㅏ이 *ai* in (27) corresponds to a PJ *aj that should have monophthongized to PR *e long ago.

What if that monophthongization had not occurred in PR? What if it had yet to occur in HCO, long after the breakup of PR?

Suppose that PJ *aj had raised to *əj in PR. Such a raising had also occurred in Japanese.

It is possible that *aj-to-*əj raising had occurred in late PJ, just prior to the breakup of Ryukyuan and Japanese. In that scenario, PJ *e and *o then similarly raised to /i/ and /u/ in OJ,[14] whereas PR *e and *o remained mid.

It is also possible that the raising may have independently occurred in Ryukyuan and Japanese. In that scenario, the raising of PJ *aj to PR *əj was initially an isolated event in early Ryukyuan, whereas the raising of PJ *aj to OJ /əj/ may have been part of a general trend in Japanese that raised PJ *e and *o to OJ /i/ and /u/.

I use the word 'initially' because there is no doubt that there were later waves of raising in Ryukyuan. PR *e and *o eventually did become modern Okinawan /i/ and /u/, but that shift occurred long after the raising of PJ *e and *o in Western Old Japanese.

I propose that PR *əj then remained unchanged for centuries until mid vowels began to rise in the years prior to HCO circa 1500. *e became a front vowel between [e] and [i] in height: i.e. [e̝] and/or [ɪ]. I will demonstrate in the next section that *o became a back vowel between [o] and [u] in height: i.e. [o̝] and/or [ʊ]. The schwa component of *əj became a vowel between [ə] and [ɨ] in height: i.e. [ə] and/or [ɨ̞]. Ideally I would like to write that vowel with a letter for a central vowel that is near-high [ɪ] and [ʊ], but the IPA has no such letter. I will hereafter refer to HCO [ə] as a near-high vowel even though its IPA symbol is for an upper mid vowel.

PR *əj cannot have remained unchanged (i.e. retaining a mid vowel) in HCO because if it were still [əj] at the end of the fifteenth century CE, the compiler(s) of HC could have written it with H ㅖ *əj*, a letter absent from the text. In reality, they generally wrote the HCO reflex

14 The similarity did not extend to final position. PJ final *aj rose to OJ /əj/, whereas PJ final *e and *o did not raise and remained mid /e/ and /o/ in OJ (Frellesvig, Whitman 2008, 22). Unless specified otherwise, the term 'OJ' in this paper refers only to Western Old Japanese, the most documented variety of OJ. See Kupchik (2011) for the details of vowel raising in Eastern Old Japanese dialects.

of PR *əj as H –ᅴ *ɯj*. I will write that reflex as /əj/ with /ə/ to indicate that its vowel is higher than the schwa of PR *əj.

Spellings like (33)-(35) for HCO reflexes of PJ forms with *aj indicate that HCO /əj/ had begun to merge with /e/.

(33) the third syllable of H 쏘믜디 *stomɯjti* 'morning': HCO /tsɨtoməjtɪ/ < /tsɨtoməjtəj/ < PR *tutomətəj < PJ *tuto-ma-j-taj (HC43)

(34) H 코몌 *k^{h}omjəj* 'rice': HCO /komɪ/ < /komɘj/ < PR *komɘj < PJ *kəmaj (HC80)

(35) H 티 *t^{h}i* 'hand': HCO /tɪ/ < /təj/ < PR *təj < PJ *taj (HC148)

(33) may not be evidence for the merger of HCO /əj/ and /e/ if its final syllable was from PJ *te rather than PJ *taj.

The PR *əj > HCO /əj/ scenario above accounts for all instances of H –ᅴ *ɯj* but (36)-(38).

(36) H 아랴븨란 *arjapɯjran* 'exist-CONV-POL-NEG': HCO /ʔar-j-aNpəjr-an/ < *ar-i + loan of Middle Japanese /ɸaNper/ + *-an (HC16)

(37) H 픠루 *p^{h}ɯjru* 'garlic': PR *peru (Thorpe 1983, 290) < PJ *peru (HC98)

(38) (3H 외븨 *ojpɯj* 'finger': PR *UjUbe (Thorpe 1983, 285) < PJ *əjoNpe (HC151)

As I noted in my discussion of (23), the auxiliary in (36) is probably a borrowing from Japanese. I would expect it to have been transcribed as H †-ᅡ비ᄅ- *-apir-*, †-ᅡ벼ᄅ- *-apjər-*, or †-ᅡ볘ᄅ- *-apjəjr-* reflecting HCO †/-aNpɪr-/.[15] The H –ᅴ *ɯj* of the actual transcription is either an error of the Korean transcriber or evidence for the confusion of /əj/ and /ɪ/ in HCO. If HCO /əj/ was beginning to shift to /ɪ/, /-aNpəjr-/ with /əj/ might be a hypercorrect pronunciation of †/-aNpɪr-/ with †/ɪ/. See (54) below for another potential example of hypercorrection.

(37), (38) have H –ᅴ *ɯj* instead of ㅣ *i*, ㅕ *jə*, and/or ㅖ *jəj* for HCO /ɪ/ from PR and PJ *e. If H –ᅴ *ɯj* in (37), (38) does not reflect a hypercorrect HCO /əj/ for HCO /ɪ/, it could represent an etymological HCO /əj/. But

15 H has a single letter ᅣ for *ja* that transcribes the syllable containing HCO /-j-/ 'CONV' and the initial vowel of /-aNpəjr-/ 'POL'. I use H ᅡ *a* as the transcription of the initial vowel of the auxiliary in isolation: i.e. without a preceding /-j-/. I supply prenasalization /N/ in HCO /-aNpɪr-/ on the basis of prenasalization in Middle Japanese /-ɸaNper-/. Prenasalization was inconsistently indicated in HC: e.g. /ʔaNkəj-/ 'to raise, give' was transcribed as both H 아긔- *akɯj-* without a nasal and H 앙긔- *aŋkɯj-* with a nasal. Transcriptions without nasal letters may indicate nasalized vowels: e.g. H 아긔- *akɯj-* may represent a pronunciation like [ʔãgəj-] with voicing of /k/ after a nasalized vowel. The core subset of H letters used in HC lacked the special letters for voiced obstruents used for the prescriptive Sino-Korean readings in *Tongguk chŏngun* 東國正韻 (Correct Rhymes of the Eastern Country; 1448).

how could that be possible if the Japanese cognates (39), (40) have an /i/ which cannot be from PJ *aj, the source of HCO /əj/?

If I were unaware of the HCO forms in (37), (38), I would derive the /i/ of the Japanese forms in (39), (40) from a PJ *e preserved in Thorpe's (1983, 285, 290) PR forms.

