SCHOOL OF LANGUAGES, CULTURES AND LINGUISTICS
SOAS UNIVERSITY OF LONDON

SOAS WORKING PAPERS IN LINGUISTICS

VOLUME 20, 2021

SOAS Working Papers in Linguistics

Volume 20, 2021

edited by

Adékúnmi Ọlaṫúnjí and Joseph Lovestrand

School of Languages, Cultures and Linguistics
SOAS University of London
10 Thornhaugh Street, Russell Square
London WC1H 0XG
www.soas.ac.uk/linguistics

Published by the School of Languages, Cultures and Linguistics, SOAS University of London, 10 Thornhaugh Street, Russell Square, London WC1H 0XG

ISSN: 1473-0855

Available for free download at:
https://www.soas.ac.uk/linguistics/research/workingpapers/

Contents

Editorial note iv–v

Phonology & Orthography

Lenition and metathesis in Hawu: A quantity-sensitive language 1–25
Shanti Ulfsbjorninn

Preliminary orthographic design for Ramari Dongosaro 26–48
Vasiliki Vita & Chelsea Pedro

Morphosyntax

Tense and aspect comparison between Griko, Modern Greek and Italian for the indicative 49–71
Paris Alexandros Zeikos

Reverse complex predicates or something else? The case of Hindi *de* ‘give’ and *jaa* ‘go’ 72–94
Andriana Koumbarou, Shubham K. Srivastava & Sarada B. Biswas

Temporal information in sentence-final particles: *Tse* and *keh* in Modern Shanghai Wu 95–120
Yan Jiang & Wen Wang

Sociolinguistics

Koronashokku: Loanwords in Japanese and the Covid-19 pandemic 121–141
Tom O'Neill

While we are asleep: Master/Mentor-Apprentice language learning initiative in the Ryukyus 142–164
Miho Zlazli

Reviews in Language Documentation

Interdisciplinary research in language documentation: The benefits and present limits of a more sustainable documentation methodology 165–180
Sarah Round

Linguistic triage: Documentary linguistics and language revitalization 181–189
Victoria Baldwin

A review of *Archiving for the future: Simple steps for archiving language documentation* collections 190–194
Victoria Baldwin

Editorial note

It is with great pleasure that we present the 20th volume of the SOAS Working Papers in Linguistics. This volume comes after a three-year hiatus since the 19th volume was published in 2018. As before, the volume features the work of SOAS graduate students, faculty and alumni. The ten papers in this volume cover a broad range of linguistic disciplines including phonology, syntax, sociolinguistics and language documentation. Papers that focus on particular languages include languages originating from Japan, Indonesia, Palau, Greece, Italy, India and China.

The first two papers have a phonological focus. Shanti Ulfsbjorninn applies a "Strict CV" analysis to Hawu (Malayo-Polynesian) to propose a principled motivation for a rare pattern in which all schwa vowels are stressed. In the second paper, Vasiliki Vita and Chelsea Pedro present initial research on the phonology of Ramari Dongosaro, also known as Sonsolorese, a Micronesian language of Palau. They also outline their recent efforts to engage the community of speakers in building up a consensus on how to write the language.

The next three papers contain morphosyntactic analysis. Paris Alexandros Zeikos compares the verbal morphology of Griko with Modern Greek and Italian to show that Griko (a Hellenic language spoken in Italy) essentially maintains the same verbal morphology as Greek, with the exception of a few paraphrastic conjugations that show some Italian influence. A paper by Andriana Koumbarou, Shubham K. Srivastava and Sarada B. Biswas explores a relatively undiscussed aspect of complex predicates in Hindi in which there is an apparent reversal of the word order of the light verb and the main verb in the construction. They present new data and give an initial analysis of the semantics of these rarer versions of Hindi complex predicates. In the third morphosyntax paper, Yan Jiang and Wen Wang argue against previous analyses of the Shanghainese sentence-final particle *tse* which treat the particle as a tense marker. They instead argue that the particle's primary function is discourse-related, leading to temporal interpretations in specific contexts.

The sixth and seventh papers have a more sociolinguistic focus, both writing about language use and attitudes in Japan. Tom O'Neill presents a study of how well Japanese speakers assess their own understanding of novel loanwords used in discussions of the COVID-19 pandemic. He reviews the literature on loanword use and its motivations, and demonstrates that his findings suggest that, in some cases, novel loanwords can be an impediment to communication. Miho Zlazli writes about linguistic activities involving the Ryukyuan languages of Japan. The paper calls for more reflective approaches to language documentation and revitalization, and illustrates how this is currently being implemented in the author's own work as an insider-linguist.

The final three papers of this volume are topical essays and a review on issues related to language documentation. Sarah Round presents a summary of recent explorations of interdisciplinarity in language documentation, covering the motivations as well as perceived obstacles and possible solutions. Vicky Baldwin summarizes the literature on the relationship between language documentation and revitalization and, in another paper,

provides a concise review of the recently released online learning tool for language documenters called Archiving for the Future.

As part of the editorial process, each contributor to this volume gave feedback on one other submitted paper, and in addition we recruited another person to give expert feedback for each paper. We did not anonymise the review process, and we were pleased that all of the reviewers provided insightful and constructive feedback in a spirit of collegiality. We would like to thank each of the reviewers who gave of their own time to be a part of this process: Peter Austin, Mike Cahill, Vera Ferreira, Maria Flouraki, Weifeng Han, Seiko Harumi, Steve Parker, Julia Sallabank, Naresh Sharma, Aung Si, Martha Tsutsui Billins and Radu Voica.

SOAS Working Papers in Linguistics is an unfunded, volunteer-run open access publication. We thank each of the contributors for their sustained efforts through multiple rounds of comments and proofreading, and for their understanding and patience throughout the editorial and formatting process.

Adékúnmi Ọlaṫúnjí and Joseph Lovestrand
Editors

SOAS Working Papers in Linguistics. Volume 20 (2021): 1–25

Lenition and metathesis in Hawu: A quantity-sensitive language

Shanti Ulfsbjorninn

s.ulfsbjorninn@deusto.es

Abstract
Hawu shows highly unusual distributions of strength/weakness. Firstly, it appears unique in restricting schwa to stressed positions, excluding it elsewhere. Secondly, Hawu has undergone rampant intervocalic lenition of consonants, except after schwa where consonants are immune to lenition and are automatically geminated. This situation creates unusual synchronic structural contrasts. Hawu seems to be unique, appearing to violate a linguistic universal, however, this will be revealed to be due to its phonology-to-phonetics mapping known as virtual length. From a revised phonological perspective, Hawu's strength distribution is actually commonplace: a quantity-sensitive language with metrical bolstering, like Italian. This offers an explanation for, firstly, the otherwise unrelated lack of (C)ə́V word-shapes; and, secondly, the strange diachronic condition on metathesis, that it only applies across consonants: *$V_{\alpha}CV_{\beta}$ > $V_{\beta}CV_{\alpha}$ - *ika > [ə́k:i] 'tie/bind' vs. *VαVβ **(> VβVα) *bua > [búe] **[bə́u] 'fruit'. Both facts are explained via quantity restrictions imposed by Hawu's metrical system.

Keywords: phonology, Austronesian, Italo-Romance, quantity-sensitivity, metathesis, Strict CV

1. Introduction

Hawu (Glottolog code: sabu1255) is a Malayo-Polynesian language spoken by approximately 100,000 people on Savu and Raijua of the Lesser Sunda Islands. The closely related language, Dhao is spoken by some 5,000 people on the nearby island of Ndao. Much of what is discussed in this paper is true also for Dhao, but there are significant differences, and the languages are no longer mutually intelligible. Literature discussing Hawu consists primarily of Blust (2008; 2012), Grimes (2006) and Walker (1982), as well as some older sources whose findings are largely incorporated into Blust (2008).[1]

Hawu remains an understudied language and the analysis that follows is a phonological one based on the pattern reported primarily by Blust (2008; 2012). This paper demonstrates, for the first time, the potentially unique aspects of its phonological system, and hopefully this will spur further work on the language. Specifically, as a reviewer mentions, careful phonetic measurements of the language would be warranted. Though these are unlikely to change the phonological aspect of the analysis that are presented in this paper, they are undoubtedly required to provide a more complete description of the

[1] Thank you to Jacklin Bunga for providing me with audio samples and information on the language, and to the editors Kúnmi Ọlátúnjí and Joseph Lovestrand for their dedication to the volume. Thanks also go to Tobias Scheer and Noam Faust for comments on the manuscript, and finally to Steve Parker for his generous and extensive comments as well as to Andriana Koumbarou whose comments helped improve the structure of the paper.

language's phonetics, and they may indeed throw up some phonologically relevant facts that are hitherto unknown.

Section 2 introduces the reader to basic facts of Hawu and puts its vowel distributions and diachronic consonant-lenition pattern in a typological perspective. It discusses the language's diachronic process of metathesis and presents its resultant synchronic phonological pattern – remarking on its typologically oddity.

Section 3 then provides an overarching explanation for the pattern; it diagnoses Hawu as a quantity-sensitive language with metrical bolstering. The distributions are, structurally speaking, remarkably similar to those of Italian (for instance). This comparison is demonstrated, and it requires understanding Hawu as a language with virtual length. This is a non-transparent mapping between phonological representations and phonetic forms which is robustly attested for Semitic languages, but which is previously unknown in analyses of Austronesian. Since virtual length is not well known, the concept is introduced for the unfamiliar reader in this section.

Section 4 then discusses Hawu's quantity system and word-minimality in light of virtual length. This is done in the framework of Strict CV. The framework's basic assumptions and mechanism are presented for the unfamiliar reader in this section.

Section 5 then provides the core of the analysis. In light of its metrical requirements, this section explains the word-shapes that are synchronically possible in Hawu. It then relates this to an otherwise apparently unrelated fact, the diachronic development of metathesis in the language. Finally, the conclusions are presented in Section 6.

2. Hawu facts and typological implications

Blust (2008) provides an invaluable and detailed diachronic account of the development of Hawu phonology. The Hawu consonant and vowel inventory is listed in (1).

(1) Consonant and Vowel inventories

p	t		k	ʔ	i	
u						
b	d	ɟ	g		e	ə
o						
ɓ	ɗ	ɗʲ	ɠ			a
m	n	ɲ	ŋ			
				h		
	r l					
w						

There are no tautosyllabic consonant clusters. Hawu has many words of the shape $(C)V_x.V_y$, but these sequences are disyllabic: [ni.ŋa.a] 'what' (Grimes 2006); an argument for this is found in Walker (1982). There appears to be only extremely marginal contrastivity of long vowels. They cannot be found, for instance, preceding a consonant: *CVVCV(CV), and when they do occur, they appear only in final position (cf. [niŋaa] 'what'), which may be ideophonic. Stress is also non-contrastive in the language and treating these few final long vowels as disyllables allows for a fixed stress system

(penultimate): [peké.e] ‘neigh’ vs. [péke] ‘tell.SG’ (Walker 1982: 7). The only vowel that is permitted in the antepenultimate syllable is /e/. The canonical word-shape is: (C)(e)(C)V(C)V (Blust 2008: 7).

2.1. Vowel quality and its position in the word

The three requirements of Hawu mentioned so far are that (a) syllables are without consonant clusters, (b) vowel-vowel sequences are disyllabic, and (c) stress is penultimate. These facts create a regular prosodic shape to words. Vowels can be located in three essentially uniform environments: the stressed position, the post-tonic position, and (optionally) the pretonic position. As is shown in (2), Hawu has some strict phonotactic restrictions regarding the distribution of vowel quality and these three canonical positions.

(2) Positions of vowels

Pretonic		**Stressed**		**Post-tonic**	
C	V	C	V	C	V
	{e}		{i, a, u, o, e, ə}		{i, a, u, o, e}

In the pretonic position, all vowels were diachronically neutralized to schwa and subsequently shifted to ‘e’ (Blust 2008: 70): *s**u**maŋəd > h**e**maŋa ‘soul/life force’, *k**a**mali > k**e**mali ‘men’s house’, *k**a**l**i**-wati (> *k**ə**l**ə**wati) > k**e**late ‘earthworm’ (Blust 2008: 69).

Schwa is restricted to stressed positions. It is inherited from *ə, which was also stressed: *bə́qak > [ɓə́k:a] ‘split’, *lə́ku > [lə́k:u] ‘fold’ (Blust 2008: 70).

Elsewhere, *ə shifted to ‘a’ if it became word-final through final consonant loss: *tanəm > dan**a** ‘bury’, *daləm > ɗar**a** ‘in/inside’, *tələn > [dəl:**a**] ‘to swallow’ (Blust 2008: 70).

In addition to ‘a’, the final position also gained ‘e’ and ‘o’ from monophthongisation of *aj: *b-in-ahi (> *bin**aj** > bin**e**) > [bən:i] ‘woman’, *beR**aj** > [wi**e]** ‘give’, and *aw: *pan**aw** > [an**o**] ‘leucoderma’ (Blust 2008: 69). The final position also gained ‘i’ and ‘u’ through metathesis (described in Section 2.4), as well as inheriting these vowels directly: *lək**u** > [lə́k:**u**] ‘fold’.

Returning to the restrictions on schwa, there is no stressed schwa in a (C)V_xV_y word structure: **(C)ə́V (Blust 2012: 7).[2] Blust (2008) attributes this to a further (more general) condition that schwa is never found in contact with another vowel (Blust 2012). A condition in the earlier historical condition that occurred prior to Proto-Sumba-Hawu where ***ə*** is deleted when in contact with a vowel: *qah**ə**lu (> *qa**ə**lu) > [alu] ‘pestle’, *ma-buh**ə**k (> *mabu**ə**k > mabuk) > [mawo] ‘drunk’ (Blust 2008: 70). This historical contingency eliminates a large potential source for (C)ə́V sequences. Because schwa in Hawu is inherited from ***ə***, Hawu sequences of the shape (C)ə́V are primarily expected to develop from *(C)əV sequences, however these had already shifted to (C)V (Blust 2008). There is no condition banning prevocalic Vs (*keli > [kei] ‘dig/dig up’), however, the

[2] Grimes (2006) does give a few such forms, though they are mainly sub-minimal words such as grammatical particles. These do not appear in Blust (2008). Schwa is claimed not to be found before vowels (Blust 2012: 7), perhaps with a few exceptions.

restriction on (C)əV sequences is still active. This condition appears to be synchronically enforced because there have been no subsequent vowel shifts to (C)əV, despite the ample opportunities for these sequences to form via metathesis (Section 2.4).

2.2. Typological observations

The synchronic vocalic restrictions of Hawu are summarized here.

(3) a. Schwa can only occur in stressed positions; (corollary) schwa is banned from prosodically weak positions.
b. Schwa is banned from (C)əV sequences.
c. Only ‘e’ is allowed in pretonic position.

While it is common in related languages of the region to neutralize pretonic vowels to a vowel other than schwa, it is not common for this to be done in a language that also has a schwa vowel.

Timugon Murut (phonologically schwaless) limits the pretonic positions to ‘a’ (or ‘o’ under vowel harmony (Kroeger 1992; Barnes 2003)). This proceeded via an earlier schwa stage (Blust 2008: 69; cf. Barnes 2003). Hawu is particular in only having pretonic ‘e’ (Blust 2008).

When it comes to the distribution of schwa, Hawu is shown to go against typological preferences. Almost universally, schwa constitutes a phonologically weak vowel. Fittingly, it is often restricted to unstressed positions and weak positions, where it is additionally often elided (merging with zero). Due to this deleteability in weak positions, a number of languages ban word-final schwa, such as Hindi (Pandey 1990 and references therein), however, perhaps no other language than Hawu bans schwa from *all and only* prosodically weak positions.

Typologically, schwa is frequently systematically banned from stressed positions. English is one well-known example; it has pre- and post-tonic schwa, but schwa is excluded from the tonic position (Harris 1997; Gussmann 2002: 125). Many other languages share this restriction, to name a few non-related examples, one could give: Indonesian (Cohn 1989), Javanese (Ras 1982), Dutch (van Oostendorp 2000; Flemming 2007), and Tundra Nenets (Salminen 1993).

Yet other languages have schwa in all positions and only stress the schwa if no full vowel is present in the word. In these languages, the placement of fixed stress shifts away from its canonical position when the canonical position contains schwa: T’boli (Southern Mindanao, Austronesian) (Awed, Underwood & Van Wynen 2004), Tondano (Sneddon 1975) and Sye (Vanuatu) (Crowley 1998). Another common situation is for languages to have prosodically irrelevant epenthetic or intrusive schwas (for careful discussion and many examples, see Hall (2006). Finally, while there are some few languages that permit stressed schwas, such as some dialects of Catalan and Zabiče Slovene (Crosswhite 2001), all these languages *also* allow schwa in prosodically weak positions.

To my knowledge there is no language like Hawu where schwa is entirely restricted to the stressed position. Hawu seemingly violates a universal constraint against schwa being restricted to stressed positions / excluded from all weak positions.

2.3. Consonant lenition

The diachrony of Hawu is characterized by much consonant lenition and deletion; the conditions on these are informative. Consonants in the initial position were only sporadically affected by lenition, whereas final consonants were all deleted in Hawu (Blust 2008: 64). Elsewhere, the intervocalic position underwent regular and pervasive weakening, as is shown in (4). Curiously, there was only one intervocalic context where consonant lenition was resisted: following a schwa. This situation creates some unusual structural contrasts.

As shown in (4), all consonants are automatically geminated after a stressed schwa (Walker 1982). Blust (2012) hypothesizes that the geminate status of these post-schwa consonants allowed them to resist lenition via the well-known effect of geminate inalterability (Kenstowicz & Pyle 1973; Guerssel 1978; Kenstowicz 1994). The reason for this special gemination will be explained in Section 3.

(4) Intervocalic lenition and contrast in Hawu (based on Blust 2008: 64–69)

			Weak				**Strong**			
*p	>	(p ~ ∅)	*ní**p**i	>	ní	‘dream’	*ə**p**at	>	ə**p:**at[3]	‘four’
*t	>	(t ~ d)	*má**t**a	>	má**d**a	‘eye’	*bə**t**ak	>	wə**t:**a	‘split’
*k	>	(k ~ ʔ)	*paní**k**i	>	ní**ʔ**i	‘fruit bat’	*lə**k**uq	>	lə**k:**u	‘fold’
*b	>	(b ~ w)	*qá**b**u	>	qá**w**u	‘secret’	*sə́**b**u	>	hə́**ɓ:**u	‘stream’
							*tə**b**uh	>	də̃**b:**u	‘cane’
*d	>	(d/ɗ ~ r)	*tá**d**a	>	tá**r**a	‘cockspur’	*kə́**d**en	>	kə́**ɗ:**u	‘stand’
*ɟ	>	(ɗ ~ r)	*pá**ɟ**ay	>	pa**r**e	‘rice’	*qapə́**ɟ**u	>	pə́**ɗ:**u	‘gall’

Typologically, post- and inter-vocalic spirantisation is extremely common, but it is far less common for the quality of the vowels to be a determining factor. Conditions such as ‘lenite after labial but not coronal vowels’ is entirely unheard of. Moreover, it is not standardly expected for intervocalic lenition to be blocked in post-tonic position. Indeed, post-tonic intervocalic positions are one of the weakest possible environments (Harris 1997). Consider the English pattern: [tʰɒ́m] ‘Tom’, [tʰɒ́ʔə] ‘totter’, [ǽʔəm] ‘atom’, [ətʰɒ́mık] *[əʔɒ́mık] ‘atomic’.[4]

2.4. Metathesis

Hawu is perhaps best known for regular diachronic metathesis (Blust 2008; 2012).[5] The generalization can be stated as in (5).

(5) Metathesis generalisation

a.	***If***	V_2[–high]	***prec***	C	***prec***	V_1[+high]
	Then	V_1[+high]	***prec***	C	***prec***	V_2[–high]

[3] Blust (2008) and subsequent work never marks consonantal phonetic length of geminates because it is entirely predictable, this will be marked consistently in this paper as their length is key to the analysis.
[4] English examples of post-tonic intervocalic lenition after schwa cannot be supplied because schwa is not strong enough to occupy this position, again highlighting Hawu’s unusual strength distributions.
[5] This has been reanalysed as pseudometathesis (feature spreading) by Lysvik (2015). The argument in this paper speaks to either proposal.

b. ***And*** V_1 shifts to [ə]
C geminates

Though there is one exceptional form (*binə > [bə́n:i] ‘female/woman’), the [–high] vowel in the generalisation is almost exclusively ‘a’ (Blust 2008: 71). The metathesized word-shapes are therefore practically always iCa or uCa (ibid.). The order of the high and non-high vowel is reversed (6a) and in all cases, the metathesized ‘a’ becomes schwa once it is stressed.

(6) Metathesis

a. *iCa > əC:i

*miŋak	(> *miŋ**a**)	>	[mə́ŋ:i]	‘fatty, oily’
*lim**a**		>	[lə́m:i]	‘five’
*piɟa	(> *pir**a**)	>	[pə́r:i]	‘how (many)’[6]

b. *uCa > əC:u

*pusəj	(> *uh**a**)	>	[ə́h:u]	‘navel’
*kudən	(> *ur**a**)	>	[ə́r:u]	‘cooking pot’
*sukat	(> *suk**a**)	>	[hə́k:u]	‘measure’

In addition to these facts, there is a curious condition on Hawu metathesis: the C shown in generalization (5) is obligatorily part of the context. Crucially (but very strangely), metathesis can only apply if a consonant intervenes: (a) *buaq (> *bu.a) > [bu.e] **[bə́.u] or **[bá.u] or **[bu(:)] ‘fruit’, (b) *liaŋ (> *lia) > [li.e] **[lə́.i] or **[lá.i] or **[li(:)] ‘cave’ (Blust 2008: 71; 2012).

There is no natural explanation for metathesis obligatorily occurring across a consonant. Especially because in Hawu all V_xV_y sequences are disyllabic. This means that the metathesis is being blocked specifically by an empty onset (C-slot in the syllable structure). This point is particularly evident considering the pseudo-metathesis account presented in Lysvik (2015).[7]

(7) Empty onset blocks metathesis

(a) *bula > [bə́l:u] ‘forget’ (b) *bu.a > [bu.e] **bə́u ‘fruit’ (no metathesis)

[6] Here (and elsewhere) stops are shown to lenite before metathesis, however, many other forms show the metathesised form without lenition. In all cases, the consonant is automatically geminated after schwa (Walker 1982; Blust 2012).
[7] Though this paper is focused on metathesis in general and does not discuss CV.V forms

Lysvik (2015) represents this VV metathesis as final vowel loss plus feature spreading (7a). Whatever the mechanism, (7b) further shows how arbitrary it is for metathesis to be blocked by an empty onset.

2.5. What requires analysis?

The Hawu facts that require discussion and reanalysis or explanation are summarized in (8).

(8) Explananda

a. Schwa, a universally weak vowel, can only occupy the stressed position. It is excluded from prosodically weak positions. (This distribution appears unique to Hawu.)

b. Schwa triggers gemination of a right-adjacent C. Diachronically this allowed C to resist lenition, unlike after any other vowel: [lə́k:u] ‘fold’ vs. [níʔi] ‘fruit bat’.

c. There are no Cə́V sequences. Metathesis is blocked in $CV_{[+high]}.V_{[-high]}$ forms. Metathesized final /a/ becomes schwa in stressed position.

3 Hawu as a quantity-sensitive language and its schwa

3.1 Schwa in Hawu

I start by addressing explanandum (a), the unusual distribution of schwa. It is helpful to distinguish between phonetic and phonological definitions of schwa. Phonetically, schwa refers to a vowel quality approximated by the IPA symbol: [ə]. Phonologically, however, a schwa is a ‘featureless vowel’, or a null set dominated by a nucleus (a.k.a. an empty nucleus (Charette 1991)). The phonetic interpretation of this empty nucleus is language specific. Examples are shown in (9).

(9) Empty nuclei and phonetic interpretation (for more, cf. Scheer 2004)

	Phonetic Quality	**Language**	**Source**
a.	[ə]	French	(Charette 1991)
		Tocharian	(Koller 2008)
		Upper Chehalis (Salish)	(Kinkade 1998)
		Malayalam	(Sadanandan 1999)
		Chukchi, Itelman, Karo Batak, Ladahki, Malay, Wolof	(DeLacy 2002 Lombardi 2002)
b.	[ɨ]	Moroccan Arabic	(Kaye 1990)
		Turkish	(Charette & Göksel 1998)
		Pattani Malay	(Topintzi 2010)
c.	[ɯ]	Japanese	(Ito & Mester 1995)

d.	[i]	Yawelmani	(Archangeli 1984)
		Arabic dialects (Lebanese)	(Abdul-Karim 1980)
		Yoruba, Samoan	(Uffmann 2007)
		Tongan	(Kitto 1997)
		Cantonese	(Yip 1993)
		Navaho	(Lombardi 2002)
e.	[e]	Hebrew	(Bolozky 2005)
		Central Italian (Word-Final)	(Repetti 2012)
		Gengbe (at least in initial position)	(Lombardi 2002)
		Sawai (in final position)	
f.	[ʌ]	Tundra Nenets	(Lombardi 2002) (Kavitskaya & Staroverov 2008)
g.	[a]	Axininca, Lardil, Marathi, Balochi, Brahui, Sundanese, Wapishana, Iraqw, Klamath, Dakota, Coos, Mabalay Atayal	(Lombardi 2002)
h.	[ɔ]	Bengali	(Heimisdottir 2013)

3.1.1. Schwa allophony in Hawu

The hypothesis I pursue for Hawu is that only *phonetic* schwa is restricted to stressed positions. Phonological schwa can occur in all positions: pretonic, stressed and post-tonic positions, however, phonetically it manifests differently in pretonic and post-tonic position, not unlike Italian, with its initial [i] epenthesis and its final [e] epenthesis (Repetti 2012).

I claim there are three allophones of schwa. Schwa surfaces as [e] pretonically, [ə] in stressed position, and [a] post-tonically. Diachronically, the development of pretonic e-schwa and post-tonic a-schwa result from neutralisations to or from a schwa stage. I argue these have since only changed their phonetic quality rather than their featural identity.

Hawu's pretonic [e] originates from the neutralization of all vowels to schwa: *s**u**maŋed (> *h**ə**maŋa) > [h**e**maŋa] 'soul/life force', *k**a**mali (> *k**ə**mali) > [k**e**mali] 'men's house', *k**a**l**i**-wati (> *k**ə**l**ə**wati) > [k**e**late] 'earthworm' (Blust 2008: 69). This first step in this change, the neutralization to schwa, is consistent with lenition (and loss of vocalic distinctive features) in unstressed positions such as the pretonic (Crosswhite 2001; Harris 1997).

The final position also underwent neutralization, though a less drastic one. Word-final schwa *ə merged with final *a: *tan**ə**m > [dan**a]** 'bury', *dal**ə**m > [ɗar**a]** 'in/inside', *t**ə**l**ə**n > [d**ə**l:**a**] 'to swallow' (Blust 2008: 70). The merger with 'a', is attested in a round-about way. It is widely accepted that [ə] can be the phonetic expression of an unstressed 'a' (cf. Malagasy). The final schwa of Modern London English[8] is actually phonetically

[8] This is usually referred to as Multicultural London English, a term I object to as I find all English is multicultural.

pronounced with a low vowel: [bɹávɐ] ‘brother’. This is also a feature that it shares with other English varieties, New Zealand (one variant) (Bauer & Warren 2004), Fiji (Tent & Mugler 2004), [ʌ] in Pakistani English (Mahboob & Huma Ahmar 2004) and many more. The hypothesis for Hawu is that unstressed [a] is a ‘schwa’, that is, still phonologically featureless. Phonologically, schwa is not restricted to the stressed position; it is found in all positions.

(10) Distribution of schwa and phonetic interpretation

		Pretonic		**Stressed**		**Post-tonic**	
		C	V	C	V	C	V
If	Phonology:		{ }		{ }		{ }
Then	Phonetics:		[**e**]		[**ə**]		[**a**]

Consequently, Hawu does not have a unique distribution of vocalic strength and weakness, rather Hawu has the typologically ordinary pattern of permitting less contrasts in pretonic and post-tonic positions, as well as hosting a larger array of contrasts in stressed position (cf. Crosswhite 2001). This analysis is confirmed by the alternations from metathesis that will be explored later: suk**a**t (> *suk**a**) > [h**ə́**k:u] ‘measure’.

3.2. Gemination, lenition and schwa

Although schwa has no phonological positional restrictions, there is one strong phonotactic restriction on a stressed schwa, stated in (11). Its discussion takes us directly into the heart of the second explanandum (8b).

(11) **If** schwa is in stressed position, **then** it must be in a closed syllable (preceding a geminate)

This distribution of schwa is similar to a family-wide preference in Salish languages to ban schwa from open syllables (Kinkade 1998: 209). These languages often feature əC.C vs. á.CV alternations such as this one from Upper Chehalis: s-p**ə́l**xw-t-n & sp**ála**xw-t-n, s-p**ə́l**’xw-n & p**ála**xw-n ‘doctor, cure.transitive’ (Kinkade 1998: 206). Like Hawu, schwa is permitted to be stressed but only in closed syllables.

The previous section demonstrates that, in stressed position, schwa is restricted to closed syllables. This observation has wide reaching implications that speak to the second explanandum (8b), repeated here for convenience.

(12) Explanandum (b)

> Schwa triggers gemination of a right-adjacent C. Diachronically this allowed C to resist lenition, unlike after any other vowel: [l**ə́k:**u] ‘fold’ vs. [n**íʔ**i] ‘fruit bat’.

The relevant structural contrast is: $\acute{V}C_{(weak)}$ vs. $\acute{ə}C_{(strong)}$. Abstracted from its quality, the pattern can be interpreted as: $V_{(strong)}C_{(weak)}$ vs. $V_{(weak)}C_{(strong)}$. This configurational contrast is highly reminiscent of the weight trade-offs between Vs and Cs in languages with Metrical Lengthening (i.e. Standard Italian).

(13) Metrical Lengthening in Standard Italian[9]

	$V_{(strong)}C_{(weak)}$				$V_{(weak)}C_{(strong)}$		
a.	[ví:peɾa]	**[vípeɾa]	'viper'	vs.	[dít:eɾo]	**[dí:t:eɾo]	'fly genera'
b.	[muɾé:na]	**[muɾéna]	'moray eel'	vs.	[pét:o]	**[pé:t:o]	'chest'

Standard Italian shows this structural strength-based opposition through phonetic length: (V:C vs. VC:). Tuscan (Italian) goes further by maximizing this strength-based opposition by expressing it with both phonetic length and segmental quality (strong = stop vs. weak = fricative) (Marotta 2008; Ulfsbjorninn 2017).

(14) Gorgia Toscana and length

a.	/pɔ́ko/	[p**ɔ́:x**o]	'few/little'
b.	/kɔ́k:o/	[k**ɔ́k:**o]	'coconut'

The Tuscan pattern reveals a structural contrast based on both length and strength of the consonant. In stressed positions, long vowels are accompanied by lenited singletons and short vowels accompany strong, unlenited consonants. The parallel with Hawu is sketched in (15).

(15) Strength and length distributions in stressed positions in Tuscan and Hawu

a. Tuscan

ɔ́:x		**ɔ́k:**	
V [long]	C [continuant]	V [short]	C [occlusive]
Strong	*Weak*	*Weak*	*Strong*

b. Hawu

íʔ		**ə́k:**	
V [full]	C [short]	V [schwa]	C [long]
Strong	*Weak*	*Weak*	*Strong*

The Tuscan contrast shown in (15a) demonstrates that a strong long vowel is offset by a weak lenited non-occlusive consonant. Conversely, Tuscan also has a weak short vowel that is compensated for by a strong consonant that is long and can be occlusive.

It is not the phonetic property of the vowel that causes the strengthening of the consonant. That misunderstanding underlies the apparent paradox embodied in explanandum (b): "schwa (the weakest of vowels) causes the strengthening of consonants." In fact, it is the strength inherent to the stressed position that is not satisfied by a weak vowel. Neither the short vowel of Tuscan, nor the schwa of Hawu can satisfy stress. In both languages, the stressed position requires more in the way of phonological weight than a short vowel or a schwa can provide. In both languages, however, a stressed position can contain a weak

[9] This refers to Standard Italian 'Variety A'. Vowel length is not restricted to open syllables in penultimate position.

vowel, but it is compensated for by the strength of an adjacent consonant. I propose therefore that, like Tuscan Italian (and also Norwegian, Icelandic and many other languages), Hawu is quantity-sensitive with metrical lengthening/bolstering.

3.2.1. Metrical lengthening/bolstering in Hawu

I will present the argument here that, like Tuscan, Hawu has a prosodic condition on the number of objects in the stressed position. Expressed in standard terms, all words must contain a syllable dominating two moras. In Hawu, the first of these two moras corresponds to a schwa, while the other comes from the geminate: (l**ə**$_{\mu}$**k**$_{\mu}$)$_{\sigma}$.ku ‘fold’. In order to be explicative, this weight requirement is minimal and maximal. The stressed position must minimally and maximally contain two units of weight: (a) Italian - [k**á:**ne] **[káne] *cane* ‘dog’ & [k**án**:e] **[ká:n:e] *canne* ‘spliffs’ (b) Hawu - [l**ə́k**:u] **[lə́ku] ‘fold’.

For Tuscan this minimal-maximal requirement is phonetically evident in all forms, but in Hawu this phonological status must be inferred. However, we know it must be the case with an argument that follows logically from something we can directly observe: the schwa’s behaviour under stress. Even remaining agnostic about the representation, the argument is sketched out in (16).

(16) Argument for minimally ‘heavy rimes’ in Hawu

a. **Observation**	In Hawu, a schwa is not enough, on its own, to occupy a stressed position (it obligatorily induces gemination).
b. **Axiom**	Consistent with autosegmental principles: Where *x* is melody and *y* is a position/skeletal slot/mora, *x* may associate to *y*. If *x* associates to *y*, then *x* has the value of 1*y*, if *x* associates to *y* and another *y* it has the value of 2*y*.
b′. **Corollary**	Schwa cannot occupy half a skeletal slot, or half a mora, or half a unit of weight, it is either linked or not linked.
c. **Conclusion**	We know a schwa occupies minimally one position *and* we know that the schwa on its own is not enough to satisfy the requirements of a stressed position.
d. **Observation**	A schwa and a geminate satisfy the requirement together.
e. **Inference**	The geminate occupies a second position.
f. **Conclusion**	A stressed position requires minimally two positions: C**ə́**$_{\mu}$C$_{\mu}$.CV, CV́.CV = CV$_{\mu\mu}$.CV

In both languages, a coda consonant or geminate will be present in the phonetic form. From this, the learner can deduce the weight it provides to its syllable. In addition to this, as an extra phonetic cue, Italian /CV́CV/ forms are bolstered by a prolonged duration of the vowel: [CV́:CV].

Hawu must be the same because, as is demonstrated in (16), stressed positions must contain a heavy syllable. However, this vowel does not have to be phonetically marked by extra duration, evidently because Hawu has phonetic [CV́CV] sequences. However,

Hawu is only different from Tuscan because what Tuscan does with duration, Hawu does with vowel quality.

In stressed position, a full vowel quality [a,e,i,o,u] corresponds to a phonologically heavy nucleus, in standard terms, full phonetic vowel quality indicates a bimoraic phonological object. It behaves like a heavy vowel of Tuscan, regardless of its phonetic length.

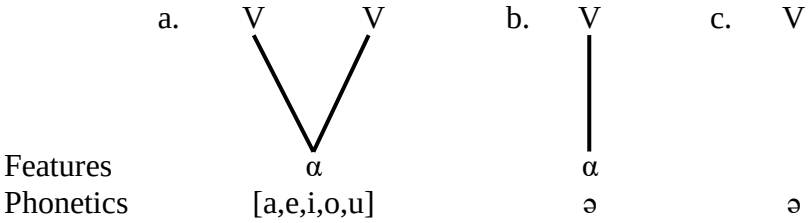
A phonetic schwa, on the other hand, does not correspond to a phonological heavy/bimoraic object. Since schwa is monomoraic, it behaves like a short vowel of Tuscan and it is obligatorily followed by a geminate in its second stressed position. This way, in Hawu, just as in Tuscan, all stressed positions are phonologically heavy: C**ə$_{\mu}$C$_{\mu}$** or C**[á/é/í/ó/ú]$_{\mu\mu}$**.

This mapping of phonological weight to vowel quality rather than duration is known as virtual length. Because not all readers may be familiar with this concept, I introduce virtual length briefly in the following subsection.