(39) OJ /piru/ < PJ *peru (not OJ */pəjru/[16] < PJ *pajru) 'garlic'

(40) Middle Japanese /ojoNpi/ < OJ *əjoNpi[17] < PJ *əjoNpe[18] (not Middle Japanese */ojoNpe/ < OJ *əjoNpəj < PJ *əjoNpaj) 'finger'

However, I am aware of the HCO forms (37), (38), so I have to somehow reconcile them with the Japanese forms (39), (40). How can I explain the anomalous correspondence of HCO /əj/ to Japanese /i/ in 'garlic' and 'finger'? Perhaps HCO /əj/ and OJ /i/ are reflexes of a PJ *ej. In the Ryukyuan branch of Japonic, PJ *ej merged with *əj which then became HCO /əj/. On the other hand, in the Japanese branch of Japonic, PJ *ej became OJ /i/, possibly merging with pre-OJ *e along the way. I contrast these two paths of development in Figure 1:

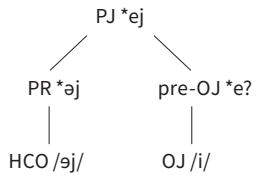

Figure 1 The development of Proto-Japonic *ej in Haedong chegukki Okinawan and Old Japanese

Normally pre-OJ *e should remain /e/ in OJ in word-final position (Frellesvig, Whitman 2008, 22). If PJ *ej became pre-OJ *e, I cannot explain why that *e raised to */i/ at the end of 'finger'. Perhaps PJ *ej became OJ /i/ in all positions without a transitional pre-OJ *e phase in final position. Whether PJ *ej also became pre-OJ *e in medial position is unclear.

I rewrite (37), (38) as (41), (42), incorporating the sound changes in Figure 1.

16 I follow Miyake's (2003) interpretation of OJ B-type *e* but rewrite his non-IPA 'əy' as /əj/. Other interpretations of OJ B-type *e* are possible. What matters is not the precise phonemic or phonetic value of OJ B-type *e* but the fact that OJ *piru* 'garlic' has /i/ rather than B-type *e*.

17 This word is not attested phonographically in Old Japanese.

18 Middle Japanese /i/ can also be derived from OJ /ɨ/, a merger of PJ *i, *uj, and *oj, but I rule out those sources since none of them correspond to Thorpe's PR *e in 'finger' **[tab. 2]**.

(41) H 피루 *pʰɯjru* 'garlic': HCO /Fəjru/ < PR *pəjru < PJ *pejru (HC98)

(42) H 외븨 *ojpɯj* 'finger': HCO /ʔʊjʊNpəj/ < PR *ojoNpəj < PJ *əjoNpej (HC151)

Although the H transcription of (42) could be taken at face value to represent an HCO †/ʔʊjpəj/, I interpret it as HCO /ʊjʊNpəj/ for two reasons.

First, (42) is the only instance of H ㅚ *oj* in HC. There is no other evidence to suggest that HCO had syllables ending in /ʊj/. It is more likely that HCO /ʊjʊ/ was misheard as *ˣ*/ʊj/. The long vowel /iː/ in the modern Okinawan form /ʔiːbi/ may be a compression of H /ʊjʊ/ that combines the palatality of /j/ with the length of two /ʊ/.

Second, the modern Okinawan form /ʔiːbi/ has a /b/ from PR *Np. Ideally the intermediate HCO stage between /b/ and *Np should have been transcribed in H as †-ㅁㅂ- *-mp-*. However, the H transcription of HCO does not consistently indicate prenasalization, so (42) is probably an instance of unwritten prenasalization.

I am hesitant to reconstruct PJ *ej to account for only two items, (41) and (42). Unfortunately, it is unlikely that I will ever find other examples of pre-*Omoro sōshi* OOK /əj/ corresponding to OJ /i/ because HC is the sole known source of potential evidence for /əj/, and it only contains transcriptions of a small fraction of the OOK lexicon.

I could be adventurous and reconstruct OJ *ej in (43) in which OJ has a word-final /i/ rather than /e/ corresponding to Thorpe's PR *e.

(43) OJ /kɨri/ (not *ˣ*/kɨre/) 'fog': PR *kire (Thorpe 1983, 288) < PJ *kujre[19]

I would rewrite (43) as (44), reconstructing the PR form with an *əj from PJ *ej.

(44) OJ /kɨri/ 'fog': PR *kirəj < PJ *kujrej

I predict that the unattested HCO cognate of OJ /kɨri/ would be (45).

(45) H †기릐 *kirɯj* 'fog': HCO †/kirəj/ < PR *kirəj < PJ *kujrej

The scenario above cannot account for Thorpe's (1983, 288) two other reconstructions of PR words for 'fog': PR *ki and *kiro 'fog'.

If PR *ki is from a PJ root *kuj, what is the function of the suffix *-rej?

I was tempted to reconstruct PJ *kujro-j which would regularly become OJ /kɨri/. PR *kiro would be from *kujro sans *-j. The trouble is that PJ *kujro-j with *-j should become PR *ˣkiri which would then become modern Okinawan *ˣ*/tɕiː/ since *r was lost before *i (Thorpe

19 I reconstruct PJ *u following Martin (1987, 451): "the /u/ is confirmed by Hachijō *kuri*".

1983, 98). The actual Okinawan form is /tɕiri/ with a /ri/ from a PR *re or *rəj, not *ri.

Going beyond native Japonic words, I am aware of only a single Sino-Japanese form which might reflect a direct shift from pre-OJ *ej to /i/ in final position (46).

(46) 高句麗 *Kōkuri* < OJ */kaukuri/ < pre-OJ *kaukorej? < Early Middle Chinese *kaw kəw lɛj^{h} 'Koguryo'

The absence of other Sino-Japanese forms with /i/ corresponding to Middle Chinese *ɐj-like rhymes may imply that all Sino-Japanese forms other than (46) were borrowed after the raising of PJ *ej in pre-OJ.

I will conclude this section by returning to the three anomalous transcriptions with unique H vowel letters (26-28) and reexamining them through the lens of my new framework.

I would expect the HCO word for 'eye' to be †/məj/, transcribed as H †믜 *mɯj*. But the actual transcription of 'eye' is (47).

(47) H 뮈 *muj* 'eye': PR *məj < PJ *maj (HC143)

If H ᅱ*uj* is not an error for H ᅳᅵ *ɯj*, it may represent [ɵj] or [ʉj] with a rounded allophone of the vowel of /əj/ after a labial onset.

I would expect the HCO word for 'box' to be †/kəj/, transcribed as H †긔 *kɯj*. But the actual transcription of 'box' is (48).