3.2.2. Virtual length

Virtual length is the term for a certain kind of phonology-to-phonetics mapping. At the phonological level, it involves bipositional structures (one-to-many relations) that define *phonological length* (Scheer 2014), defined by *bipositionality* (occupying two positions). Virtual length is a situation where the bipositionality of a phonological structure is phonetically translated as anything other than increased duration. There is a large phonetic range of properties that have been argued to correlate with phonological bipositionality (see Scheer 2014 for a summary). The type of mapping that relates to Hawu is one robustly and importantly synchronically attested in various Afro-Asiatic languages. A number of these languages are shown to map phonetic schwa to a monopositional/short vowel, while vowels of a 'full' phonetic quality [a,e,i,o,u] are phonologically bipositional, attached to two positions of syllable structure (despite being phonetically short). For the argumentation as relates to Semitic and Kabyle Berber see Lowenstamm (1991; 2011)(1991, 2011) and Ben Si Saïd (2011), respectively.

(17) Virtual length (Scheer 2014)



The diagram in (17) shows how a bipositional phonological object, in opposition with a monopositional object, can be phonetically interpreted as a short vowel with a full vowel quality. Meanwhile, a monopositional structure is interpreted as schwa. A featureless schwa vowel would also be the phonetic interpretation of a single V position, except that the V is not attached to phonological features. This difference is phonetically neutralized, but the distinction can be shown by various processes depending on the language such as

(17b) revealing its quality when the vowel is lengthened unlike (17c), which is featurally empty, and which can alternate with zero (see the literature above for examples).

4. Hawu vowel quality and phonological weight

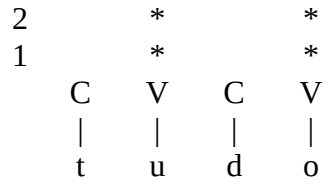
The metrical framework used here is based on Ulfsbjorninn (2014), a Strict CV grid-theory of word stress that builds on Scheer and Szigetvari (2005). Because it is a rather new approach, the metrical model is introduced in the following subsection. Then we will return to discussion of the weight requirements and phonetic interpretation of that weight in Section 4.2.

4.1. Strict CV metrics

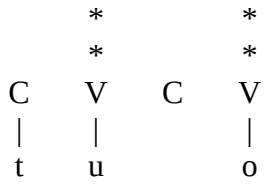
This approach models phonological quantity without using syllables or moras. It is a grid system where metrical structure is projected directly from the skeletal tier (CV). Only V-slots can metrically project. The projection of empty V-slots is parametrically controlled, whereas filled V-slots inherently project a metrical position. Filled V-slots project to Line 2. Empty V-slots (including the second position of long vowels and diphthongs) project to Line 1 (Ulfsbjorninn 2014). The different height of projection reflects their head-dependent status. The various configurations/word shapes are shown below in (18). The forms sketched there are not shown with the final step – the winning projection that determines weight (those are in 19). For a recent publication in this framework, see Faust & Ulfsbjorninn (2018).

(18) Basic projections with hypothetical forms. Light (a-b) and Heavy (c-e)

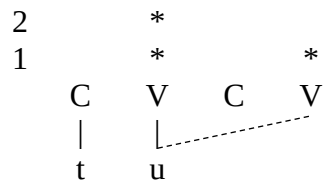
a. CVCV /tudo/



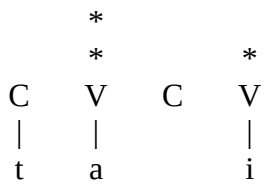
b. $CV_x.V_y$ /tu.o/ *hiatus*



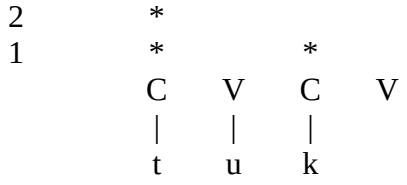
c. CV_xV_x /tu:/ *long vowel*



d. CV_xV_y /tai/ *diphthong*



e. CVC /tuk/



What is traditionally taken as a "heavy syllable" is driven by Incorporation. In this model quantity is the projection of a V-slot according to its phonological environment. This is

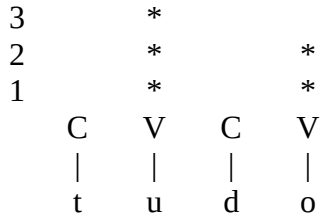
an asymmetric relationship between nuclei; a filled V-slot (Line 2) gains a level in the projection if it linearly precedes a dependent V-slot (Line 1). When this occurs, it is said that a filled V-slot incorporates another V-slot; this has a function of "identifying" empty V-slots (Faust & Torres-Tamarit 2017) and distinguishing hiatuses from diphthongs.

This is shown in (19). The forms in (19a-b) are phonologically light, while (19c-e) are the phonologically heavy ones, where the winning V-slot occurs before a dependent V-slot.

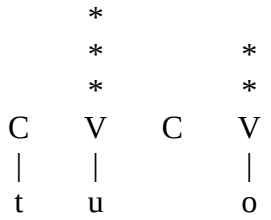
In the structures beneath, for the reader's visual convenience the grid-slots that are achieved by V-slot incorporation (phonological quantity) are shown with an alpha index. This is purely to assist the reader; the indices have no ontological status in the representation.

(19) Incorporation and projection

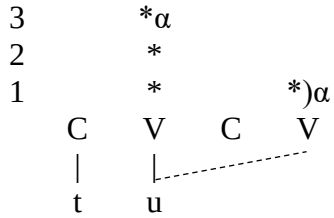
a. CVCV /tudo/



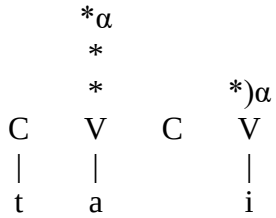
b. $CV_x.V_y$ /tu.o/ *hiatus*



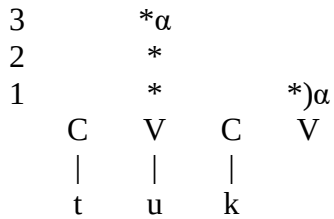
c. CV_xV_x /tu:/ *long vowel*



d. CV_xV_y /tai/ *diphthong*



e. CVC 'tuk' *closed syllable*



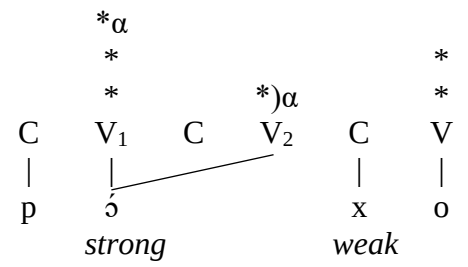
I hypothesise that full quality vowels of Hawu, when located in the stressed position of open syllables, are underlyingly bipositional. They are in effect "bimoraic" but their bimoracity is expressed through vowel quality rather than increased duration.[10]

[10] I assume this is true only for vowels in the stressed position, though it's not clear what evidence could be used to test the weight of non-stressed positions. It is reasonable to assume, I believe, that a learner would, in absence of evidence elsewhere, assume that a virtual length interpretation of vowel quality applies only in stressed position.

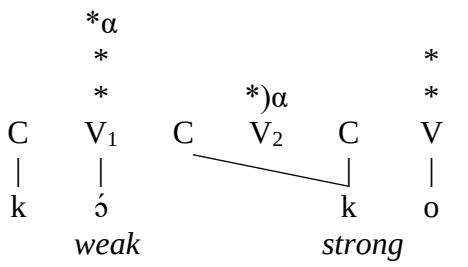
(20) Weight and strength in Tuscan and Hawu

a. Tuscan

[pɔ́:xo] ‘few/little’

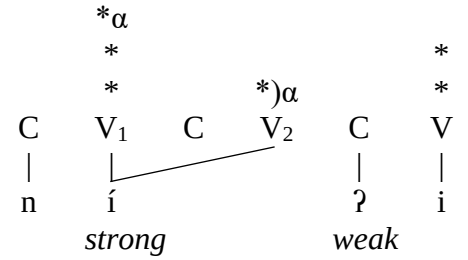


[kɔ́k:o] ‘coconut’

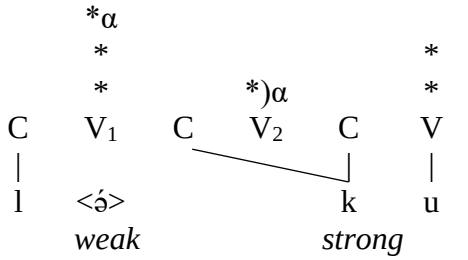


b. Hawu

[níʔi] ‘fruit bat’



[lə́k:u] ‘fold’



As the comparison between Hawu and Tuscan shows, in both languages, and in all forms, the V-slot that projects main stress (V_1) must incorporate a V position to its right (V_2). Having incorporated a V-slot at Line 1, the incorporating V-slot (V_1) gains a grid-mark and projects to Line 3.

4.2. Word Minimality in Strict CV Metrics

Reaching Line 3 is the core metrical prerequisite of Hawu; it is equivalent to standard treatments of minimal word conditions (e.g. LexicalWordProminence (Prince & Smolensky 1993)).

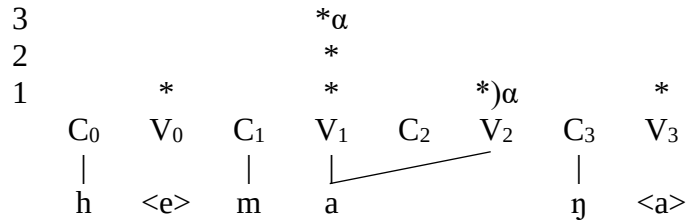
(21) ‘Minimal word’ Domain-hood condition in Hawu

A Domain must contain a metrical head (V-slot) that projects to Line 3

Head:	Penultimate V-slot
Incorporation (obligatory)	V-slot at Line 2 incorporates a V-slot at line 1

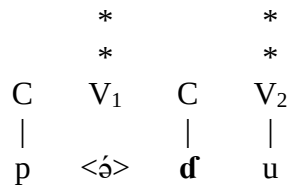
As explained in Section 3.2.2, vowels with a full phonetic quality are underlyingly bipositional (a.k.a. bimoraic). Although they are expressed as phonetically short, this phonological bipositionality allows them to satisfy the metrical conditions of the minimal word (see (20b) in analogy with Tuscan (20a)). A further Hawu example is shown beneath. Epenthetic vowel qualities are shown in angle brackets <>, they follow the rule of schwa allophony explained in Section 3.1.

(22) /hmáŋ/ [heməáŋa] ‘soul/life force’



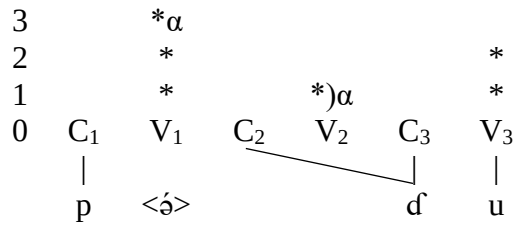
The word shown in (22) has a pretonic metrically weak (CV_0) corresponding to its e-schwa allophone. Then it has the metrical head (CV_1) with its (full quality) bipositional vowel, followed by the metrically weak (CV_3) and its a-schwa allophone. In (22), the metrical head is V_1. At the metrical level, the head incorporates the position (V_2). The incorporated (CV_2) is occupied melodically by the bipositional vowel that spreads across V_1 and V_2. The fact that V_1 incorporates V_2 allows V_1 to project one level further to Line 3, thereby satisfying the metrical condition of domain-hood. Schwa, on the other hand, is the phonetic expression of an empty nucleus, a single V-slot. On its own, a schwa cannot project to L3. As shown in (23), this results in a word shape that does not meet the minimal word condition.

(23) Ill-formed word **[pə**ɗ**u][11]



A word of the shape CəCV is illicit because its metrical head (V_1) has not incorporated a position. The domain-hood condition requires that V_1 incorporates a position. The CV position can be supplied by epenthesis (if needs be) (Larsen 1998; Scheer 2004; Bucci 2013), however, the schwa, being featurally empty, has nothing to spread into that empty CV. So as not to leave the position empty, a consonant (C_3) can instead spread into the position. Gemination ensues. This is shown in (24); C_3 fills the empty (CV_2) and forms a geminate. Meanwhile, the schwa’s position, the domain-head (V_1) incorporates the V-slot (V_2) and projects to Line 3 satisfying the domain-hood condition.

(24) Geminate satisfying weight of stressed schwa, [pə**ɗ**ːu] ‘gall’



[11] Stressed positions are assumed to project to line 2 despite being (potentially) empty. This may be how underlyingly stressed V slots are marked in the lexicon.

In the absence of any other heterosyllabic configuration (there are no other "coda-onset" sequences), the geminate is the only structure of Hawu that can follow a schwa. It is for this reason that a stressed schwa can only ever be found in a "closed syllable".

This quantity-sensitive account of Hawu explains (automatic) gemination after a schwa as well as its absence after any full vowels. These are phonologically bipositional (bimoraic), as such they are large enough to incorporate a V-slot and satisfy the quantity requirement on domain-hood. This analysis fully resolves explanandum (b).

5. Implications for word-shapes and metathesis

The previous section reveals Hawu to be a quantity-sensitive language and concludes with a definition of the metrical condition licensing its words. These metrical conditions also explain two other facts about the language that have been grouped together as explanandum (c): firstly, that (C)ə́V words are illicit, and secondly that metathesis has the strange condition that it must apply over a filled onset/consonant.

5.1. Motivating the ban on (C)ə́V

As has already been shown, (C)ə́CV words are illicit in Hawu (shown in (25)). In these structures, the metrical head does not project high enough to meet the "minimal word" domain-hood condition. When the schwa is a metrical head, it needs to be bolstered by incorporating a dependent CV. However, because schwa is featureless it cannot spread, ruling out forms such as **(C)ə́:CV. This is resolved with gemination, (C)ə́C:V is a licit word structure (shown in (26)). In this form, the dependent CV (CV_2) is present and it gets featurally filled by a consonant. From this discussion, the restriction on **(C)ə́CV words falls out automatically.

(25) Illicit structure of **(C)ə́V words

2		*		*
1		*		*
0	C	V_1	C_2	V
	|			|
	p	<ə́>		u

The metrical head (V_1) in structure in (25) cannot project sufficiently to meet the quantity condition on domain-hood. Consequently, an extra empty CV should be inserted. The difficulty, however, lies in how to fill the empty structure that is metrically required. This is shown in (26); the large span of featurally-empty pieces of structure are shaded grey. The incorporation of V_2 by V_1 and its projection to Line 3 is counterfactual here because the structure is not licit.

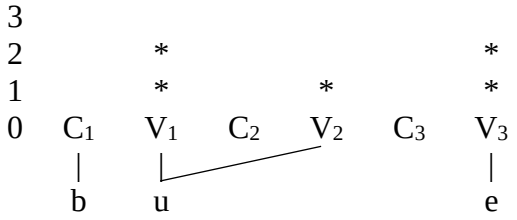
(26) Empty structure in **(C)ə́V word

3		*α				
2		*				*
1		*		*)α		*
0	C_1	V_1	C_2	V_2	C_3	V_3
	|					|
	p	<ə́>				u

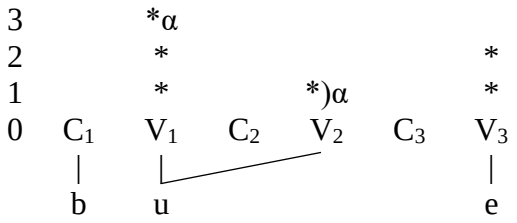
As we have already seen, the schwa in V_1 is featurally empty, therefore it is not able to spread and occupy the empty structure (CV_2). If (26) were a (C)ə́CV word, the content of C_3 would spread into C_2 forming a geminate, but in (26) C_3 is empty, so that may not spread either. The emptiness of C_3 means that the usual means by which schwa-headed words to gain weight are missing. There is simply nothing local to spread into all the empty structure.[12,13] This discussion shows that there is no way of licensing structures such as (26), thereby neatly explaining why (C)ə́CV are unattested in Hawu. Conversely, $CV_x.V_y$ words are fully permissible. Their dependent CV is occupied by bipositional, full-quality vowels; their structures are shown below.

(27) Empty structure in CV.V word [bu.e] ‘fruit’

a. Underlying form



b. Computed form



As is shown in (27), the spreading of the full quality vowel leaves only the onsets C_2 and C_3 as empty categories. These are straightforwardly licensed by their adjacent filled nuclei.

5.2. Explaining why metathesis is blocked by empty onsets

The discussion so far handles the first part of explanandum (c): the impermissibility of (C)ə́CV word shapes. The second part of explanandum (c) relates to the strange condition that metathesis is blocked over an empty onset (C-slot). As we discussed in Section 2.3, Blust (2008; 2012) describes Hawu metathesis as a swapping of the order of vowels when

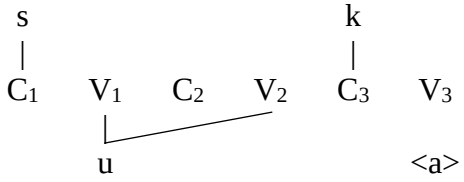
[12] The ‘u’ of V_3 cannot spread to V_2. V_1 is the metrical head and /u/ spreading would create an ill-formed iamb (that is, the designated metrical head would not reach its required projection because its position was coopted by another V.

[13] Government Phonology and Strict CV have developed strong and simple formal conditions on the licensing/permissibility of empty structure. The precise formal details are highly technical and their exposition here would not help the reader accept the analysis. Therefore, I limit exposition of formal mechanisms constraining empty structure to a description in this footnote. Readers interested in the details may consult the references here. Successive units of empty structure cannot be left unfilled due to the (universal) phonological ECP condition (Kaye, Lowenstamm & Vergnaud 1990). Empty onsets parametrically count as empty categories. In some languages their emptiness is marked (requiring filling or licensing), and in others they are essentially transparent. Hawu is a language where onsets are relevant to the ECP (Charette 2003; Pagliano 2003; Ulfsbjorninn 2014; Faust 2015; Faust, Lampitelli & Ulfsbjorninn 2018; Ulfsbjorninn 2020; Ulfsbjorninn forthcoming). Without going into the details here, the outcome of these conditions would be that forms such as (25, 26 and 27a) could not simultaneously fill C_2 and V_2. This means that the dependent CV cannot be licensed in **(C)əV shaped words. Consequently, this word shape cannot legitimately meet the “minimal word” condition of the language.

V_1[+high] precedes V_2[–high]: *suka > [hə́k:u] ‘measure’. In practice, metathesis almost exclusively occurs to uCa and iCa sequences. According to the analysis in this paper, the final [a] in Hawu is the phonetic interpretation of an empty nucleus. It is a featureless schwa vowel in final position. Blust (2008) notices a further oddness of the metathesis rule he describes. When ‘a’ is metathesized, it becomes schwa in the stressed position: *suk**a** > [hə́k:u] ‘measure’. This finds a ready explanation in the present account because both stressed schwa and final [a] are phonologically featureless. Since both schwa and final [a] are contextually determined phonetic interpretations of the same featureless vowel, the [a]-schwa alternation is actually expected. As we demonstrate in (28), this account confirms and even simplifies Lysvik’s (2015) analysis of metathesis.[14]

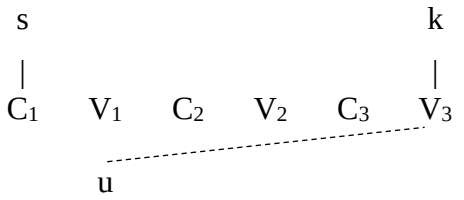
(28) Pseudo-metathesis with my schwa analysis (based on Lysvik 2015)[15]

a. *suka

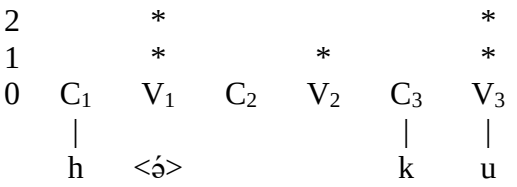


b. Metathesis

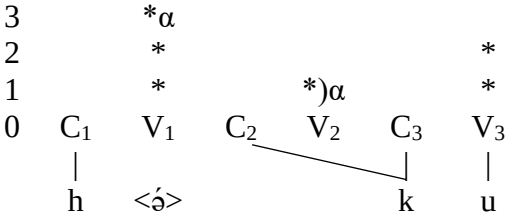
i. Step 1



ii. Outcome[16]



d. Metrical bolstering and the modern form [hə́k:u] “measure”



[14] The mechanism behind metathesis in Hawu, a diachronic process, is not at issue in this paper. Readers are encouraged to read Lysvik (2015). However, our analysis does lend support to Lysvik (2015) and other accounts that show metathesis operating via feature spreading rather than position swapping. Here the precise mechanism is not at issue.

[15] I present the consonants and vowels on separate tiers as Lysvik (2015) does for convenience.

[16] “s to h” is a process that appears very recent, the etymological *s is preserved for instance in the name of the island name, Savu, because at the point of Hawu’s earliest documentation (Wijngaarden 1896), the language still had [s].

The allophonic schwa analysis that is proposed in Section 3.1.1, and the metrical condition on stressed positions (Section 4.2) have conspired to prevent (C)ua or (C)ia from metathesizing into modern Hawu. The outcome of these forms should be **(C)ə́V, but this form fails to meet the quantity requirement of a stressed position and there is no way to repair it through spreading (as shown in Section 5.1).

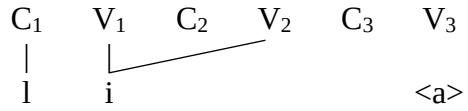
(29) Metathesis is blocked by an empty onset (Blust 2008: 71)

a.	*buaq	(> bu.a)	>	[bu.e]	**[bə́.u]	**[bá.u]	**[bu(:)]	‘fruit’
b.	*liaŋ	(> lia)	>	[li.e]	**[lə́.i]	**[lá.i]	**[li(:)]	‘cave’

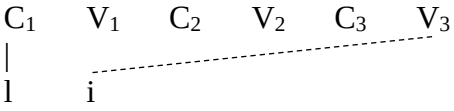
The following diagrams sketched in (30) show that a metathesis derivation cannot operate on (C)ua or (C)ia stems. This is because the output of the process finds no way to be licensed.

(30) *liaŋ (> *lia) > [li.e] **[lə́.i], **[lá.i], **[li(:)] ‘cave’

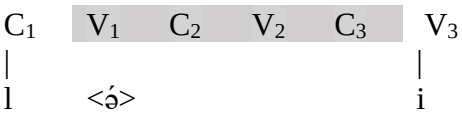
a. *lia



b. Metathesis



c. **Outcome



The derivation reaches an impasse with structure in (30c). There are multiple empty positions to license or fill through spreading: the dependent CV (C_2V_2) and the empty onset C_3. Schwa is featureless; it cannot spread. Likewise, C_3 is empty, so this position is also unable to spread so as to fill dependent CV_2. There is simply nothing local to spread to fill all the empty positions.[17]

6. Conclusion

This paper started by showing that Hawu appears to present rare (if not unique) distributions of strength and weakness at the phonetic level. These rare conditions were summarized as follows: (a) schwa, a universally weak vowel, can only occupy the stressed

[17] Most, if not all(?), of the final a-schwa of Cua and Cia sequences was shifted to ‘e’. This could be analysed as ‘e’ or ‘i’-insertion. If that is the case it would seem that there is a diachronic filling of this final position with actual phonological material in lieu of metathesis. This change might suggest the diagnosis of the underlying cause of metathesis, though more research is required to confirm this is the case and establish the cause.

position. It is excluded from prosodically weak positions; (b) schwa triggers gemination of a right-adjacent consonant. Diachronically this allowed C to resist lenition. Synchronically it appears that the position after a schwa is a strong position, while the position after all other vowels is weak: [lə́**k:**u] ‘fold’ vs. [ní**ʔ**i] ‘fruit bat’; (c) there are no (C)ə́V sequences and metathesis is blocked in (C)$V_{[+high]}$.$V_{[-high]}$ forms. Metathesized final ‘a’ becomes schwa in stressed position.

The paper explained that these facts were all connected to a central organising fact of Hawu phonology that had so far gone unnoticed: Hawu is a quantity-sensitive language. The vowel/ə phonotactics of Hawu were shown to be strictly analogous to those of Tuscan Italian and its metrical lengthening/bolstering and lenition. Hawu was shown also to have metrical bolstering. The reason why Hawu’s quantity-sensitivity went unnoticed is because, in vowels, phonetic duration (length) is not the phonetic correlate of phonological bipositionality. Instead, consistent with virtual length of the Afro-Asiatic type, bipositionality/bimoraicity is correlated to full vowel quality. The quantity requirement of stressed positions is shown to be satisfied by either (a) full phonetic quality or (b) a schwa followed by a geminate. We then showed that this diagnosis rules out words of the shape *(C)ə́V and, simultaneously, it explains the lack of metathesis in *(C)ua or *(C)ia word-shapes.

Language specific phonetic mappings can often obscure the structural similarities in the phonology of languages. Though they look rather dissimilar, Hawu is phonologically similar, and in fact, typologically of the same type, as Tuscan Italian, Norwegian, Icelandic and other quantity-sensitive languages with metrical lengthening/bolstering. This conclusion has important implications for phonological theory; because, under the standard analysis, Hawu appears to have a unique distribution of strength and weakness, with its unique condition that schwa can be exclusively found in a metrically strong position. In fact, under this reanalysis, Hawu’s phonological system is typologically common, what is very rare is its language specific phonetic mappings.

References

Abdul-Karim, Kamal. 1980. *Aspects of the phonology of Lebanese Arabic*. University of Illinois at Urbana-Champaign PhD dissertation.

Archangeli, Diana. 1984. *Underspecification in Yawelmani phonology and morphology*. MIT PhD dissertation.

Awed, Silin A, Lillian Underwood & Vivian Mildred Van Wynen. 2004. *Tboli-English dictionary*. Manila: SIL Philippines.

Barnes, Jonathan. 2003. The phonetic roots of phonological typology: Final syllable vowels. MIT Speech Communication Group, ms. https://www.bu.edu/linguistics/UG/barnes/Final%20Vs%20MIT%20Speech%20Group%20.pdf.

Bauer, Laurie & Paul Warren. 2004. New Zealand English: Phonology. In Bernd Kortmann & Edgar W. Schneider (eds.), *A Handbook of Varieties of English*, vol. 1: Phonology, 580–602. De Gruyter.

Ben Si Saïd, Samir. 2011. The interaction between structure and melody: The case of Kabyle nouns. In Kamila Debowska-Kozowska & Katarzyna Dziubalska-Koaczyk (eds.), *On words and sounds: A selection of papers from the 40th PLM, 2009.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Blust, Robert. 2008. Is there a Bima-Sumba subgroup? *Oceanic Linguistics*, 47(1). 45–113.
Blust, Robert. 2012. Hawu vowel metathesis. *Oceanic Linguistics* 51(1). 207–233.
Bolozky, Shmuel. 2005. The role of casual speech in evaluating naturalness of phonological processes: The phonetic reality of the schwa in Israeli Hebrew. *SKASE Journal of Theoretical Linguistics* 2(3). 1–13.
Bucci, Jonathan. 2013. *Raddoppiamento Fonosintattico induit par l'accent et réduction vocalique en Italie: Perspectives phonologique et dialectologique*. Université Nice Sophia Antipolis PhD dissertation.
Charette, Monik. 1991. *Conditions on Phonological Government*. Cambridge University Press.
Charette, Monik. 2003. Empty and pseudo-empty categories. In Stefan Ploch (ed.), *Living on the edge: 28 papers in honour of Jonathan Kaye*, 465–480. De Gruyter Mouton.
Charette, Monik & A. Göksel. 1998. Licensing constraints and vowel harmony in Turkic languages. In Eugeniusz Cyran (ed.), *Structure and interpretation: Studies in phonology*, 65–89. Lublin: Folium.
Cohn, Abigail C. 1989. Stress in Indonesian and bracketing paradoxes. *Natural Language & Linguistic Theory* 7(2). 167–216.
Crosswhite, Katherine. 2001. *Vowel reduction in Optimality Theory*. New York: Routledge.
Crowley, Terry. 1998. *An Erromangan (Sye) grammar*. Honolulu: University of Hawaii Press.
DeLacy, Paul. 2002. *The formal expression of markedness*. UMass Amherst PhD dissertation.
Faust, Noam. 2015. Eroded prefixes, gemination and guttural effects in Tigre: An account in CVCV phonology. *Natural Language & Linguistic Theory* 33(4). 1209–1234.
Faust, Noam, Nicola Lampitelli & Shanti Ulfsbjorninn. 2018. Articles of Italian unite! Italian definite articles without allomorphy. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 63(3). 359–385.
Faust, Noam & Francesc Torres-Tamarit. 2017. Stress and final /n/ deletion in Catalan: Combining Strict CV and OT. *Glossa: A journal of general linguistics* 2(1).
Faust, Noam & Shanti Ulfsbjorninn. 2018. Arabic stress in strict CV, with no moras, no syllables, no feet and no extrametricality. *The Linguistic Review* 35(4). 561–600.
Flemming, Edward. 2007. The phonetics of schwa. MIT, ms. http://web.mit.edu/flemming/www/paper/schwaphonetics.pdf.
Grimes, Charles E. 2006. Hawu and Dhao in eastern Indonesia: Revisiting their relationship. In *Tenth International Conference on Austronesian Linguistics, 17-20 January 2006*. Palawan, Philippines: Linguistic Society of the Philippines and SIL International. https://www.sil.org/resources/archives/6393 (6 August, 2021).
Guerssel, M. 1978. A condition on assimilation rules. *Linguistic Analysis* 4(3). 225–254.
Gussmann, Edmund. 2002. *Phonology: Analysis and theory*. Cambridge: Cambridge University Press.
Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23(3). 387–429.

Harris, John. 1997. Licensing inheritance: An integrated theory of neutralisation. *Phonology* 14(3). 315–370.
Heimisdottir, Linda Osp. 2013. The phonetics of Bengali consonant clusters. *Cornell Working Papers in Phonetics and Phonology* 1–13.
Ito, Junko & R. Armin Mester. 1995. Japanese phonology. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 817–838. Cambridge, MA: Blackwell.
Kavitskaya, Darya & Peter Staroverov. 2008. Opacity in Tundra Nenets. In Natasha Abner & Jason Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, 274–282. Somerville, MA: Cascadilla Proceedings Project.
Kaye, Jonathan. 1990. 'Coda' licensing. *Phonology* 7(1). 301–330.
Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7(2). 193–231.
Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Cambridge, MA: Blackwell.
Kenstowicz, Michael & Charles Pyle. 1973. On the phonological integrity of geminate clusters. In Michael J. Kenstowicz & Charles W. Kisseberth (eds.), *Issues in phonological theory*, 27–43. De Gruyter.
Kinkade, M. Dale. 1998. How much does a schwa weigh? In Ewa Czaykowska-Higgins & M. Dale Kinkade (eds.), *Salish languages and linguistics*. Berlin: De Gruyter Mouton.
Kitto, Catherine. 1997. *Epenthesis and Polynesian loanwords*. University of Auckland MA thesis.
Koller, Bernhard. 2008. An approach to Tocharian schwa-zero alternations in Government Phonology. University of Vienna, ms.
Kroeger, Paul. 1992. Vowel harmony systems in three Sabahan languages. In Peter W. Martin (ed.), *Borneo Research Council Proceedings Series*, 279–296. Kota Kinabalu: Borneo Research Council.
Larsen, Bergeton Uffe. 1998. Vowel length, "raddoppiamento sintattico" and the selection of the definite article in Italian. In Patrick Sauzet (ed.), *Langues et grammaire II-III, phonologie*, 87–102. Departement SDL (Universite de Paris 8).
Lombardi, Linda. 2002. Markedness and the typology of epenthetic vowels. University of Maryland, College Park, ms. http://roa.rutgers.edu/files/578-0203/578-0203-LOMBARDI-0-1.PDF.
Lowenstamm, Jean. 1991. Vocalic length and centralization in two branches of Semitic (Ethiopic and Arabic). In Alan S. Kaye (ed.), *Semitic studies in honor of Wolf Leslau on the occasion of his eighty-fifth birthday*, 949–965. Wiesbaden: Harrassowitz.
Lowenstamm, Jean. 2011. The phonological pattern of phi-features in the perfective paradigm of Moroccan Arabic. *Brill's Journal of Afroasiatic Languages and Linguistics* 3(1). 140–201.
Lysvik, Julian K. 2015. Vowel compensation and its role in metathesis. *UCL Working papers in Linguistics* 27. 1–9.
Mahboob, Ahmar & Nadra Huma Ahmar. 2004. Fiji English: Phonology. In Bernd Kortmann & Edgar W. Schneider (eds.), *A Handbook of Varieties of English*, vol. Volume 1: Phonology, 1003–1016. De Gruyter.
Marotta, Giovanna. 2008. Lenition in Tuscan Italian (Gorgia Toscana). In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition*, 235–272. Berlin, New York: Mouton de Gruyter.

Oostendorp, Marc van. 2000. *Phonological projection: A theory of feature content and prosodic structure*. Berlin: De Gruyter Mouton.
Pagliano, Claudine. 2003. *L'épenthèse consonantique en français. Ce que la syntaxe, la sémantique et la morphologie peuvent faire à la phonologie*. University of Nice PhD dissertation.
Pandey, Pramod Kumar. 1990. Hindi schwa deletion. *Lingua* 82(4). 277–311.
Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2.
Ras, J.J. 1982. *Inleiding tot het modern Javaans [Introduction to modern Javanese]*. Den Haag: M. Nijhoff.
Repetti, Lori. 2012. Consonant-final loanwords and epenthetic vowels in Italian. *Catalan Journal of Linguistics* 11. 167–188.
Sadanandan, Suchitra. 1999. *Malayalam phonology: An Optimality Theoretic approach*. University of Southern California PhD dissertation.
Salminen, Tapani. 1993. On identifying basic vowel distinctions in Tundra Nenets. *Finnisch-Ugrische Forschungen* 51. 177–187.
Scheer, Tobias. 2004. *A lateral theory of phonology*. Vol. 1: What is CVCV, and why should it be? Berlin: Mouton de Gruyter.
Scheer, Tobias. 2014. Spell-out, post-phonological. In Eugeniusz Cyran, & Jolanta Szpyra-Kozłowska (eds.), *Crossing phonetics-phonology lines*, 255–275. Newcastle upon Tyne: Cambridge Scholars Publishing.
Scheer, Tobias & Péter Szigetvári. 2005. Unified representations for stress and the syllable. *Phonology* 22(1). 37–75.
Sneddon, James N. 1975. *Tondano phonology and grammar* (Pacific Linguistics Series B). Department of Linguistics, Research School of Pacific Studies, The Australian National University.
Tent, Jan & France Mugler. 2004. Fiji English: Phonology. In Bernd Kortmann & Edgar W. Schneider (eds.), *A Handbook of Varieties of English*, vol. 1: Phonology, 750–779. De Gruyter.
Topintzi, Nina. 2010. *Onsets: Suprasegmental and prosodic behaviour*. Cambridge: Cambridge University Press.
Uffmann, Christian. 2007. *Vowel epenthesis in loanword adaptation*. Tübingen: Max Niemeyer Verlag.
Ulfsbjorninn, Shanti. 2014. *A field theory of stress: The role of empty nuclei in stress systems*. SOAS University of London PhD dissertation.
Ulfsbjorninn, Shanti. 2017. Bogus clusters and lenition in Tuscan Italian: Implications for the theory of sonority. In Geoff Lindsey & Andrew Nevins (eds.), *Language Faculty and Beyond*, vol. 14, 278–296. Amsterdam: John Benjamins Publishing Company.
Ulfsbjorninn, Shanti. 2020. Segment–zero alternations in Galician definite article allomorphy: Floating consonants at the left-edge of morphemes. *Acta Linguistica Academica* 67(1). 155–170.
Ulfsbjorninn, Shanti. forthcoming. A phonological reanalysis of morphological segment deletion and de-affrication in Ik. *The Linguistic Review*.
Walker, Alan T. 1982. *A Grammar of Sawu* (NUSA: Linguistic Studies in Indonesian and Languages in Indonesia). Badan Penyelenggara Seri Nusa, Universitas Atma Jaya.

Wijngaarden, J.K. 1896. *Sawuneesche woordenlijst*. The Hague: Nijhoff.
Yip, Moira. 1993. Cantonese loanword phonology and Optimality Theory. *Journal of East Asian Linguistics* 2(3). 261–291.