(48) H 카이 *k^{h}ai* 'box': PR *kəj < PJ *kaj (Martin 1987, 448) (HC128)

The aspiration of H 카이 *k^{h}ai* is not surprising since the TSS model predicts aspiration before HCO nonhigh vowels. What is surprising is H ᅡᅵ *ai* instead of H ᅳᅵ *ɯj* H ᅡᅵ *ai* cannot be explained away as a slip of the brush because it bears no graphic resemblance to H ᅳᅵ *ɯj* apart from the shared grapheme H ᅵ *i*. H 카이 *k^{h}ai* appears to represent HCO /kaj/ [k^{h}aj] which on a phonemic level looks exactly like PJ *kaj, even though PJ *aj should not have been preserved in HCO or even in PR.

The modern Okinawan descendant of (48) is /ke:/ 'chest for clothes' with /e:/ which is usually from secondary *aj or borrowed *aj that postdate the raising of PJ *aj to PR *əj, not primary *aj.

If 'box' had primary *aj, it would have been †/kəj/ in HCO which would then have become †/ki:/ in modern Okinawan like (49), a word that is unfortunately not attested in HC.

(49) Okinawan /ki:/ 'hair' < HCO †/kəj/ < PR *kəj < PJ *kaj (Martin 1987, 447)

Is 'box' a *sui generis* instance of the retention of PJ *aj as late as HCO? Or could it be a borrowing from an otherwise unknown Japonic

Table 2 Some sources of Okinawan front vowels with Japanese correspondences

PJ	PR	HCO	Okinawan	Pre-OJ	OJ	Japanese
*i	*i	/i/	/i(:)/[1]	*i	/i/	i
*uj				*uj	/ɨ/, /i/[2]	
*oj[3]				*oj?[4]		
*əj	*e	/ɪ/ [ę] ~ [ɪ]		*əj	/ɨ/, /i/, /əj/, /e/[5]	i, e
*e				*e	/i/, /e/[6]	
*aj	*əj	/əj/		*aj	/əj/, /e/[7]	e
*ej			/i/[8]	*i, *e?[9]	/i/	i
(none)[10]	(none)	/aj/	/e:/	(none)	/ai/	ai
*api[11]	*api			*api	/api/	
*a:j	*a(:)j	/a(:)j/		*aj	/əj/, /e/[12]	e

1 /i:/ at the end of monosyllabic words. One could regard that length as the result of a phonological rule and phonologize monosyllables with short /i/: e.g. Okinawan /ki/ [ki:] < PR *ke < PJ *kəj 'tree'.

2 Late pre-OJ *ɨ fronted to OJ /i/ except after /k Nk p Np m/.

3 A clear-cut example of PJ *oj is Serafim's (2008, 92) PJ *sungoi- or *songoi- 'to exceed', which I reconstruct in my notation as *suNko-j-. The second vowel of the root is preserved in Middle Japanese /suNko-s-/ 'to spend (time)' which appears to be a borrowed eastern archaism without the regular *o-raising in Western Old Japanese /suNku-r-/ 'to exceed'. The Western OJ cognate of modern *sugo-s-* may be attested in *Man'yōshū* 4318 as 須…之弖 /su … site/. The missing second man'yōgana has been assumed to be 具 /Nku/, presumably on the basis of Western OJ /suNku-r-/ and Middle Japanese /suNku-s-/, but in theory the lost character could have also represented /Nko/. Igarashi (1969) excludes /su … site/ from his dictionary of OJ phonographic spellings.

4 It is unclear whether */oj/ was still distinct from */uj/ in pre-OJ.

5 Late pre-OJ *ɨ and *əj respectively fronted to OJ /i/ and /e/ except after /k Nk p Np m/. Frellesvig and Whitman (2008) reconstruct PJ *ɨi (= *ɨj in my notation) as a source of OJ /ɨ i/ and PJ *əi (= *əj in my notation) as a source of OJ /əj e/. The Frellesvig-Whitman hypothesis neatly accounts for a split in OJ reflexes but at the price of other complications that are beyond the scope of this paper which is concerned primarily with HCO and secondarily with PR, not PJ. A non-Frellesvig-Whitman account of the different reflexes could involve dialects: e.g. PJ *kəj 'tree' became /kɨ/ in Western OJ but /kəj/ in Eastern OJ with a mid vowel like PR *ke in the southwest. I suspect /ɨ i/ are the regular Western OJ reflexes of PJ *əj, whereas Western OJ forms with /əj e/ from PJ *əj may be borrowings.

6 PJ *e remained mid in word-final position in OJ (Frellesvig, Whitman 2008, 22).

7 Late pre-OJ *əj fronted to OJ /e/ except after /k Nk p Np m/.

8 In theory, a monosyllabic word ending in PJ *ej might end in long [i:] in Okinawan, but no such word is known.

9 It is unclear whether PJ *ej became pre-OJ *e in medial position.

10 This row is for Sino-Japanese loans in HCO and OJ. There were no such loans in PJ and PR. See other rows for Sino-Japanese loans in pre-OJ: e.g. if one wants to see the development of *aj in sixth century Sino-Japanese loans, find the row with native pre-OJ *aj which became /əj/ or /e/ in eighth century OJ.

11 I provide only a single example of the PJ and PR *VCV sequences that merged as /e:/ in Okinawan. Some of those sequences merged in HCO as a secondary /aj/ (51, 52) not to be confused with the primary *aj of PJ that became HCO /əj/.

12 Late pre-OJ *əj fronted to OJ /e/ except after /k Nk p Np m/.

language that retained PJ *aj? I would rather not posit an isolated irregularity or an entire lost variety of Japonic if I can come up with a solution that involves a regular sound change within a known Japonic lineage. The final row of Table 2 contains such a solution **[tab. 2]**.

Given that vowel length has been reconstructed in PJ (e.g. by Serafim 2008), it would not be a stretch to speculate that PJ could also have long vowels before *-j. If PJ *aj had a long vowel counterpart *a:j, an earlier length distinction between *aj and *a:j may have become a quality distinction between /əj/ and /aj/ in HCO, whereas *aj and *a:j merged in Japanese **[tab. 3]**.

Table 3 A Proto-Japonic *aj: *a:j minimal pair

Gloss	PJ	PR	HCO	Okinawan	Pre-OJ	OJ	Japanese
'hair'	*kaj	*kəj	†/kəj/	/ki:/	*kaj	/kəj/	ke
'box'	*ka:j	*ka(:)j	/ka(:)j/	/ke:/	*kaj	/kəj/	(*ke*)*

* This word is now obsolete in Japanese, but the OJ word can be pronounced in modern Japanese as *ke*.