SOAS Working Papers in Linguistics. Volume 20 (2021): 26–48

Preliminary orthographic design for Ramari Dongosaro

Vasiliki Vita 675802@soas.ac.uk

Chelsea Pedro pedro35@hawaii.edu

Abstract
This paper aims at providing a detailed account of a standardisation project currently underway for Ramari Dongosaro, or Sonsorolese (ISO 639-3: sov), an endangered language spoken by less than 400 speakers (Eberhard, Simons & Fennig 2021) in the Republic of Palau. The purpose of this paper is to function as a record of the project, providing a preliminary phonological analysis, along with recommendations for an alphabet for Sonsorolese and potential applications of it. Finally, with this paper, we aim to gain input and feedback from Micronesian languages specialists and linguists specialising in standardisation.

Keywords: Micronesia, Sonsorolese, community-based standardisation, phonology, phonetics, orthography

1. Introduction

1.1. Background and context

In the middle of the west Pacific, neighbouring Indonesia, Philippines and the Federated States of Micronesia, is Palau, an independent nation-state. Its linguistic landscape is an example of diglossia, with Palauan being used in the local, everyday life and English for administrative and official issues (Matsumoto & Britain 2000: 10). Both languages have official status, with Palauan being the only national language (Matsumoto & Britain 2000: 22). This multilingual context is also home to Ramari Dongosaro, or Sonsorolese (ISO 639-3: sov), and Ramari Hatohobei, or Tobian (ISO 639-3: tox), some of the languages of the southwestern islands of the Republic.[1]

Ramari Dongosaro, or Sonsorolese (ISO 639-3: sov), is the language of Sonsorol, which is the main island of the State of Sonsorol. It belongs to the Chuukic, Micronesian group of the Austronesian family, and is part of a dialectal continuum spoken in the southwestern islands of the Republic of Palau: Sonsorol, Pulo Ana, Merir and Tobi (Grant 2017: 852). In the past, the Southwest islands presented high numbers of population (van den Berg 2014: 3). However, nowadays, the majority of islanders have migrated to the island of Koror for various reasons, such as economic, health, educational and environmental (typhoons). In the village of Echang, a mixture of Southwest islanders live and flourish, speaking Palauan and English leading to the emergence of Echangese, a mixture of Tobian, Sonsorolese, English and Palauan (Black & Black 2013; Taborosi 2018; Vita 2020).

[1] We would like to acknowledge Justin Andrew, Laura I. Miles, Frank Pedro, Lucy Pedro, Felicia Andrew, Lahaina L. Pedro, Peter W. Black and Barbara W. Black for their help, insights and recommendations in preparing the proposal that was submitted to the local authorities and their continuous work throughout this project. We would also like to thank and acknowledge our collaborators Thafaas Men's Organization, Dini Faruya Women's Association and Youth - Sonsorol State Youth.

According to Grant (2017: 853), Tobian and Sonsorolese are two of the most archaic languages spoken in the West Micronesian sprachbund, primarily on the phonological level, more closely related to nuclear Micronesian languages rather than other Chuukic languages, preserving the word-final voiceless vowels on stems. Their position in the family is still unclear, since descriptions of the languages vary from Capell's (1969) grammar, who argues they are more closely related to Ulithian (Capell 1969: 1), and van den Berg's (2014) linguistic sketch, both discussing both languages, to Vita's (2020) analysis of Tobian prosody and Grant's (2017: 853) claim that they are more closely related to Woleaian based on comparisons with other languages of the area. Regardless, there is an abundance of raw data on Kaipuleohone[2] (van den Berg 2013) and in the Endangered Languages Archive collection *Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language* (Black & Black 2014), especially for Tobian.

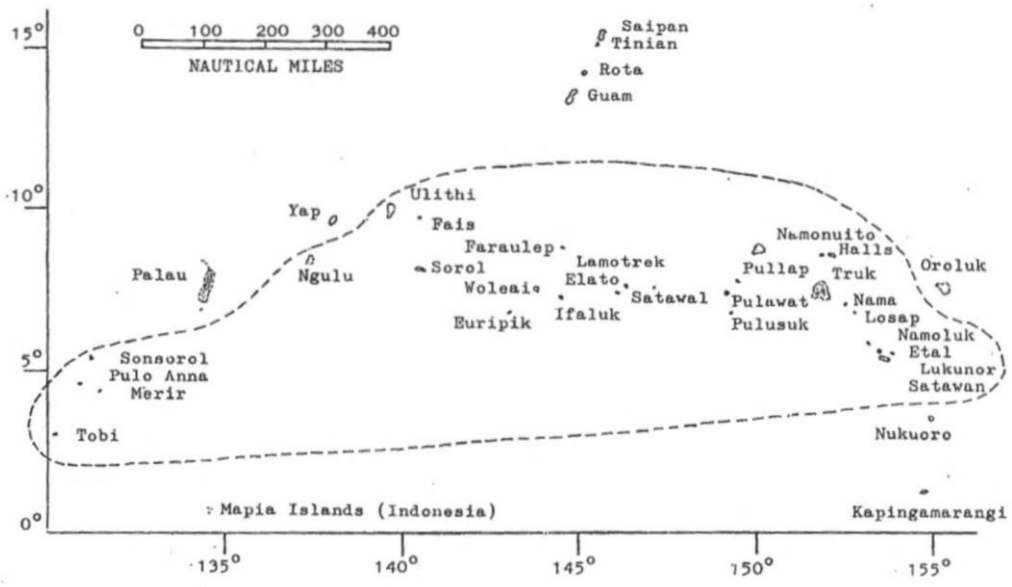


Figure 1. Boundaries of the Chuukic continuum (Quackenbush 1968)

1.2. Standardisation and language maintenance in Micronesia

A standard language has been considered as an ideal notion (Rehg 2004; Romaine 2008; Jones & Mooney 2017), yet its effects in a society are clearly visible, from the creation of resources such as grammar and dictionaries, to the implementation of policies in administration and education. For this reason, the standardisation process is usually initiated by political, social, cultural or religious motivations (Moseley 2017: 36–37) and although standardisation may concern distinct languages, it seems that ideologies, practices, discourses and beliefs travel across boundaries and borders (Milroy & Milroy 1999; Romaine 2008). Furthermore, in order for a standardisation project to be successful and for literacy in the oral language to increase, identifying domains that are not only traditional but which impact everyday life and associate the language to them are required (Casquite & Young 2017; Jones & Mooney 2017).

[2] Kaipuleohone is the digital language archive of the University of Hawai'i. http://ling.hawaii.edu/kaipuleohone-language-archive/

Standardisation and language maintenance efforts in Micronesia originated in the 1970s when a group of linguists undertook a project of documenting and describing Micronesian languages, the *Pacific Languages Development Project* (or PALI) (Rehg 2004: 499). Its goals included documentation of the languages of Micronesia, provision of training for local educators and promotion of literacy in the local language. Although the first two goals were mostly achieved, the last one of promoting literacy in the vernacular language was not, because of complex reasons, that is, inaction on the part of Micronesian educators and failure on the side of linguists to deal with previous orthographies and consult speakers about their desires and recommendations (Rehg 2004: 501–502). After an assessment of the project, Rehg (2004: 506–510) provides the following recommendations for linguists interested in undertaking literacy projects in Micronesia:

- Make sure that the phonology of the language is clear.
- Diacritics may be difficult to be accepted by Micronesians since they are used to writing in English which contains no diacritics.
- When including new graphemes, choose the ones that are familiar and user-friendly.
- Build on existing practices.
- Combine underspecification with digraphs so as to eliminate the use of diacritics.
- When working with different varieties, it is important to be politically acceptable and propose solutions that are efficient for both readers and writers.
- Before deciding, test a preliminary version of it with the general public.
- It is important to listen to the community and realise the effect and impact our work has on its structure.

Although underspecification is undesirable (Hinton 2014: 144), Regh (2004: 508) seems to be accepting it in this case considering English's influence on Micronesian politics and settling with a combination of digraphs in order to avoid diacritics. In general, considering the PALI team's experience regarding the orthographies of various Micronesian languages, it is suggested that one should not focus on a standardised spelling system but rather encourage people to produce written materials that align with the community's needs and aspirations.

The Sonsorolese community is trying to promote the use of the language in official announcements and has expressed an interest in language work such as "writing a dictionary, revive or archive for future use, keep cultures alive especially for *faifire* ('women') and work on translating the Bible and all gospels" (personal communication, SPTS meeting with Thafaas Men's Organization, Dini Faruya Women's Association and Youth - Sonsorol State Youth on 28 June 2021). We hope that this project can further encourage such initiatives and make the current work easier.

2. The project

The origins of this project stem from the relationship Vasiliki Vita developed with the former governor of the State of Sonsorol, Laura Ierago, and the curators of the ELAR collection *Documenting Ramari Hatohobei, or Tobian, a severely endangered Micronesian language* (2014), Peter and Barbara Black. It was through Laura Ierago that Vasiliki Vita got in touch with Lahaina Pedro, who is part of the administrative personnel of the State of Sonsorol office and a member of the Young Historians of Sonsorol, the

main collaborators in this project and an association of Sonsorolese youth aiming at preserving the local culture and practices. Lahaina Pedro was then the one who introduced the two writers, leading to the submission of an application for an ELDP grant in 2019. However, due to COVID-19 the grant cycle was cancelled along with the project but after a few months, the writers got together and thought up how they could help the Young Historians in their work. Through Chelsea Pedro's connections, primarily her family, and Justin Andrew, Vasiliki Vita's consultant for her SOAS MA dissertation project, they got together and prepared a proposal that was later submitted to the local authorities, the State of Sonsorol and Hatohobei offices.

2.1. General structure

The standardisation project presented in this paper is based on surveys. This is mainly because of the fact that the two communities, Hatohobei and Sonsorol, are already writing their languages, although there is great variation. In this paper we will deal with the Sonsorolese side which is moving at a different pace than the Hatohobei side, primarily due to bureaucratic specifications. The aim is to reach a common ground regarding how the language should be written. Another reason for using surveys is the fact that the Linguistic Consultants (Vasiliki Vita and Chelsea Pedro, henceforth LcCs) are significantly distant. However, Chelsea Pedro undertook an eight-week summer 2021 internship with the Sonsorol State office leading to the organisation of workshops and outreach regarding the project.

The organisation of a Language Committee (henceforth LC) was encouraged in order to handle language matters and work in the community. When final decisions are made, these surveys will be shared with the authorities for an official registration of the alphabet in the states' constitutions and other legal documents. Hence subsequent translation of official documents at the state level of Sonsorol will be facilitated.

In the survey, participating speakers will be asked questions concerning their understanding of the structure of the project, their language practices and ideology. They will then be provided with words and a text selected by the LcCs based on the existing linguistic descriptions of the two languages to listen to (Capell 1969; van den Berg 2014). Google Forms will be used for preparing the surveys due to its capability for collaborative editing. However, due to the fact that not everybody in Palau has access to a computer and airtime is expensive, printed surveys are proposed. That is, a Group of Representatives (Sonsorolese Group of Representatives, henceforth GR) will distribute the surveys to members of the community while playing the recordings on their phone.

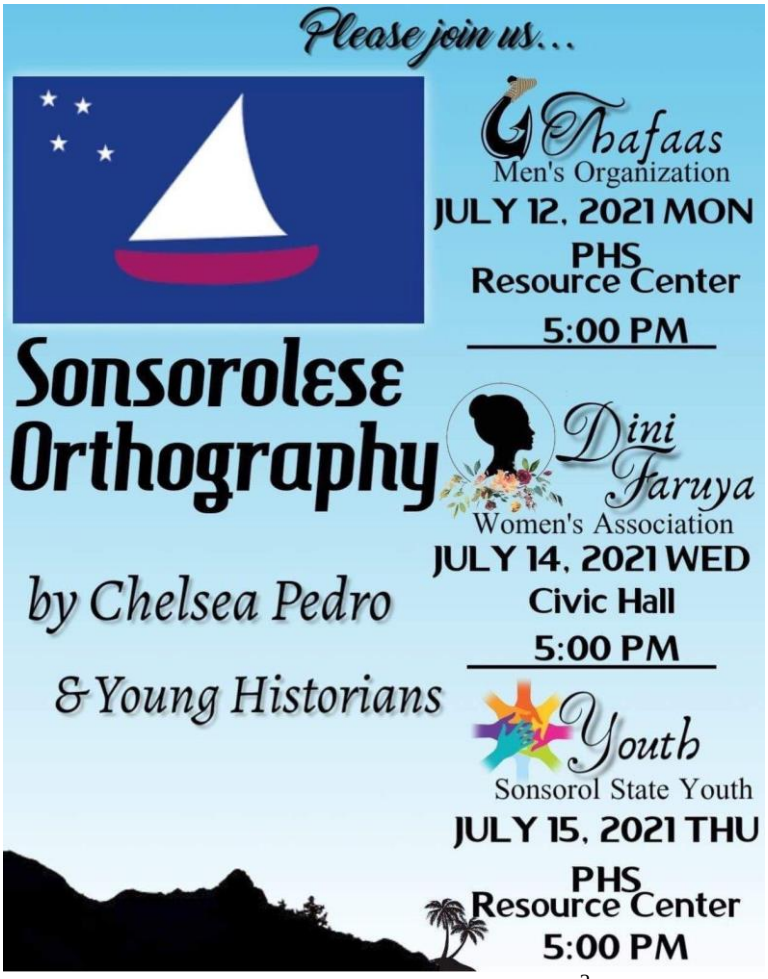


Figure 2. Survey distribution events organised in July 2021[3]

Participating speakers will then be prompted to write the words and text they are listening to based on their personal preference and intuition. Finally, the LcCs and GRs will analyse the results and propose a suggested writing system to the LC which will make further suggestions and changes and announce the final product at another cycle of events, similar to the ones in Figure 2. Before distributing, it is necessary to identify the demographics of the participating speakers. After discussing with the collaborators,[4] we have decided to make these surveys anonymous. For purposes of data protection but also involving as many participants as possible, it was decided to ask all community members above 18 to participate.

[3]Thank you to Laree Ierago for preparing this outreach poster.

[4]This project is organised in collaboration with volunteers from the Young Historians of Sonsorol (https://www.facebook.com/younghistoriansofsonsorolstate/), Thafaas Men's Organization, Dini Faruya Women's Association, the Friends of Tobi (http://www.friendsoftobi.org) and the Hatohobei State Youth Organization. Some of these volunteers will participate as members of the various groups, for example, volunteers from the Young Historians will participate as representatives, while others have assisted with communicating the project to local leaders or consulting them as regards to events on the ground, such as identifying participants, budget and other practicalities. Vasiliki Vita, Chelsea Pedro and the current collaborators have extended an invitation to local leaders to participate in this effort. The collaborators are in charge of selecting the members of the Language Committee and of the Group of Representatives.

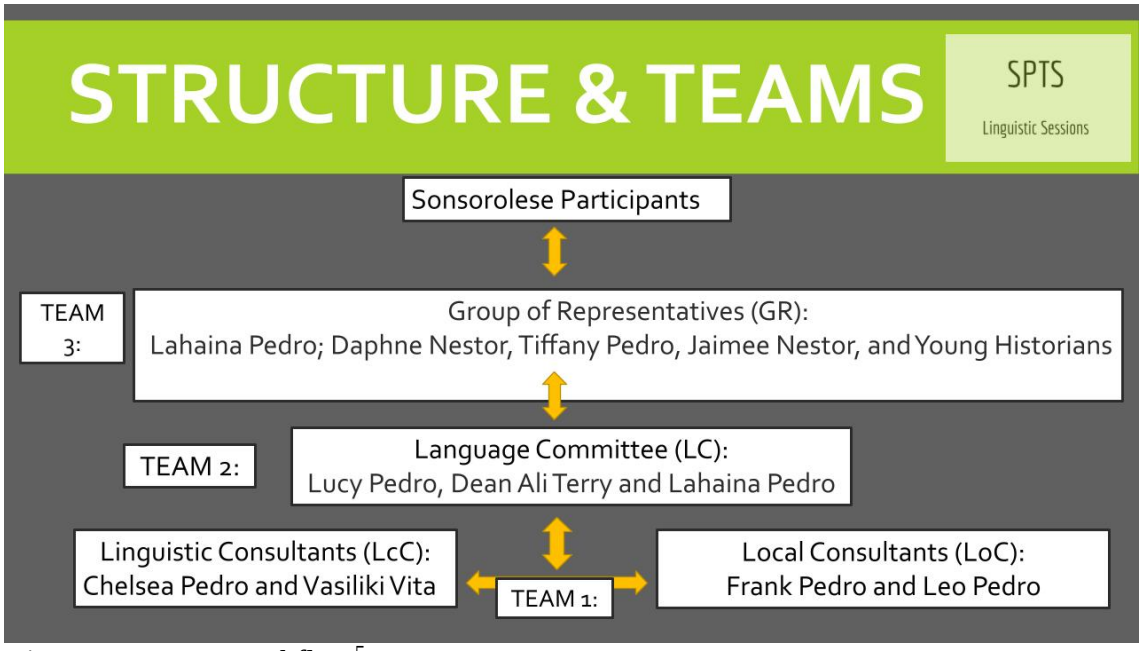


Figure 3. Project workflow[5]

2.2. Detailed steps and responsibilities

2.2.1. Linguistic and Local Consultants

Vasiliki Vita and Chelsea Pedro will function as consultants (Linguistic Consultants, LcCs) throughout the project. This means that they will work with two individuals, namely the Local Consultants (henceforth LoCs), to prepare the surveys. These individuals will be provided with a consent form explaining the purposes of the project, their tasks and how their data will be handled. An LoC could[6] be an individual that is recognised as a fluent speaker of the language and is trusted by the community.

The responsibilities of the LcCs include various steps. Before distributing, they are asked to work with the LoCs to create the surveys, pilot them and review the feedback by making any necessary changes. After distributing, they will have to analyse the results, propose solutions and discuss them with the LoCs and the LC. The responsibilities of LoCs include the provision of recordings of the vocabulary or text required for the preparation of the survey, assistance with demonstrating the differences between similar sounds, provision of possible expected writings of the vocabulary contained in the survey and evaluation of the survey drafts.

2.2.2. The Language Committee

After discussions with the collaborators, the LC will be composed of three speakers of Sonsorolese. The aim is for this committee to continue working on making decisions related to the language after the end of the particular project. The recommended individuals joining this committee are as follows:

[5] Thank you to Lincy Lee Marino for preparing this diagram as part of the Linguistic Sessions outreach event by the Young Historians of Sonsorol in July 2021.