I write vowel length in parentheses in PR and HCO /ka(:)j/ in Tables 3 and 4 since I do not know if vowel length was phonemic before the coda /j/ in PR or HCO.[20] The fact that 'box' was transcribed with two syllabic letter blocks as H 카이 *kʰai* may suggest that to Korean ears, the word sounded as long as two Korean syllables, unlike HCO /aj/-syllables transcribed with single syllable letter blocks (50-53).

(50) H 패 *pʰaj* 'bow': HCO /Faj/ < Late Middle Chinese 拜 *pà:j (HC63)

(51) H 소내 *sonaj* 'vinegared dish': HCO /sʊnaj/ < pre-HCO *so-no-apəj (HC99)

(52) H 캐 *kʰaj* 'spoon': HCO /kaj/ < PJ *kapi (Martin 1987, 433) (HC122)

(53) H 대 *taj* 'table': HCO /Ntaj/ < Early Middle Chinese 臺 *dəj (HC138)

The last remaining anomaly is (54).

(54) H 아ᄅᆡ *arʌj* 'older sister': PR *ane < PJ *ani-me 'older brother-female' (Martin 1987, 382) (HC6)

20 Vowel length was certainly phonemic in open syllables: e.g. Okinawan /mu:ku/ 'bridegroom' retains the long vowel of PJ *mo:ko (Serafim 2008, 86).

I hypothesize that (54) may be an error for ʜ †아릐 *arɯj* representing a hypercorrect pronunciation †[ʔanəj] of ʜᴄᴏ /ʔane/.[21] The short horizontal stroke of ʜ ㅢ *ɯj* could have been miscopied as the dot of ʜ ㆍㅣ *ʌj*.

It is also possible that Martin's etymology requiring an irregular reduction of ᴘᴊ *-ime- to ᴘʀ *-e is incorrect. The ʜᴄᴏ form could have been /ʔanəj/ from ᴘʀ *ʔanəj going back to a ᴘᴊ *anaj. The resemblance between ᴘᴊ *anaj 'older sister' and ᴘᴊ *ani 'older brother' would then be coincidental or the result of analogical alteration of one form to resemble the other.

4 Reflexes of ᴘʀ *o in *Haedong chegukki*

A three-way contrast between high labial *u and mid labial *o and non-labial *ə in ᴘᴊ was reduced to a two-way contrast between high labial *u and mid labial *o in ᴘʀ. ᴘʀ *o then raised to ʜᴄᴏ /ʊ/, eventually merging with ʜᴄᴏ /u/ **[tab. 4]**. I will explain my choice of the symbol /ʊ/ later.

Table 4 Major native sources of Okinawan /u/

ᴘᴊ	ᴘʀ	ʜᴄᴏ	Okinawan	Pre-ᴏᴊ	ᴏᴊ	Japanese
*u	*u	/u/ [u] ~ [ʐ̩][1]	/u(:)/[2], /i/	*u	/u/	u
*o	*o	/ʊ/ [o̞] ~ [ʊ]	/u(:)/	*o	/u/, /o/[3]	*u*, *o*
*ə				*ə	/ə/, /o/[4]	o

1 ʜᴄᴏ /tu Ntu su Nsu/ were pronounced [tsʐ̩ n(d)ʐ̩ sʐ̩ n(d)ʐ̩]. /N/ may also have surfaced as nasalization of a preceding vowel. Pulleyblank's (1991) Late Middle Chinese *ʐ̩ was borrowed into Korean as /ʌ/, so it is likely that ʜ ㆍ ㅅ after sibilants transcribes ʜᴄᴏ [ʐ̩]. ʜᴄᴏ [ʐ̩] later merged with [i]. A similar shift of *ʐ̩ to [i] occurred after Cantonese sibilants. I could reconstruct ʜᴄᴏ [ɨ] or [ɯ] instead of [ʐ̩], but I would expect an ʜᴄᴏ [ɨ] or [ɯ] to be transcribed with the ʜ high vowel letter ㅡ ɯ, not the ʜ low vowel letter ㆍ ʌ.

2 /u:/ at the end of monosyllabic words. One could regard that length as the result of a phonological rule and phonologize monosyllables with /u/: e.g. Okinawan /mu/ [mu:] < ᴘʀ *mo < ᴘᴊ *mə 'seaweed'.

3 ᴘᴊ *o remained mid in word-final position in ᴏᴊ (Frellesvig, Whitman 2008, 22).

4 Pre-ᴏᴊ *ə rounded to /o/ after labials in most ᴏᴊ texts. *Kojiki* 古事記 (Records of Ancient Matters) (710 CE) is the only ᴏᴊ text retaining a distinction between /ə/ and /o/ after /m/.

21 I tentatively project the initial glottal stop of modern Okinawan back into ʜᴄᴏ. Although the ʜ transcription lacks the special glottal stop letter ㆆ *ʔ*, that does not necessarily mean ʜᴄᴏ lacked a glottal stop. That ʜ letter was only rarely used in early ʜ texts in Korean and became obsolete by the late fifteenth century, so the transcriber may not been aware of its existence. However, the fact that *k- in Chinese transcriptions of ᴏᴏᴋ occasionally corresponds to a zero initial in ʜᴄ transcriptions led Lin (2015, 147) to reconstruct a glottal stop /ʔ/ in ᴏᴏᴋ. Could the Chinese *k-transcriptions represent a uvular allophone [q] of /ʔ/? Could [q] be an archaic pronunciation of /ʔ/? Could the Japanese doublet *are* ~ *kare* 'that' contain two reflexes of an earlier *q in different dialects: zero and /k/?

Lin (2015, 190) reconstructs /o/ for OOK, but I prefer to write that vowel as /ʊ/ to remind the reader that it was higher than PR *o.

I do not reconstruct /ʊ/ as a mid /o/ because it was transcribed with H ㅜ *u* as well as H ㅗ *o* **[tab. 5]**.

Table 5 Frequencies of transcriptions of Old Okinawan reflexes of PR *o

PR	HCO	H	H transcription	Tokens	Percentage
*o	/ʊ/	ㅜ	u	33	33.7%
		ㅠ	*ju*[1]	5	5.1%
		ㅗ	o	50	51.0%
		ㅛ	*jo*[2]	9	9.2%
		ㅘ	waj	1	1.0%

1 This letter transcribes the HCO glide-vowel sequence /jʊ/. There is no H letter for /j/.

2 This letter transcribes the HCO glide-vowel sequence /jʊ/. There is no H letter for /j/.

The sole example of H ㅘ *waj* is (55). H ㅘ *waj* may either be an error for H ㅚ *oj* or a transcription of HCO [wɔ] with an unusually low allophone [ɔ] of /ʊ/, perhaps due to dissimilation before /j/, a glide resembling the high vowel /i/.