[6] We would rather say ‘could’ than ‘should’ here because we have also consulted a speaker who is not traditionally fluent but whose insights have been accepted by the rest of the collaborators. By including a young speaker who mixes the languages of their repertoire, we hope to further engage young people in language work and promote an atmosphere of acceptance.

- A local leader to function as a bridge between this volunteer initiative and the state
- An individual who is recognised as a fluent speaker of the language and/or an education specialist and/or a teacher
- A young individual, aged 20-40, who identifies as a speaker of the language

The purpose of including a young individual is to encourage them to participate in decision making but also learning and experiencing how language decisions in the future are/can be made (see also footnote 4).

The responsibilities of the LC include various steps. Before distributing the survey, they are asked to function as the piloting group for the surveys and provide feedback and recommendations. They are also asked to distribute the final draft of the surveys to the GRs and assist them with any issues that might arise. Finally, after the analysis of the results by the LcC and GRs, they will be asked to review the findings, provide feedback on the recommendations of the LcC and GRs, announce the results, collect participant feedback from the GRs and work together with LcC and GRs to finalise the orthography.

2.2.3. The Group of Representatives

The creation of a Sonsorolese Group of Representatives (GRs) aims at assisting the Language Committee with distributing the surveys to the participating speakers. This group will be composed of three individuals. Since these individuals will oversee distributing the surveys, it is recommended for young people who identify as speakers of the language to join this group.

The responsibilities of the GRs include various steps. Before distributing, they are asked to provide feedback and recommendations to the LC and prepare a promotion strategy to inform participants (see Figure 2). During distributing the final draft of the surveys to the participants, they are asked to handle participant interaction and questions/concerns that might arise. After distributing, they collect the answers and report participant feedback and assist with the analysis of results. After the results have been announced by the LC, they collect participant feedback and report to the LC.

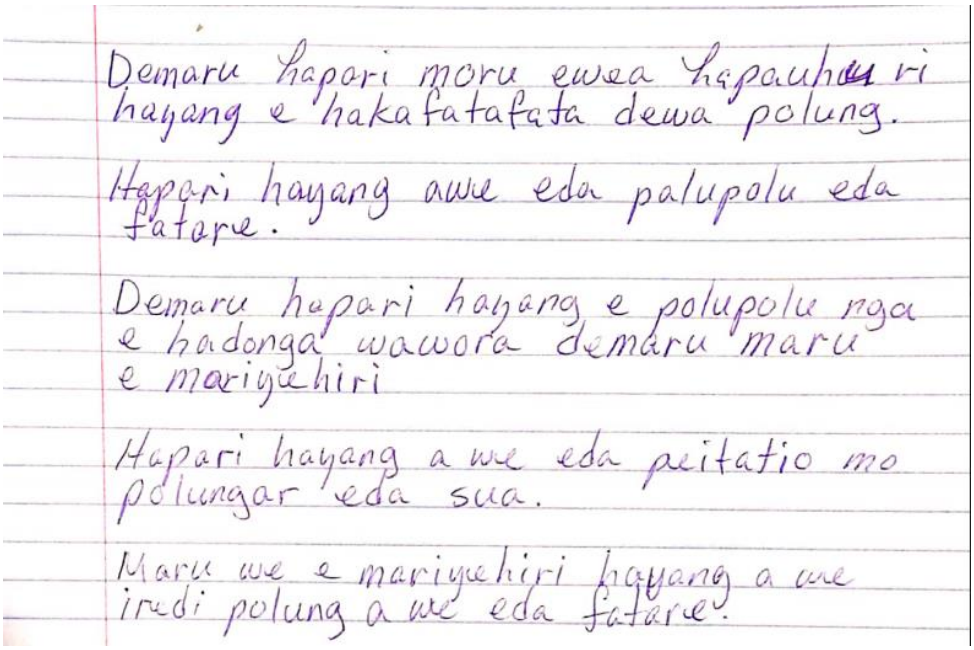
2.3. The survey

The survey[7] aims at gaining input from speakers on how they wish to write, or already write, Sonsorolese. The survey is divided into three parts. In the first part, speakers are asked about their language practices, how they would use the writing system and who they think is the appropriate individuals/organisations/other to decide on a standard writing system. The reason for including this part is to discover how speakers think of their language and how they could potentially use the writing system. If the responses in this first part of the survey do not showcase an interest in furthering work on the language and extend the social contexts in which it could be used, we as linguistic consultants will not push this project forward.

[7] Find a draft of the survey, here: https://forms.gle/oXuwW82EASH5VYVUA

In the second part of the survey, speakers are asked to listen to particular words. Some of these words were selected based on the existing linguistic analyses of the language (Capell 1969; van den Berg 2014). Others were selected based on questionnaires created by consulting Frank Pedro, the main LoC, and recommendations for other languages of the area, such as Woleaian and Saipan Carolinian (Sohn & Tawerilmang 1976; Sohn 1984; Jackson 1984). The reason for including this part is to identify how the majority of speakers would spell each sound of the language and take that into account when making final decisions. We wish to make these results public at the end of the project so that speakers are aware of how the majority has chosen to spell each sound and the reasons behind our recommendations.

In the third part of the survey, speakers are asked to listen to an audio recording in van den Berg's (2013) collection of Rayme Ierago telling the bird story[8] and write what they are listening to. Although this will make the survey longer, we believe that this is an interesting story that will encourage participants to consider ways of spelling the various sounds and words of Sonsorolese. The main reason for including this task is to be able to compare participants and identify again how the majority spells, particularly, subject markers and other proclitics. Furthermore, the fact that the majority of the Sonsorolese-speaking population is literate makes this an ideal introductory task for future literacy projects.


Demaru hapari moru ewea hapauhau ri hayang e hakafatafata dewa polung.

Hapari hayang awe eda palupolu eda fatare.

Demaru hapari hayang e polupolu nga e hadonga wawora demaru maru e mariyehiri

Hapari hayang a we eda peitatio mo polungar eda sua.

Maru we e mariyehiri hayang a we irudi polung a we eda fatare.

Figure 4. Frank Pedro's writing of Rayme's bird story

In February 2021, we held four sessions in total with Frank Pedro via Zoom. The purpose of these sessions was to prepare the survey, test our hypotheses, choose the appropriate vocabulary to be included in the survey and gain some initial input on how speakers of Sonsorolese think about their language. Zoom was used because of the possibility of recording the sessions,[9] of sharing one's screen and of having multiple participants on

[8] Find the recording here: https://scholarspace.manoa.hawaii.edu/handle/10125/32511

[9] Find all sessions:

screen at the same time. Sessions were conducted using a questionnaire,[10] while Frank Pedro has signed a consent form[11] regarding revealing his name in all documentation of the project, making the recordings available and using his insights and productions for the purposes of preparing these surveys.

Currently the survey is in online format, that is, Google Forms. The Sonsorol Language Committee (LC) functioned as the focus group, testing the survey which was finalized after the pilot group identified potential problems and made recommendations. Once the survey was finalized, the online format was used for expatriate speakers of the Sonsorolese community, and the survey was transformed into a printed version for speakers who live in Palau. In the online format, the audio files are embedded, while in the printed version the GRs played the audio files of each word for part two and the story file for part three during the events (see Figure 2). A list of potential participants had been prepared ahead of time.

2.4. Proposed analysis

After collecting the surveys, the findings will be analysed based on the sound system of the language. The words, phrases and texts included in the survey aim at identifying specific sounds and phonological rules. Since both Vasiliki Vita and Chelsea Pedro are recent graduates and, as mentioned before, the phonological descriptions of the languages include Capell's (1969) grammar and van den Berg's (2014) linguistic sketch, the data we gathered are based on such earlier analyses of the language, we are either confirming or debunking their analyses using their wordlists (primarily van den Berg 2014) or the questionnaire (see footnote 8) we have prepared using examples found in Oda's (1977) phonology chapter of Pulo Annian, a related variety, and Sohn (1984).

Google Form's auto-generated spreadsheet with the survey results will be used for the analysis of the findings. In effect, under each word or sentence that speakers are asked to transcribe, the investigated sound (vowel, consonant or diphthong), will be inserted in a column under the word and then the LcCs and selected members of the GRs will assist in counting the answers. Community members and volunteers were presented with the sounds of Sonsorolese, from a linguistic point of view, during the events (see Figure 2). That is, the LcCs prepared a presentation of the linguistic analysis that is also presented in this paper, accommodating linguistic terminology and examples (such as phonology and phonetics, orthography, scripts and alphabets, the IPA, its symbols and uses, single and geminate consonants, single and long vowels, diphthongs, diacritics and digraphs) to ones that are appropriate for the audience.

RStudio (R Core Team 2013) will be used for the final presentation of the results. RStudio is a programming language used for statistics and statistical data visualisation. The benefits are that it will allow the graphing of the survey results making it more

https://drive.google.com/drive/folders/13tcrmKQsn38cooGSTYS4agR82SQFzYUi?usp=sharing

[10] The questionnaire is a combination of examples used in Oda's (1977) phonology chapter and Sohn's papers in Bender's (1984) Studies in Micronesian Linguistics. Find the questionnaire below: https://drive.google.com/file/d/1gK2qRxMtsi-odi-c0s62Js20USOlPOfI/view?usp=sharing

[11] Find the original consent form here: https://docs.google.com/document/d/1vVRqLrj2L6WJcOwtbMMUw5nlzLhbWop3AGkDVMB2fJg/edit?usp=sharing

understandable for our collaborators and allowing its use in official statistical documents of the state. Some preliminary conditions (to be revised according to LC and GRs' recommendations as we move forward):

- If the majority of participants agree on a specific letter for a sound, then that letter will be used.
- If it is 50-50 between two letters for the same sound, then the LcCs and the LC will make recommendations for the particular sound.
- If the participants decide on a symbol or letter that the LcCs would not recommend, then this will be discussed in a community meeting to reach a consensus.

3. The sounds of Sonsorolese

As mentioned earlier, the analyses and the vocabulary used for them are based on earlier work done by van den Berg (2014) and Capell (1969). In this analysis, we have merely described the sounds in more detail and confirmed or debunked previous claims.

3.1. The vowels

As far as vowels are concerned, we agree with van den Berg (2014: 14) that there are seven main vowels in Sonsorolese. Most of these vowels also occur as long vowels.

Table 1: Sonsorolese main vowels

IPA	word (translation)
i	/i:tĕ/ (name)
u	/bʸu:ŋŭ/ (flower)
ʉ	/ŋʉ:ŋŭ/ (chew)
ɛ	/m:ɛ:tɐ/ (what)
ə	/xəɣə-/ (tie up)
o	/xosou/ (rain cloud)
ɐ	/xɐmʸɐsŭ/ (grab)

The first vowel is /i/, which is pronounced with the tongue in the front of the mouth, as in the Sonsorolese word for 'name', /i:tɐ/ and English 'see'.[12] The next vowel is /u/, which is pronounced with the tongue in the back of the mouth, as in the Sonsorolese word for 'flower', /bʸu:ŋŭ/ and General American English 'fruit', while the third one is a vowel that is pronounced with the tongue in the centre of the mouth /ʉ/, as in the Sonsorolese word for 'chew', /ŋʉ:ŋŭ/. The fourth main vowel is /ɛ/, which is pronounced with the tongue in the near front of the mouth and only one of the lips open, that is, the upper lip is covering the teeth, as in the Sonsorolese word for 'what', /mɛtɐ/ and English 'head'. The next sound is /o/, as in the Sonsorolese word for 'rain cloud', /xosou/, which is pronounced with the tongue in the back of the mouth and the lips in a near neutral position.

The schwa /ə/ is the next phoneme, which is pronounced with the tongue in the centre of the mouth and the lips in mid positions, meaning that they are not fully open or closed,

[12] Find all audio files here:
https://drive.google.com/drive/folders/1mk4EcZ4cRB7j68KvM23t0csCVOcZSo0N?usp=sharing

rather in a neutral position, as in the Sonsorolese word for ‘tie-up, fasten’, /xəɣə-/ and the first vowel in the English word ‘again’. Van den Berg (2014: 15–16) argues that /ə/ may not in fact be phonemic but rather an allophone of another phoneme because native speakers seem unaware of the particularity of this sound in stressed positions (its spelling varies between ‘a’ and ‘o’) and because vowels are regularly reduced in running speech. However, we would argue that /ə/ is phonemic because even in careful speech speakers pronounce the /ə/ in /xəɣə-/ (tie up), the same example van den Berg (2014: 16) used. In contrast, a word used by older speakers when addressing children in Sonsorolese seems to be /xɐɣɐ/, while /ɣˡɛɣˡɛ-/ means ‘to saw, to cut something’ and /xoɣo/ means ‘to use something’.

Finally, the vowel sound /ɐ/ is pronounced with the tongue in the centre of the mouth and the lips near open, as in the Sonsorolese word for ‘grab’, /xɐmʸɐsŭ/ and Australian English for ‘calm’. There is variation as far as /ɐ/ is concerned. As van den Berg (2014: 16) notes, in many cases it is pronounced in the centre of the mouth, as in ‘palm’, especially when in open syllables and monosyllabic words, such as the Sonsorolese word for ‘eye’, /ma:t/, while when a low vowel is followed by either /i/ or /e/, it occurs as /æ/. Evidence for this explanation of /æ/ comes from Woleaian, for which a rule exists which states that /ɐ/ is raised before /i/, /e/ and /a/ (Sohn & Tawerilmang 1976: 18), yet we were not able to confirm it.

As far as the short final vowels are concerned, we have identified four (Table 2). As van den Berg (2014: 17) notes /ə/ does not appear in word-final position.

Table 2: Short final vowels

IPA	word (translation)
ĭ	/piɾisĭ/ (dog)
ŭ	/ŋu:ŋŭ/ (chew)
ŏ	/ɛ-mɐoxŏ/ (it-good)
ĕ	/i:tĕ/ (name)

Finally, seven diphthongs have been identified (Table 3). Similar diphthongs have been identified in van den Berg (2014: 17), apart from /ɛo/, /ɐo/ and /ɐʉ/, with the last being described as a VCV-sequence ‘ɐwŭ’. However, we argue here that /ɐʉ/ is a diphthong, as this sequence is not evident in /fɐʉwᵉ/ (four). Furthermore, /ɐo/ is also described in Capell (1969: 8).

Table 3. Diphthongs

IPA	Description	word (translation)
ɛi	a combination of /ɛ/ and /i/	/fɛitɐ/, (to do/how)
ou	a combination of /o/ and /u/	/bʸuwou/, (to go out)
ɛo	a combination of /ɛ/ and /o/	/ðɛow/, (one)
ɐɛ	a combination of /ɐ/ and /ɛ/	/jɐwɐɛɣˡ/, (their mouths)
ɐʉ	a combination of /ɐ/ and /ʉ/	/fɐʉwĕ/ (four)
ɐo	a combination of /ɐ/ and /o/	/e-ða-p:ɐo-:/ (it-TAM-pound-OBJ)

3.2. The consonants

Sixteen single consonants and six geminates have been identified for Sonsorolese, which correspond to a great degree to the ones identified by van den Berg (2014: 18).

Table 4. Consonants

IPA	Description	word (translation)
p	voiceless bilabial plosive	/piɾisĭ/ (dog)
pʸ	voiceless velarised bilabial plosive	/xɐpʸesi/ (heat up)
bʸ	voiced velarised, bilabial plosive	/bʸuːŋŭ/ (flower)
m	voiced bilabial nasal	/maːɾŭ/ (animal)
mʸ	voiced velarised, bilabial nasal	/mʸɐːɾə/ (man)
ŋ	voiced velar nasal	/ŋʉːŋŭ/ (to chew)
f	voiceless labio-dental fricative	/fɐuwĕ/ (four)
x	voiceless velar fricative	/xuːbʸe/ (leg)
ɾ	voiced alveolar tap	/ʉːɾŭ/ (to drink)
w	voiced labial-velar approximant	/wɐʉtŭ/ (to hit)
j	voiced palatal fricative	/jɐːwĕ/ (mouth)
s	voiceless dento-alveolar fricative	/siːmʸ/ (head)
t	voiceless dento-alveolar plosive	/taːŋi/ (to cry)
ð	voiced dental fricative	/ðuːxi/ (to open)
k	voiceless velar plosive	/kɐkːɐnɛ/ (to see)
ɣˡ	voiced lateral-fricativised velar	/jɐfɐɣˡ/ (shoulder)

Oda (1977: 18) mentions that bilabials can be velarised and rounded in Pulo Annian and this is how van den Berg (2014: 19) also described them, thus the same occurs with /bʸ/ in /bʸuːŋŭ/, ‘flower’. Yet, this has not been confirmed for the rest of the bilabial consonants. As far as /ɣˡ/ is concerned, van den Berg (2014: 19–20) mentions that it functions as the correspondent /x/ in Tobian. Although this may be true for certain examples, it could be argued that both sounds exist in Sonsorolese and are used independently. Although we have described this consonant as voiced, it is worth noting that because of the fricativisation, at certain points, it comes out as devoiced. As far as an alveolar nasal, /n/, is concerned, van den Berg (2014: 22) does not describe it as phonemic, and we would agree with this analysis by arguing that /n/ emerges after tap deletion (see Section 4). Nevertheless, this is still preliminary, and there are certain examples, such as /nɐwɛɾ/, ‘no’, inhibiting us from making any conclusive remarks.

Table 5. Geminate consonants

IPA	word (translation)	IPA	word (translation)
m:	/mːɛːtɐ/ (what)	m	/mɐːɣŭ/ (animal)
s:	/sːɐ/ (blood)	s	/siːmʸ/ (head)
ŋ:	/ŋːɐðĕ/ (to breathe)	ŋ	/ŋʉːŋŭ/ (to chew)
k:	/kːɛŋi/ (sharp)	k	/kɐkːɐnɛ/ (to see)
f:	/fːɛxi/ (offering)	f	/fitɛɣ/ (to work)
t:	/tːɐɾĕ/ (to dream)	t	/taːŋi/ (to cry)

Regarding geminate consonants, and particularly the geminate /p:/, it seems that it appears under certain conditions. For example, in /pao/, ‘to pound’, there is a single /p/,

as well as in /pi:jɞ̆/, ‘sand’. Yet, in /eðap:ao:/, ‘he then pounded it’, there is a geminate /p:/. It should be noted that it is strange for the word /pao/ to be pronounced on its own without any proclitics or object markers, thus further investigation is needed. As far as the rest of the geminate consonants are concerned, they appear most often in word-initial position and precede the near front /ɛ/ and the central /ɐ/ (Table 5). In comparison, vowels which occur in word-initial position are usually lengthened when pronounced on their own (van den Berg 2014; Vita 2020), thus something similar might be occurring in the case of certain consonants as well. For this reason, the next steps would be to investigate their occurrence mid-speech and words in which they precede different vowels than the near front /ɛ/ and the central /ɐ/.