(55) H 왜쳐*wajtsʰjə* 'be-CON-SUB': HCO /wʊice/ < PR *wor-i-te (HC24)

The same morpheme can appear in HC with both H ㅜ *u* and H ㅗ *o* (56-7):

(56) H 피츄 *pʰitsʰju* (HC1) ~ 피죠 *pʰitsjo* (HC29) 'person': HCO /Fitʊ/ < PR *pito < PJ *pitə

(57) H 누 *nu* (HC112) ~ 루 *ru* (HC86) ~ 노 *no* (HC63) ~ 로 *ro* (HC80) 'genitive marker': HCO /nʊ/ < PR *no < PJ *nə

The inability to settle upon a single vowel symbol indicates that neither symbol was a perfect match for the HCO vowel which may have been between Korean /u/ and /o/. I symbolize that vowel phoneme as /ʊ/. Although HCO /ʊ/ may have been pronounced as low as [o̞], /o̞/ is difficult to type and easily confused with /o/ without a raising diacritic. Moreover, the symbol /ʊ/ reminds the reader that HCO /ʊ/ was higher than PR *o. Writing the HCO vowel as /o/ would give the false impression that PR *o had remained unchanged for centuries. An unchanged mid vowel would not have been transcribed with H ㅜ *u*, which is also the default transcription of the HCO reflex of PR *u: e.g. (58).

(58) H 파무 *pʰamu* 'snake': HCO /FaNpu/ < PR *paNpu (cf. Thorpe 1983, 332) (HC165)

PR *u has a special transcription as H · ㅅ after the coronal obstruents /s Ns t Nt/: e.g. (59).

(59) H 아숨비 *asʌmpi* 'play-IMP': HCO /ʔasuNpɪ/ [ʔasn̩mbɪ] < PR *asuNpe (cf. Thorpe 1983, 317)[22] < PJ *asoNpe (HC24)

(60) is the only form in HC containing both H · ㅅ and H ㅜ *u*.

(60) H ᄉᆞ우 *sʌu* 'vinegar': Okinawan /si:/ ~ /su:/, Japanese *su*, OJ */su/ (HC92)[23]

I could mechanically interpret (60) as HCO /suu/ or /suʊ/ with two vowels in different syllables following regular transcription conventions, but I suspect it is either a composite of (61) and (62) or an error for (62).

(61) H †ᄉᆞ *sʌ* 'vinegar': HCO /su/ [sz̩] < Japanese /su/ < PJ *so

(62) H †수우 *suu* 'vinegar': HCO /sʊ/ < PR *so < PJ *so

(61), the ancestor of modern Okinawan /si:/, is a borrowing from Japanese that underwent both the *o > /u/ shift in Japanese and the *u > [z̩] shift in Okinawan.

(62), the ancestor of modern Okinawan /su:/, is the native word for 'vinegar'. It could not have had *u in pre-HCO because it did not undergo the *u > [z̩] shift.

The use of two syllabic blocks to transcribe (60) may have indicated a phonetic long vowel [ʊ:] in HCO. Nonetheless, I do not reconstruct a phonemic long vowel in HCO or its ancestors since I am uncertain whether vowel length was distinctive in HCO monosyllabic words.

One final anomalous form is (63) which contains the only instance of H — *ɯ* in HC.

(63) H 크지 *kʰɯtsi* 'mouth': HCO /kuti/ < PR *kuti (Thorpe 1983, 308) < PJ *kutuj (HC145)

H — *ɯ* happens to match the phonetic value [ɯ] of modern Japanese /u/, but that fact is of no relevance for the reconstruction of OOK circa 1500.

Initially I regarded H — *ɯ* in (63) as an error for H †ㅜ *u* transcribing HCO †/u/ from PR and PJ *u. However, if the word had /u/ in HCO,

22 Thorpe's PR *u for PJ *o is irregular. Serafim (2008, 84) reconstructs PJ 'to play' with an *o that raised to *u prior to what I reconstruct as a shift of *u to HCO [z̩] after coronal obstruents.

23 Although the OJ word for 'vinegar' is not attested phonographically, the fact that the *kanji* 酢 for 'vinegar' is a phonogram for OJ /su/ implies that 'vinegar' was also /su/ in OJ.

I would not expect the preceding consonant to be aspirated. The TSS model predicts aspiration only before nonhigh[24] vowels: i.e. /a ɪ ʊ/ which are lower than high /i u/. H ㅋ k^h may then represent an HCO aspirated [k^h] before a nonhigh vowel resembling H †ㅜ *u*: i.e. HCO /ʊ/ from PR and PJ mid *o. Serafim (2008, 84) reconstructs PJ ?*kotoj 'mouth'[25] which would regularly become HCO /kʊti/ according to the correspondences in Table 6. H †쿠지 k^h*ɯtsi* 'mouth' would be a phonetic transcription of HCO /kʊti/ [k^hʊtɕi]. Accidentally omitting the vertical stroke of H †ㅜ *u* from the first letter block would result in H 크지 k^h*ɯtsi*, the transcription attested in HC. If Serafim and I are correct,[26] I can rewrite (63) as (64) with a revised PR form that is transitional between my HCO phonemicization and Serafim's PJ reconstruction.

(64) H 크지 k^h*ɯtsi* 'mouth': HCO /kʊti/ < PR *koti < PJ *kotoj (Serafim 2008, 84) (HC145)

The first *o of PJ *kotoj matches the mid vowel of 古次 Early Middle Chinese *kɔ' ts^hi^h,[27] a transcription of a toponym element widely regarded to be the word for 'mouth' in a Para-Japonic language of Koguryŏ.[28] The common ancestor of PJ and that Para-Japonic language (Proto-Macro-Japonic?) may have had a mid vowel like *o in its word for 'mouth'.

Unfortunately, my PR *koti does not predict the forms of 'mouth' in Table 7 which point to Thorpe's PR *kuti. Compare those forms with those for 'egg' from Thorpe's (1983, 62) PR *koga.

24 Although HCO /ɪ ʊ/ are near-high vowels, they could condition aspiration like the low vowel /a/. I place all three of those HCO vowels into a 'nonhigh' category. I reserve the term 'high' for the true high vowels /i u/.

25 I have rewritten Serafim's *y as IPA *j for consistency with other forms in this paper. I do not know why Serafim writes a question mark before his *kotoy.

26 Here I only refer to Serafim's (2008) PJ reconstruction. Serafim (2008, 84-5) also proposes a "First Vowel-Raising" of "many" PJ *o in words such as 'mouth' long before his "Second Vowel Raising" of "all remaining mid vowels". (It is unclear why he hyphenates "First Vowel-Raising" but not "Second Vowel Raising".) Serafim (2008, 84) reconstructs 'mouth' as *kutuy after First-Vowel Raising. That form would become HCO †/kuti/ with a high vowel /u/ that should have blocked aspiration of the preceding /k/. On the other hand, I propose a single gradual raising of mid vowels corresponding to Serafim's Second Vowel Raising.