Van den Berg (2014: 20) also notes that geminate forms are the results of morphological processes. For example, when a verb begins with /x/ and needs to be reduplicated, it might result in a sequence of /x/ + V + /x/, where the vowel disappears leading to the emergence of [k]. However, we were not able to confirm this.

4. Phonological rules

As far as the phonological rules of Sonsorolese are concerned, it seems that the language follows the patterns of other languages of the area, in particular, Pulo Annian and Woleaian. As mentioned before, our analysis is based on a questionnaire prepared based on Oda’s (1977) analysis of the syntax of Pulo Annian, as well as Sohn’s (1984) *An orthographic design for Woleaian,* and Sohn & Tawerilmang’s (1976) *Woleaian-English Dictionary*.

4.1. High Front vowel lowering

The first vowel of the first person singular object marker suffix /-ai/ (van den Berg 2014: 37) becomes [ɛ] after a stem-final /ɐ/ (1). In (2), the central rounded /ʉ̆/ is elongated because of the addition of the suffix, which resurfaces the devoiced vowel (see Section 4.6), assimilating into a high back rounded vowel /u/.[13]

(1)	a.	/fɐðɞ̆/	b.	/xɐ-fɐðɞ̆-ai/	→	[xɐfɐðɐjɛi]
		live		TR-birth-1SG.OBJ		‘to give birth to me’
(2)	a.	/wautʉ̆/	b.	/wautu-ai/	→	[wautu]
		hit		hit-1SG.OBJ		‘to hit me’

Van den Berg (2014: 32) also notes certain stem-dependent rules too, such as, final /ɐ/ becoming /ɛ/ before /-i/, as in (3), final /ɐ/ becoming /o/ before /-mʷu/, as in (4) and final /ʉ/ becoming /u/ before /-mʸu/, as in (5).

(3)	a.	/ruma-i/	b.	[rumɛi]
		drink-1SG.POSS		‘my drink’
(4)	a.	/i:mʸɞ̆-m/	b.	[i:mʸomʸᵘ]
		house-2SG.POSS		‘your house’

[13] Abbreviations in all examples follow the Leipzig Glossing Rules.

(5)	a.	/ɾɐŭ-m/	b.	[ɾɐumʸŭ]
		child-2SG.POSS		‘your child’

4.2. Glide Epenthesis

The palatal glide /j/ is inserted between identical unrounded vowels (7, 8), and between an unrounded vowel and any vowel (6). In (6), both glide epenthesis and high front vowel lowering are present. Examples (9) and (10) with rounded vowels show when this rule does not occur.

(6)	a.	/fɐðĕ/	b.	/xɐ-fɐðĕ-ai/	→	[xɐfɐðɐjɛi]
		live		TR-live-1SG.OBJ		‘to give birth to me’
(7)	a.	/fitɛɣĭ/	b.	/fitɛɣĭ-(i)/[14]	→	[fitɛji:][15]
		build		build-3SG.OBJ		‘to build it’
(8)	a.	/wɛxitɛxĭ/	b.	/wɛxitɛxĭ-(i)/	→	[wɛxitij]
		turn		turn-3SG.OBJ		‘to turn it’

(9)	a.	/ŋu:ŋŭ/	b.	/ŋu:ŋŭ-(i)/	→	[ŋu:tu̟]
		chew (without swallowing)		chew-3SG.OBJ		‘to chew it’
(10)	a.	/u̟:ɾŭ/	b.	/u̟:ɾŭ-(i)/	→	[u̟ɾu̟mi]
		drink		drink-3SG.OBJ		‘to drink it’

The labio-velar /w/ is inserted between non identical vowels, one of which is rounded, as in example (11), provided the second vowel is not high, as in example (12) where a glide is not inserted.

(11)	a.	/tou/	b.	/tou-(i)/	→	[towu]
		poke		poke-3SG.OBJ		‘to poke it’
(12)	a.	/ɣˡoŋoɣˡoŋŏ/	b.	/ɣˡoŋoɣˡoŋŏ-i/	→	[ɣˡoŋoɣˡoŋo]
		hear		hear-3SG.OBJ		‘to hear it’

4.3 Diphthongization before suffixes

When the third person plural possessive suffix is preceded by the high front vowel /i/ or the near-low central vowel /ɐ/, it turns into a diphthong (13, 14, 15). Van den Berg (2014: 32) describes it as an insertion of either /i/ or /ɛ/. However, in the cases we have observed so far it has been manifested as /ɛ/. Examples (16) to (18) show that with the central rounded /ŭ/ in final position in the stem, the same diphtongization does not occur.

(13)	a.	/jɐ:wĕ/	b.	/jɐ:wĕ-ɣˡɐ/	→	[jɐwɐɛɣˡ]
		mouth		mouth-3PL.POSS		‘their mouth’

[14] Van den Berg (2014: 37) describes the 3rd person singular object marker suffix as -ja, -wa or null but we think that it is -i or null. Since we are still uncertain, we present it in parenthesis.
[15] This is how the speaker pronounced it during the elicitation session. It is not usually elongated.

(14)	a.	/jɐɾɐɾixɐɾ/ tongue	b.	/jɐɾɐɾixɐɾ-ɣˡɐ/ → tongue-3PL.POSS	[jɐɾɐɾixɐɾɐɛɣˡ] ‘their tongue’
(15)	a.	/ðiɾɛðiɾ/ mother	b.	/ðiɾɛðiɾ-ɣˡɐ/ → mother-3PL.POSS	[ðiːɾɐɛɣˡ] ‘their mother’
(16)	a.	/xumɣusŭ/ hand	b.	/xumɣusŭ-ɣˡɐ/ → hand-3PL.POSS	[xumɣusuɣˡ] ‘their hands’
(17)	a.	/taɣˡʉxŭ/ back	b.	/taɣˡʉxŭ-ɣˡɐ/ → back-3PL.POSS	[taɣˡʉxʉɣˡ] ‘their back’
(18)	a.	/fɐðʉxŭ/ head	b.	/fɐðʉxŭ-ɣˡɐ/ → head-3PL.POSS	[fɐðʉxʉɣˡ] ‘their head’

4.4. Tap nasalization

The voiced alveolar flap /ɾ/ becomes an alveolar nasal /n/ when the suffix starts with /ɾ. Thus, this rule applies to a noun ending in /ɾV/ when it is followed by a construct suffix, /-ɾi/ ‘of’ (19), a possessive suffix /-ɾa/ (21), or a demonstrative /ɾa/ (23), but not in examples (20), (22) and (24). The /ɾ/ of the suffix is not retained. Thus, two adjacent taps are pronounced as a long nasal, /ɾɾ/ → [nː]. It could be argued that the nasal here the realization of a geminate rhotic tap.[16] The final vowels /i/ and /a/ are sometimes deleted (see section 4.6).

(19)	/mʸɐɾĭ-ɾi ðoŋosaɾŏ/ man-of Sonsorol	→	[mʸɐnːi ðoŋosaɾŏ] ‘a man from Sonsorol’
(20)	/sɐo-ɾi pɐnəʉ/ person-of Palau	→	[sɐwɛɾ pɐnəʉ] ‘a person from Palau'
(21)	/ðiɾɛðiɾ-ɾa/ mother-3SG.POSS	→	[ðiːnːĕ][17] ‘her mother’
(22)	/uːfĕ-ɾi fɛifiɾ-ɾɐ/ clothes-of girl-DEM	→	[ufɐɾ fɛifinɐ] ‘that girl’s clothes’
(23)	/mʸɐɾĭ-ɾa/ man-DEM	→	[mʸɐnːɐ] ‘that man there’
(24)	/ɾiwɛis-ɾa/ child-DEM	→	[ɾiweisiɾɐ] ‘that kid’

4.5. High vowel rounding/backing before glide

As van den Berg (2014: 31) notes, the construct suffix /-ɾi/ ‘of’ is pronounced [ru] when it appears before /w/, with /ɾ/ once again becoming /n/ as in (25).

[16] It is important to note that we do not yet have any conclusive remarks about the nature of /n/ in general.
[17] The vowel is lengthened because of careful speech.

(25) /mʸɐɾʲ-ɾi woɾɛjɐi/ → [mʸɐn:u woɾɛjɐi]
man-of Woleai ‘man from Woleai’

(26) /mumu-ɾi wɛiɾɛŋ/ → [mumu ɾu wɛiɾɛŋ]
kingdom-of heaven ‘the kingdom of heaven’

4.6. Final vowel devoicing

Van den Berg (2014: 13) notes that possibly Chuukic words ended in vowels. However, with the passage of time, many dialects have shortened long vowels and/or reduced short unstressed vowels, leading to either them becoming voiceless or disappearing completely. We argue here that Sonsorolese is doing both, with final devoicing happening after a rounded vowel and at a phrase boundary. The degree of devoicing seems to vary depending on the degree of stress or carefulness given by the speaker. The so-called ‘devoiced’ vowel may be half-voiced or whispered or is not phonated although the speech organs are set in position for articulation of the vowel. A simple vowel following a consonant or glide is devoiced before a phrase boundary (27). Voiceless word-final vowels are sensitive to the sound environment, if a word follows them closely, they become voiced (28).

(27) ŋ:ɛoɾ ɣɐpɐɾiɛɾʸi wʉoɾ **fɐɣʉɣʉɾ**
ŋ:ɛ-oɾ ɣɐpɐɾiɛɾʸi wʉoɾ **fɐɣʉɣʉ-ɾi**
CONJ-DEM small.bird on.top.of **head-of**
‘and there was a baby bird on it’

(28) **ufɐɾ** fɛifinɐ
u:fɐ̆-ɾi fɛifiɾ-ɾɐ
clothes-of girl-DEM
‘that girl’s clothes’

Although this may seem like a low-level, postlexical rule, final vowels are important to remember because they dictate which vowel to use when adding a suffix. Thus, it would be good practice to spell them out in order for people to remember them. This is one of the main issues with literacy in Sonsorolese and Tobian. Speakers might not remember the final devoiced vowel and/or are not sure how to spell these usually short final vowels when writing the word with a suffix.

4.7. Low vowel raising and fronting

A single low vowel, /ɐ/, is fronted and raised to /ɛ/ between two high unrounded vowels /i/, /ɪ/ (31).

(29) a. /ɟɐfɐɣˡɐ̆/ b. /ɟɐfɐɣˡɐ-ɾi/ → [ɟɐfɐɣˡɐɾi]
shoulder shoulder-of ‘shoulder of’

(30) a. /ɟɐfɐɣˡɐ̆/ b. /ɟɐfɐɣˡɐ̆-ɾ/ → [ɟɐfɐɣˡɐɾ]
shoulder shoulder-3SG.POSS ‘his shoulder’

(31)	a.	/i:mʸɞ̆/ house	b.	/imʸɞ̆-ɾi/ house of	→	[imʸɛɾi] ‘house of’
(32)	a.	/i:mʸɞ̆/ house	b.	/i:mʸɞ̆-ɾ/ house-3SG.OBJ	→	[imʸɐɾ] ‘his house’

4.8. Vowel rounding and/or backing

An unrounded vowel /i/, /ɪ/, /ɛ/, /ɐ/ becomes rounded when followed by a rounded vowel (33, 34, 35). With the resurfacing (see Section 4.6) of the underlying vowel /ɞ̆/ because of the addition of the suffix /-ɾi/, in (33), the unrounded vowel /ɨ/ becomes /u/. In (35) it is not just rounding that is present, but also monophthongisation with the diphthong /ɛo/ being deleted completely and the back rounded vowel /o/ emerging.

(33)	a.	/ɟɐɾɨðɞ̆/ ghost	b.	/ɟɐɾɨðʉ-ɾi/ ghost-of[18]	→	[ɟɐɾuðʉɾi] ‘ghost of’

(34) /bʸi-wow/ → [bʸuwou]
go-out ‘to go out’

(35) /ðɛow-bʸoŋɨ̆/ → [ðobʸoŋɨ̆]
one-night ‘one night’

Van den Berg (2014: 27) also notes that /ɛ/ becomes /ə/ before /ʉ/ if it is followed by /-uwɐ/ (36).

(36) /ðɛɣˡi-uwɐ/ → [ðəɾu:wɐ]
one-two ‘three’

5. Proposed orthographic design

As Rehg (2004: 510) describes the stages of standardization, it could be argued that Sonsorolese is at stage 3, where the speakers employ writing for a variety of functions but there is no widespread agreement concerning how words should be spelled or what letters to use. For this reason, we are focusing on surveys and tallies to count what the majority is using. Regardless, we have certain recommendations. Although Rehg (2004: 507) does not recommend diacritics, we believe that they would be useful in order to remain accurate and represent the differences between single and long vowels and consonants.

Thus, we recommend using the macron above the letter, as in ā, to signify long vowels and geminate consonants. Since the Bible Translation team is using the umlaut, if it comes up in the survey answers we would be willing to recommend it instead of the macron. Next we recommend using the circumflex above vowels, as in â, to signify central vowels like the schwa, /ʉ/ and devoiced vowels at the end of the word. If a central vowel that already has a circumflex is long then we would recommend using the umlaut to signify the fact that this vowel is not only central but also long, as in /ŋʉ:ŋʰ/ ‘chew’, ngüngû. Subject markers and other proclitics are recommended to be written separately from the

[18] Used only for objects, not people.

main verb phrase, while suffixes are recommended to be conjoined at the end of the word. When writing a dictionary, it is recommended to always include the devoiced vowel at the end of the word.

Table 6. Proposed Alphabet for Sonsorolese

IPA	Orthography	IPA	Orthography
p	p	i	i
pʸ	pw	u	u
bʸ	bw	ʉ	û
m	m	ɛ	e
mʸ	mw	ə	ê
ŋ	ng	o	o
f	f	ɐ	a
x	h	ɨ	î
ɾ	r	ʉ	û
w	w	ɐ	â
j	y	ə	ê
s	s	ɐi	ai
t	t	ɛi	ei
ð	d	oʉ	oû
k	k	ɛo	eo
ɣˡ	ghl	ɐɛ	ae
m:	m̄	ɐu	au
s:	s̄	ɐo	ao
ŋ:	n̄g		
k:	k̄		
f:	f̄		
t:	t̄		

In effect, we propose the use of three diacritics:

- The macron (¯) for long vowels and geminates
- The circumflex (ˆ) for central and final vowels
- The umlaut (¨) for central long vowels

Our reasons for choosing diacritics rather than digraphs are various. First of all, it could be argued that digraphs have not been widely accepted in Micronesia and although people may be using them there are many cases of communities wishing for a revised version of their grammar or dictionary (Taborosi Danko, personal communication, November 26, 2019). In fact, this is reasonable considering how largely spoken languages like French and Greek have committees deciding on the orthography and revisions of it at certain points in time. Consider, for example, the case of double σ for κλασσικός, 'classic', in Greek, which was transformed into κλασικός in 1976 when it was postulated for loan words in Greek to be written in a simpler manner (Saradakos 2018).

Furthermore, diacritics have already been introduced to the communities through the Bible translation teams' work and, from communication with the SIL linguist, they have

been well-received (Paulus Kieviet, personal communication, January 6 2021). In addition, if diacritics are not used speakers will have to deal with an alphabet of more than 30 letters in order to be phonetically accurate. Although Regh (2004: 506) points out that one should be sure about the phonology of the language before diving into a design of orthography for it, this contradicts with choosing underspecification. We posit that it would be better to use diacritics than digraphs for matters of economy. Diacritics would make the letters look closer to what people are already using and make them seem more special, differentiating from Palauan and English. The main goal would be to create an alphabet that is easy and comfortable for both readers and writers, while preserving unique linguistic features.

Moreover, suggesting diacritics could also reveal attitudes and ideologies towards various languages that exist in their immediate environment. Many of the indigenous languages in the Pacific are seeing a renaissance, meaning they are receiving attention and support from state authorities. Those closest to Micronesia are Hawaiian and Māori, both of them using diacritics. Perhaps by choosing diacritics the Sonsorolese might see it as aligning themselves with Hawaii and the Hawaiian renaissance, a common destination for education and immigration, or by not choosing to use diacritics they align with the rest of Micronesian nations. After all, Rehg (2004: 515) mentioned an incident where Micronesians showed admiration toward Waikīkī. Although this admiration was interpreted as being towards English as a key element for societal development, perhaps this admiration could shift towards Hawaiian after Micronesians become introduced to the efforts of the Hawaiian revitalisation movement. Either way, through this survey and our next steps we hope to build on existing practices and make recommendations accordingly.

Finally, it is important to remember that this is a preliminary design that will become final after speakers allow it and after consulting with the Language Committee and Group Representatives. We have no intention of implementing the recommendations presented in this paper without the consent of the speakers, and this is evident by the fact that we introduce diacritics in the third part of the survey asking speakers how they would feel about them. During the workshops about the linguistic aspects of the language, diacritics and their use were also introduced as part of terminology rather than as one of our recommendations. That way, we hope to encourage speakers to consider beforehand on their own the various ways that their language could be written.

6. Applications

The desired outcome of the project described in this paper would be a standardised orthography for the variety of Sonsorol that will be used by community members in education, administration and local life. This desired outcome could then lead to community involvement in the documentation of Sonsorolese and the subsequent use of this material for the development of teaching materials which could be important in showcasing how this writing system could be applied in real-life contexts.

This next phase of the project includes a documentation project with revitalisation in mind (Sugita 2007; Amery 2009; Nathan & Fang 2009). That is, a collection of audio and video materials comprising both linguistic and paralinguistic data of culturally specific practices (such as fishing and canoe building), as well as recordings of naturally occurring

conversations in everyday life, are proposed. Members of the team will be trained in collecting metadata so that they can continue the work of archiving material for posterity with all materials described using the agreed-upon writing system.

Based on the literature (Rehg 2004; Jones & Mooney 2017; Casquite & Young 2017), collaboration between LcC, LC and educators will be required. These educators need to have a high level of fluency in the target language and be able to write in English, Palauan or both. Following Mosel (2011), small theme-based dictionaries using WeSay (2020) are proposed. Its primary distinction from other software is that "it has been specifically designed for relatively unskilled native speakers of under-resourced languages, empowering them to be active contributors and creators of dictionaries" (Albright & Hatton 2007: 192). By combining recordings and the aforementioned dictionaries, the goal would be to create a corpus as a resource to be used for future work as well as an online database available to researchers and members of local and expatriate communities.