27 I cite Pulleyblank's (1991) Early Middle Chinese readings in lieu of the unknown local (Sino-Koguryŏ?) readings which I presume were similar.

28 The Koguryŏ toponym 甲比古次 is a phonetic transcription of a non-Chinese name later replaced by 穴口郡 'cave mouth prefecture', a Chinese-style name that may be a translation of the earlier name. I use the term 'Para-Japonic' to refer to relative(s) of Japonic once spoken on the Korean peninsula. The choice of the phonogram 次 *ts^hi^h indicates that a syllable corresponding to Serafim's PJ *toj had shifted to something like *tsi in a Para-Japonic language long before pre-HCO */ti/ became HCO /ti/ [tɕi] and Middle Japanese /ti/ became modern Japanese [tɕi].

Table 6 Nonmatching onsets for 'mouth' and 'egg' in Ryukyuan varieties as recorded in Thorpe (1983)

Ryukyuan variety	'mouth'	'egg'
Aden	/kuci/	/huãã/
Yamatoma		/xoga/
Shodon	/ku·cj/	/kʻuga/
Taketomi	/huci/	/kuNga/

If 'mouth' were PR *koti, I would expect 'mouth' to have the same onsets as 'egg'.

On the other hand, the aspiration in HCO is not the only potential evidence for a mid vowel in PR *koti. Naze and Yuwan have /kʻ/ in 'mouth' as well as 'egg', implying that 'mouth' and 'egg' both once began with *ko **[tab. 7]**.

Table 7 Matching onsets for 'mouth' and 'egg' in Ryukyuan varieties as recorded in Thorpe (1983)

Ryukyuan variety	'mouth'	'egg'
Naze	/kʻuci/	/kʻuga/
Yuwan		

I could try to explain the different onsets for 'mouth' by reconstructing a PR *kowti that became *koti in some Ryukyuan varieties and *kuti in others. I would then have to project PR *ow back into PJ. This *ow would be the back counterpart of the *ej that I reconstructed for PJ. However, there may also be less drastic solutions: e.g. borrowing within Ryukyuan.

Aside from the troubling case of 'mouth', aspiration was clearly associated with the nonhigh vowels /a ɪ ʊ/. However, that association was not absolute.

Medial voiceless stops before HCO /a/ were almost always unaspirated in H transcription: e.g. medial /k/ in (65) and medial /t/ in (66).

(65) H 사가나 *sakana*: HCO /sakana/ 'fish' < loan of Middle Japanese /sakana/ (HC16)

(66) H 카다나 *kʰatana*: HCO /katana/ 'sword' < loan of Middle Japanese /katana/ (HC129)

(67) is an exceptional case of a medial aspirated stop before /a/.

(67) H 취타지 *tsʰujtʰatsi*: HCO /tuitati/ 'first day' < loan of Middle Japanese /tuitati/ (HC11)

HCO morphemes with voiceless stops before /ɪ ʊ/ were not consistently transcribed with H aspirated consonant letters (68-73).

(68-70) share the subordinative converb /-tɪ/.

(68) H 타졔 *tʰatsjəj* 'leave-CON-SUB': HCO /tat-i-tɪ/[29] < PR *tat-i-te < PJ *tat-i-te (HC8)

(69) H 왜쳐 *wajtsʰjə* 'be-CON-SUB': HCO /wʊ-i-tɪ/ < PR *wor-i-te < PJ *wər-i-te (HC24)

(70) H 랃디 *ratti* 'become-CON-SUB': HCO /nat-tɪ/ < PR *nar-i-te < PJ *nar-i-te (HC23)

(71) is a more conservative pronunciation of 'rice' than (72). The former lacks the aspiration and nonetymological /ɪ/ of the latter.

(71) H 고믜 *komɯj* 'rice': HCO /komɘj/ < PR *komɘj < PJ *kəmaj (HC83)

(72) H 코몌 *kʰomjəj* 'rice': HCO /komɪ/ < /komɘj/ < PR *komɘj < PJ *kəmaj (HC80)

The H transcription of (75) has aspiration even before a high vowel /i/.

(73) H 피츄 *pʰitsʰju* (HC1) ~ 피죠 *pʰitsjo* (HC29) 'person': HCO /Fitʊ/ < PR *pito < PJ *pitə

That aspiration is not evidence for reconstructing 'person' with a non-high /ɪ/ in HCO or *e in PR. The HCO reflex of PR *p was generally transcribed with aspiration as H ㅍ *pʰ* regardless of the following vowel, though exceptional cases of H ㅂ *p* also exist: e.g. (74-5).

(74) H 비ᄌᆞ쟈 *pitsʌtsja* 'sheep-TOP': HCO /FituNci-ja/ [pɸitsɻndʑija] < PJ *pituNsi-pa (HC163)

(75) H 오부시 *opusi* 'many-FIN': HCO /ʊpʊ-si/ < PJ *əpə-si (HC14)

Clearly the aspiration of H ㅍ *pʰ* reflects a phenomenon of a different nature than that of other H aspirated letters. What is not so clear is what H ㅍ *pʰ* represented. Ming Mandarin (MM) transcriptions of the OOK voiceless bilabial obstruent have *p-, *pʰ-, *f-, and even *xu[30] corresponding to H ㅍ *pʰ* (Lin 2015, 141) (76-7).

29 I interpret H 타졔 *tʰatsjəj* as a disyllabic transcription of a trisyllabic [tʰatɕitɕẹ]. Although the second vowel /i/ was omitted from the transcription, I assume it was present to condition the palatalization of the following /t/ indicated by H ㅈ *ts*. The Korean transcriber may have heard a sequence of two similar syllables [tɕitɕẹ] as a single syllable [tɕẹ]. Compare (68) with palatalized /t/ to (70) in which the second /t/ was not palatalized because /i/ was lost.

30 Lin (2015) writes only *x-, but I regard the following *-u- as part of the transcription of the OOK voiceless bilabial obstruent.

(76) MM 扒 *paʔ ~ 華 *xua 'tooth': H 파 *pʰa*: HCO /Fa/ (HC153) < PR *pa < PJ *pa (Martin 1987, 394)

(77) MM 必祿 *piluʔ ~ 皮祿 *pʰiluʔ ~ 非祿 *fuiluʔ 'day': H 피루*pʰiru*: HCO /Firu/ (HC47) < PR *piru < PJ *piru (Martin 1987, 409)

There is also a strange instance of a word-initial velar stop *k- in one MM transcription of (80), a word absent from HC.

(78) MM 谷古里 *kuʔkuli 'rejoice-INF:'[31] OOK /Fʊkʊr-i/ < PR *pokor-i < PJ *pokər-i 'take.pride.in-INF' (Martin 1987, 692)

I would have expected MM †xuʔkuli with *x-. The first syllable may have been confused with the similar-sounding second syllable.

Lin (2015, 142) reconstructs OOK /ɸ/, but I do not think a fricative would be transcribed with an unaspirated stop in both Korean and Chinese. I prefer to write the HCO phoneme as /F/ with a capitalized non-IPA symbol to indicate that I do not know its precise pronunciation. /F/ may have had an allophone like [pɸ] resembling MM *p, *pʰ-, and *f- and Korean /pʰ/ but lacking a precise match in either language. Another allophone may have resembled the 'obsolescent' Shuri Okinawan onset that Serafim (2008, 81) writes as [ɸ(w)]. MM 華 *xua (77) may have been the closest available match for an HCO mora like [ɸwa] since there was no MM syllable **fua*.

There is no need to reconstruct exotic consonant clusters like [pɸ] or [ɸw] in HCO at other points of articulation. Most of the allophones of OOK /k t/ had exact counterparts in MM and Korean with the exception of [tɕ] and [tɕʰ], the palatalized allophones of /t/, which could only be approximated with alveolar fricatives. I list the frequencies of H transcriptions of HCO /k t/ before /ɪ ɘj ʊ/ in **[tab. 9]**.

31 This verb appears in several Chinese glossaries (Lin 2015, 222) before MM 烏鴉沒 *uiamuʔ or 烏牙沒 *uiamuʔ, possibly OOK /ʔʊja-mʊ/ 'parent-FP'. The Chinese gloss for the phrase is 給賞 'to give an award', so the OOK phrase may have meant something like 'even parents rejoice (at their child receiving an award)'. The meaning 'rejoice' is literary and archaic in Shuri Okinawan (Martin 1987, 692). Here I have projected 'rejoice' back into OOK, but I am not sure if /Fʊkʊr-/ already had that meaning when the transcription first appeared in the 1561 edition *Shi Liuqiu lu* 使琉球錄 (Record of Ambassadors to the Ryukyus).

Table 9 Frequencies of hangul transcriptions of НСО /k t/ before /ɪ ə j ʊ/

н	**/ɪ/**	**/əj/**	**/ʊ/**
ㄱ- *k-*	0	0	9
-ㄱ- *-k-*	1	11	0
ㅋ- k^h-	0	0	1
-ㅈ- *-ts-*	3	0	2
-ㅊ- $-ts^h$-	1	0	2
ㄷ- *t-*	0	0	1
-ㄷ- *-t-*	11	0	3
ㅌ- t^h-	4	0	2

It is hazardous to draw hard conclusions from a small number of tokens.

The absence of н aspirated letters before /əj/ may simply be an artifact of a limited sample. A larger sample might have revealed that aspiration could also have occurred before /əj/. I would expect the near-high vowel /ə/ to condition aspiration like the other near-high vowels /ɪ ʊ/.

It is most likely that aspiration was not obligatory before near-high /ɪ ə ʊ/, contrary to the TSS model. However, I should also point out that aspiration is not binary. Perhaps НСО aspiration was not as strong as Korean aspiration and thus was not always salient to Korean ears. But that hypothesis would not explain the consistent use of н aspirated letters for the word-initial stops /k t/ before /a/.

I propose that such stops were the first to be aspirated and that aspiration gradually spread to voiceless obstruents before nonhigh vowels other than /a/. In НСО, medial /k t/ were unaspirated before /ɪ ə j ʊ/ with the exception of palatalized /t/ which was transcribed both with and without aspiration: e.g. (79).

(79) н 피츄 p^hits^hju (нс1) ~ 피죠 p^hitsjo (нс29) 'person': НСО /Fitʊ/ [pɸit$ɕ^h$ʊ] ~ [pɸit$ɕ_Q$]
< PR *pito < PJ *pitə

Aspiration may have spread not only within OOK but also from Shuri as a prestigious trait, taking root and persisting in the periphery but eventually perishing in Okinawan itself.

5 Conclusions

What began as a survey of H transcriptions of PR mid vowel reflexes in HCO led to my proposals of PR *əj and *a(:)j and a much later date for aspiration before nonlow vowels. Although those proposals may appear plausible within the limited context of HC, they remain to be tested against the evidence of Chinese transcriptions of OOK, the native OOK orthography of the *Omoro sōshi*, and modern Ryukyuan varieties. Until such testing occurs, my even more radical proposals of PJ *ej and * a:j can only be extremely tentative.

In the course of this study, I have also made more moderate discoveries of words that should be reconstructed with mid vowels in PR and PJ. However, limitations of space prevent me from discussing those words here. I hope to delve into them in a future publication.

Abbreviations and Symbols

ˣ	erroneous form
†	expected form
*	reconstruction
ATTR	attribute marker
CONV	converb
GEN	genitive
FIN	final verbal form
FP	focus particle
H	hangul
HC	*Haedong chegukki* 海東諸國紀
HCO	*Haedong chegukki* Okinawan
INF	infinitive
IPA	International Phonetic Alphabet
MM	Ming Mandarin
NEG	negative
OJ	Old Japanese (Western unless otherwise specified)
OOK	Old Okinawan of *Omoro sōshi*
PJ	Proto-Japonic
POL	polite
PR	Proto-Ryukyuan
SUB	subordinative converb
TOP	topic
TSS	Thorpe/Serafim/Shinzato

Bibliography

Frellesvig, B.; Whitman, J. (2008). "Evidence for Seven Vowels in Proto-Japanese". Frellesvig, B.; Whitman, J. (eds), *Proto-Japanese: Issues and Prospects*. Amsterdam: John Benjamins, 15-41.

Igarashi Jin'ichi 五十嵐仁一(1969). *Jōdai kanazukai jiten* 上代仮名遣字典 (A Dictionary of Ancient Kana Usage). Tōkyō: Shōgakukan.

Kupchik, J. (2011). *A Grammar of the Eastern Old Japanese Dialects* [PhD Dissertation]. Honolulu: University of Hawai'i at Mānoa.

Lin, C. (2015). *A Reconstruction of Old Okinawan: A Corpus-Based Approach* [PhD Dissertation]. Honolulu: University of Hawai'i at Mānoa.

Martin, S. (1987). *The Japanese Language Through Time*. New Haven: Yale University Press.

Miyake, M. (2003). *Old Japanese: A Phonetic Reconstruction*. London: Routledge Curzon.

Pulleyblank, E. (1984). *Middle Chinese: A Study in Historical Phonology*. Vancouver: University of British Columbia Press.

Pulleyblank, E. (1991). *Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: University of British Columbia Press.

Robinson, K. (2012). "Mapping Japan in Chosŏn Korea: Images in the Government Report *Haedong chegukki*". *Korean Studies*, 36, 1-30.

Serafim, L. (2008). "The Uses of Ryukyuan in Understanding Japanese Language History". Frellesvig, B.; Whitman, J. (eds), *Proto-Japanese: Issues and Prospects*. Amsterdam: John Benjamins, 79-99.

Serafim, L.; Shinzato, R. (2021). *The Language of the Old-Okinawan "Omoro Sōshi": Reference Grammar, with Textual Selections*. Leiden: Brill.

Starostin, S. (1989). *Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy* Реконструкция древнекитайской фонологической системы (Reconstruction of the Old Chinese Phonological System). Москва: Nauka.

Tanaka Takeo 田中健夫 (1991). *Kaitō shokokuki: Chōsenjin no mita chūsei no Nihon to Ryūkyū* 海東諸国紀 朝鮮人の見た中世の日本と琉球 (Record of Countries Across the Sea to the East: Medieval Japan and the Ryukyus as Seen by Koreans). Tōkyō: Iwanami bunko.

Thorpe, M. (1983). *Ryūkyūan Language History* [PhD Dissertation]. Los Angeles: University of Southern California.

Ca' Foscari Japanese Studies

1. Mariotti, Marcella; Miyake, Toshio; Revelant, Andrea (eds) (2014). *Rethinking Nature in Contemporary Japan: Science, Economics, Politics*. History and Society 1.

2. Miyake, Toshio (a cura di) (2014). *Mostri del Giappone. Narrative, figure, egemonie della dis-locazione identitaria*. Religion and Thought 1.

3. Calvetti, Paolo; Mariotti, Marcella (eds) (2015). *Contemporary Japan: Challenges for a World Economic Power in Transition*.

4. Moro, Daniela (2015). *Writing Behind the Scenes: Stage and Gender in Enchi Fumiko's Works*. Arts and Literature 1.

5. Mariotti, Marcella; Nobuo, Kazashi (eds) (2017). *New Steps in Japanese Studies: Kobe University Joint Research*.

6. De Antoni, Andrea; Raveri, Massimo (eds) (2017). *Death and Desire in Contemporary Japan: Representing, Practicing, Performing*. Religion and Thought 2.

7. Ruperti, Bonaventura; Vesco, Silvia; Negri, Carolina (eds) (2017). *Rethinking Nature in Japan: From Tradition to Modernity*. Arts and Literature 2.

8. Bulian, Giovanni; Nakano, Yasushi (eds) (2018). *Small-scale Fisheries in Japan: Environmental and Socio-cultural Perspectives*. History and Society 2.

9. Favi, Sonia (2018). *Self Through the Other: Production, Circulation and Reception in Italy of Sixteenth-Century Printed Sources on Japan*. Arts and Literature 3.

10. Mariotti, Marcella; Novielli, Roberta; Ruperti, Bonaventura; Vesco, Silvia (eds) (2018). *Rethinking Nature in Post-Fukushima Japan: Facing the Crisis*. Arts and Literature 4.

11. Berndt, Enno (2018). *J-Economy, J-Corporation and J-Power since 1990: From Mutual Gain to Neoliberal Redistribution*. History and Society 3.

12. Rivadossi, Silvia (2020). *Sciamani urbani. Rintracciando un nuovo discorso nel Giappone contemporaneo*. Religion and Thought 3.

13. Pappalardo, Giuseppe; Heinrich, Patrick (eds) (2020). *European Approaches to Japanese Language and Linguistics*. Linguistics and Language Education 1.

14. Bulian, Giovanni; Rivadossi, Silvia (eds) (2021). *Itineraries of an Anthropologist. Studies in Honour of Massimo Raveri*. Religion and Thought 4.

Per acquistare | To purchase:
https://fondazionecafoscari.storeden.com/shop

15. Negri, Carolina (2021). *Tra corte, casa e monastero. La vita di una donna nel Giappone del Medioevo*. Arts and Literature 5.

16. Mantelli, Alessandro (2021). *E-learning sostenibile per la didattica del giapponese. Progettare per l'apprendimento autonomo*. Linguistics and Language Education 2.

17. Novielli, Maria Roberta; Ruperti, Bonaventura; Vesco, Silvia (eds) (2021). *Italy-Japan. Dialogues on Food*.

18. Dal Corso, Elia (2022). *Materials and Methods of Analysis for the Study of the Ainu Language. Southern Hokkaidō and Sakhalin Varieties*. Linguistics and Language Education 3.

19. Negri, Carolina; Tommasi, Pier Carlo (eds) (2022). *Images from the Past: Intertextuality in Japanese Premodern Literature*. Arts and Literature 6.

20. Gusso, Massimo (2022). *Italia e Giappone: dal Patto Anticomintern alla dichiarazione di guerra del luglio 1945. Inquiete convergenze, geopolitica, diplomazia, conflitti globali e drammi individuali (1934-1952)*. History and Society 5.

21. Rivadossi, Silvia; Franceschini, Cecilia; Ruperti, Bonaventura (2023). *L'estetica del vuoto. Vuoti, pieni, suoni, silenzi, presenze, assenza*. Arts and Literature 7.

22. Lapis, Giovanni (2023). *Religion, Education, and the 'East'. Addressing Orientalism and Interculturality in Religious Education Through Japanese and East Asian Religions*. Religion and Thought 5.

23. Zampol D'Ortia, Linda (2024). *A Failing Mission? Salvation in the Jesuit Mission in Japan Under Francisco Cabral*. Religion and Thought 6.

Japan's ecological variety seems to parallel the multitude of different languages and dialects attested in the Japanese archipelago. In addition to standard Japanese and its dialects, there exist other Japonic languages such as Ryūkyūan and Hachijō, as well as non-Japonic varieties like Ainu. In this volume, four articles explore the importance of a philological approach to sources for historical linguistics: "Adopting a Philological Approach Toward *Chishi*" by Étienne Baudel; "Elements of Sakhalin Ainu Phonetics, Phonology, and Morphosyntax in Bronisław Piłsudski's Corpus of Ainu Folklore" by Elia Dal Corso; "The Language of Miyako Oral Traditions" by Aleksandra Jarosz; "Reflexes of Proto-Ryukyuan Mid Vowels in *Haedong Chegukki*" by Marc Miyake.

Università
Ca'Foscari
Venezia