Despite the fact that monolingual dictionaries have been suggested by the community, most young speakers are semi-speakers and might be discouraged by a monolingual dictionary. Sonsorolese-English dictionaries are, thus, proposed, since English has high status, with varieties of English emerging among the youth, like Palauan English and Echangese. Naturally occurring language conversations and recordings from elicitation sessions collected for the purposes of creating the multimodal documentation mentioned before could be used for the production of a pedagogical grammar.

Finally, although the transference of ESL practices is not recommended for teaching endangered languages (Hinton 2011; Penfield & Tucker 2011) and the languages of Micronesia in particular (Yunick 2000), the accommodation of such techniques in the Sonsorolese context is proposed for various reasons. First, most Sonsorolese, like Tobian, youths have fluency in English and Palauan due to intermarriage and the structure of the education system in Palau (Tibbetts 2002: 10), so they have already reached a degree of literacy. Furthermore, Palau Community College (PCC 2013) offers an Education Program which familiarises potential teachers with lesson planning, integrating technology into teaching and learning, and other techniques. Additionally, by applying a collaborative model for the creation of material, the issues mentioned in Yunick (2000) can be confronted.

7. Conclusion

In conclusion, this paper deals with a project that aims at the production of a standardised orthography for the language of Sonsorol in the Republic of Palau. Although we are facing various issues and limitations, from physical distance, because of inability to travel due to COVID-19 and lack of resources, to bureaucracy and issues of ownership and authenticity, we are slowly and steadily working on building relationships, using the resources available and informing speakers of the possibilities of using the local language in multiple contexts.

Our next steps include the organisation of seminars and workshops. Topics for these workshops include: first, a description of the project and its specific phases and second, the proposed orthography and its linguistic aspects. We recognise that using diacritics in

nowadays' digitally driven world may present certain obstacles for the community, but we are working with young individuals and hope to make this transition easier during these workshops as well.

In effect, the general aim is to acquire a clearer idea of the desires of the Sonsorolese community, how to adapt our workflow and re-evaluate our methodology and goals. By collaborating with local authorities, closing the gap between the domains of language use, which are clearly represented and divided in the Palauan society, and expanding them to official administrative documentation and education seems faster and more legitimate.

Considering Rehg's (2004) recommendations, it seems that it is the people of Micronesia who are responsible for conserving their linguistic heritage. In this case as well, this project hopes to encourage not only the individuals volunteering, but survey participants as well, to consider their language practices. As linguists, our position lies in supporting such efforts and assisting with any knowledge that is relevant. It is the speakers who are ultimately the ones to persevere the global wave of language change and continue using and speaking their languages.

References

Albright, Eric & John Hatton. 2007. WeSay: A tool for engaging native speakers in dictionary building. In D. Victoria Rau & Margaret Florey (eds.), *Documenting and revitalizing Austronesian languages*, 189–201. Honolulu: University of Hawaii Press.

Amery, Rob. 2009. Phoenix or relic? Documentation of languages with revitalization in mind. *Language Documentation & Conservation* 3(2). 138–148.

Berg, Esther van den. 2013. Hatohobei and Sonsorol. *Kaipuleohone: University of Hawai'i Digital Language Archive*. https://scholarspace.manoa.hawaii.edu/handle/10125/32448 (15 April, 2020).

Berg, Esther van den. 2014. *Linguistic sketch of Hatohobeian and Sonsorolese: A study of phonology and morphology*. Amsterdam: Vrije Universiteit Amsterdam BA thesis.

Black, Peter W. & Barbara W. Black. 2013. Documenting Ramari Hatohobei, a severely endangered Micronesian language. Endangered Languages Documentation Programme grant proposal. http://www.friendsoftobi.org/misc/research/ELDPgrantproposal2013.pdf (13 December, 2020).

Black, Peter W. & Barbara W. Black. 2014. Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language. *Endangered Languages Archive*. http://elar.soas.ac.uk/deposit/0295 (9 September, 2020).

Capell, Arthur. 1969. *Grammar and vocabulary of the language of Sonsorol-Tobi*. Sydney: University of Sydney.

Casquite, Mansueto & Catherine Young. 2017. Hearing local voices, creating local content: Participatory approaches in orthography development for non-dominant languages communities. In Mari C. Jones & Damien Mooney (eds.), *Creating orthographies for endangered languages*. Cambridge: Cambridge University Press.

Eberhard, David M., Gary F. Simons & Charles D. Fennig. 2021. *Ethnologue: Languages of the World*. Twenty-fourth edition. Dallas, Texas: SIL International. http://www.ethnologue.com (8 December, 2020).

Grant, Anthony P. 2017. The Western Micronesian sprachbund. In Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*, 852–877. Cambridge: Cambridge University Press.

Hinton, Leanne. 2011. Language revitalization and language pedagogy: New teaching and learning strategies. *Language and Education* 25(4). 307–318.

Hinton, Leanne. 2014. Orthography wars. In Michael Cahill & Keren Rice (eds.), *Developing orthographies for unwritten languages*, 139–168. SIL International.

Jackson, Frederick H. 1984. Selecting an orthography Saipan Carolinian. In Byron W. Bender (ed.), *Studies in Micronesian linguistics* (Pacific Linguistics: Series C 80), 237–258. Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Jones, Mari C. & Damien Mooney (eds.). 2017. *Creating orthographies for endangered languages*. Cambridge: Cambridge University Press.

Matsumoto, Kazuko & David J. Britain. 2000. Hegemonic diglossia and pickled radish: Symbolic domination and resistance in the trilingual Republic of Palau. *Essex Research Reports in Linguistics* 29. 1–37.

Milroy, James & Lesley Milroy. 1999. *Authority in language: Investigating Standard English*. 3rd edn. London: Routledge.

Mosel, Ulrike. 2011. Lexicography in endangered language communities. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 337–353. 1st edn. Cambridge University Press.

Moseley, Christopher. 2017. Who owns vernacular literacy? Assessing the sustainability of written vernaculars. In Mari C. Jones & Damien Mooney (eds.), *Creating orthographies for endangered languages*, 36–53. Cambridge: Cambridge University Press.

Nathan, David & Meili Fang. 2009. Language documentation and pedagogy for endangered languages: A mutual revitalisation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 5, 132–160.

Oda, Sachiko. 1977. *The syntax of Pulo Annian: A nuclear Micronesian language*. UMI Ann Arbor.

PCC. 2013. Education (ED) Program. *Palau Community College*. https://pcc.palau.edu/education-ed-program/ (5 January, 2021).

Penfield, Susan D. & Benjamin V. Tucker. 2011. From documenting to revitalizing an endangered language: Where do applied linguists fit? *Language and Education* 25(4). 291–305. https://doi.org/10.1080/09500782.2011.577219.

Quackenbush, Edward Miller. 1968. *From Sonsorol to Truk: A dialect chain.* University of Michigan PhD dissertation.

R Core Team. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/ (20 September, 2020).

Rehg, Kenneth L. 2004. Linguists, literacy and the law of unintended consequences. *Oceanic Linguistics* 43(2). 498–518.

Romaine, Suzanne. 2008. Linguistic diversity and language standardization. In Marlis Hellinger & Anne Pauwels (eds.), *Handbook of Language and Communication: Diversity and Change*, 685–714. De Gruyter Mouton.

Saradakos, Nikos. 2018. Τα δίδυμα σίγμα του κλασικού [The twin sigmas of 'klasikos']. https://sarantakos.wordpress.com/2018/12/07/classic/ (14 July, 2021).
Sohn, Ho-min. 1984. An orthographic design for Woleaian. In Byron W. Bender (ed.), *Studies in Micronesian linguistics* (Pacific Linguistics: Series C 80), 215–236. Department of Linguistics, Research School of Pacific Studies, The Australian National University.
Sohn, Ho-min & Anthony F. Tawerilmang. 1976. *Woleaian-English dictionary*. Honolulu: The University Press of Hawaii.
Sugita, Yuko. 2007. Language revitalization or language fossilization? Some suggestions for language documentation from the viewpoint of interactional linguistics. In Peter Austin, Oliver Bond & David Nathan (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory*, 243–250. London: SOAS University of London.
Taborosi, Danko. 2018. Sonsorolese Language: Ramari Dongosaro. *Sonsorol Island*. http://www.sonsorol.com/language/ (29 December, 2019).
Tibbetts, David T. 2002. *Tobian cultural identity in the Republic of Palau*. Mangilao: University of Guam MA thesis.
Vita, Vasiliki. 2020. *Prosodic patterns in Ramari Hatohobei*. London: SOAS University of London MA thesis.
Yunick, Stanley. Jr. 2000. Linguistics, TESL, and language planning in Micronesia. *Studies in the Linguistic Sciences* 30(1). 183–200.
2020. *WeSay*. Payap Language Software & SIL International. https://software.sil.org/wesay/ (5 December, 2020).

SOAS Working Papers in Linguistics. Volume 20 (2021): 49–71

Tense and aspect comparison between Griko, Modern Greek and Italian for the indicative

Paris Alexandros Zeikos

Pzeikos@hotmail.com

Abstract

This essay investigates how Griko, an Italiot-Greek language, spoken in Italy, forms verbs in comparison to Modern Greek and Italian, by focusing on the present tense, the imperfect, the past tense, the present perfect, the past perfect and lastly, the progressive aspect, as those are the tenses and aspects found in Griko. The primary findings suggest that the verbal morphology of Griko is somewhat cognate to that of Modern Greek, with some influence from Italian, primarily in the present perfect and the progressive aspect, due to Griko being in contact with Italian. The essay concludes by stating that Griko verbal morphology functions the same way as Modern Greek morphology does, due to both languages being typologically similar and due to both languages being derived from Ancient Greek, despite long contact with Italian.

Keywords: Griko, Modern Greek, Italian, morphology, tense, aspect

1. Background information

Griko is an Italiot-Greek language spoken in Southern Italy, in the province of Lecce, in the administrative area of Apulia by the Griko people, whose name is believed to be derived from the Grecians, an Ancient Hellenic tribe which had colonised parts of Southern Italy (Douri & De Santis 2015: 71). The region where Griko is spoken is also known as Grecia Salentina encompassing around 9 villages (Calimera, Castrignano dei Greci, Corigliano d'Otranto, Martano, Martignano, Melpignano, Soleto, Sternatia and Zollino) covering an area of 143.90 km^2 (Douri & De Santis 2015: 71). The total population of Grecia Salentina in 2015 was estimated to be around 41,500 people, with the language being spoken by around 20,500 people, mostly elderly (Douri & De Santis 2015: 71).

Identifying the origins of Griko is troublesome as there are not many historical records or artefacts that can help linguists identify its ancestry. However, there are two theories, one supported by Greek linguists, the Magna Grecia theory which states that Griko is derived from the Doric dialect of Ancient Greek as spoken in the colonies of Magna Grecia, and a second one supported by Italian linguists, the Byzantine' theory which states that Griko is derived from the Hellenistic Koine spoken in the Byzantine era, which was based on the Athenian dialect (Douri & De Santis 2015: 71; Pellegrino 2016: 65).

In the region where Griko is spoken, Italian is also spoken in its regional variety alongside Salentino, an Italo-Romance variety. Salento can be described as *triglossic*, due to the speakers being able to use all three languages, with varying degrees of fluency. Due to prolonged contact between the three languages, some features from Italian and Salentino have been incorporated into Griko, affecting the lexicon, phonology, and morphology of the language to varying degrees.

The examples used in Modern Greek are derived from the author (as they are a native speaker of the language) or from grammar books, while examples in Griko are derived from the grammar book published by Karanastasis (1997) and from various authors who have conducted research in Griko. Examples in Italian are derived from grammar books as well.

2. Verbal morphology

2.1. Morphological marking

Verbs in Griko, like Modern Greek and Italian, are inflected for person, number, tense, aspect, voice and, to some degree, mood. Furthermore, all three languages are known as pro-drop languages whereby the subject is expressed through agreement on the verb, i.e., the ending of the verb indicates whether the subject is in the 1st, 2nd, or 3rd person singular or plural (Holton et al. 2012).

(1) **Modern Greek**

paíz-o bala
play-PRES.1SG football
I play football.'

(2) **Italian**

cant-o una canzone
sing-PRES.1SG a song
'I sing a song.'

(3) **Griko**

avri pam-e totzu
tomorrow go-PRES.1PL field
'Tomorrow we will go to the field.' (Golovko & Panov 2013: 75)

In the above examples, the subject pronoun is not overtly expressed as it is shown via agreement on the verb. All the suffixes in the examples indicate person, number, tense and voice. Both example 1 and example 2 are inflected for the 1st person singular active voice, indicated by the suffix *-o* for the verb *paizo* 'I play' and for the verb *canto* 'I sing', while for example 3 the verb *pame* 'we go' is inflected for the 1st person plural active present tense indicated by the suffix *-e*.

2.2. Verbal stems

2.2.1. Verbal stems in Modern Greek and Griko

Verbs in Modern Greek and Griko are made up of a stem, which carries the meaning of the verb, and an inflectional ending, which indicates the grammatical properties of the verb. Griko and Modern Greek use different verb stems for different tenses, i.e., the imperfective and perfective stem, which are differentiated by aspect semantically and by form of the stem. Imperfective and perfective stems exist for both the active and passive voice (Holton et al. 2012: 236). The verb *gráfo* 'to write' in Griko and the verb *déno* 'to tie' in Greek are used to show how the verb stems remain the same for some tenses. The imperfective stem is indicated by *gráf-* in Griko and *dén-* in Modern Greek while the perfective stem is indicated by *gráfs-* in Griko and *des-* in Modern Greek as shown in the tables below.

2.2.1.1. Imperfective stem

The following examples indicate the imperfective stems, in italics, for both Griko and Modern Greek according to the tenses found in each language.

(4) **Griko** (present)
gráf-o
write-PRES.1SG
'I write.' (Karanastasis 1997: 83)

(5) **Griko** (imperfect)
é*graf*-a
write-IMPF.1SG
'I was writing.' (Karanastasis 1997: 83)

(6) **Modern Greek** (present)
dén-o
tie-PRES.1SG
'I tie.' (Holton et al. 2012: 127)

(7) **Modern Greek** (imperfect)
éden-a
tie-IMPF.1SG
'I was tying.' (Holton et al. 2012: 127)

(8) **Modern Greek** (imperfective future)
tha *dén*-o
FUT tie.PRES.1SG
'I will tie.' (Holton et al. 2012: 128)

2.2.1.2. Perfective stem

The following tables indicate the perfective stems, in italics, for both Griko and Modern Greek according to the tenses found in each language.

(9) **Griko** (aorist)
égrafs-a
write-PST.1SG
'I wrote.' (Karanastasis 1997: 83)

(10) **Griko** (pluperfect)
eíxa *gráfs*-onta
have.PST write.PTCP
'I had written.' (Karanastasis 1997: 83)

(11) **Modern Greek** (simple past)
édes-a
tie.PST.1SG
'I tied.' (Holton et al. 2012: 127)

(12) **Modern Greek** (perfect)
éxo *des*-ei
have.PRES.1SG tie.PTCP
'I have tied.' (Holton et al. 2012: 127)

(13) **Modern Greek** (pluperfect)
eíxa *dés*-ei
have.PST.1SG tie.PTCP
'I had tied.' (Holton et al. 2012: 127)

(14) **Modern Greek** (perfective future)
tha *dés*-o
FUT tie.PRES.1SG
'I will tie.' (Holton et al. 2012: 127)

(15) **Modern Greek** (future perfect)
tha éxo *dés*-ei
FUT have.PRES.1SG tie.PTCP
'I will have tied.' (Holton et al. 2012: 127)

2.2.2. Verbal stems in Italian

Italian verbs are made up of three parts, one consisting of the root of the verb (which can be stressed or unstressed) which expresses the lexical meaning of the verb, a thematic vowel *a,e,* or *i,* which is present in some parts of the verb in the infinitive and helps identify the conjugation a verb belongs to and an inflectional ending, which gives information about the grammatical properties of the verb *cantare* 'to sing' as shown in Table 1.

Table 1: Verbal stem in Italian (Maiden & Robustelli 2007: 219)

Root	Thematic vowel	Suffix
cant-	-a-	-re

2.3. Stress in Modern Greek and Griko

Stress is defined as the relative emphasis or prominence given to a certain syllable in the word. For Modern Greek and Griko, every disyllabic or polysyllabic word must contain one stressed syllable, with the stress falling on the ultimate, penultimate, or antepenultimate syllable. Stress can change depending on the tense, and it is indicated by an accent on the vowel as shown below (Holton et al. 2012). As shown in example (16), for the *imperfect* and the *past simple*, the stress retreats to the antepenultimate syllable. As shown in example (17), for the *present*, stress falls on the last syllable of the stem of the verb.

(16) **Modern Greek** (past simple)
diá.va.s-a
read-PST.1SG
'I read'

(17) **Modern Greek** (present)
dia.vá.z-o
read-PRES.1SG
'I read'

2.4. Augment in Modern Greek and Griko

The augment is a prefix which is used to indicate past time. In Modern Greek and Griko, the augment is usually an *e-* and it is only used to form verbs for the *imperfect* and *past tense* when the stem begins with a consonant (Holton et al. 2012: 152). It is mandatory when the verb consists of a one-syllable stem and a one-syllable ending, with the augment carrying the stress of the verb as shown below.

(18) **Modern Greek**
é-graps-a
AUG-write-PST.1SG
'I wrote.' (Holton et al. 2012: 127)

Table 2: Breakdown of verb *égrapsa* in the past tense to show the augment, stem, and suffix

Augment	Stem	Suffix
é-	graps	-a

Interestingly, verbs in Griko whose original vowel has been lost, e.g., *apalíno* > *paláino* 'to soften', also take the augment *e-* which suggests that Griko is more conservative, as shown in the following examples.

(19) **Griko** (present)
paláino
soften.PRES.1SG
'I soften.' (Karanastasis 1997: 80)

(20) **Griko** (imperfect)
e-pálina
AUG-soften.PST.1SG
'I softened.' (Karanastasis 1997: 80)

2.5. Conjugations

2.5.1. Modern Greek

Modern Greek verbs fall into 2 categories or conjugations, one being the 1st conjugation or *paroxytone* verbs where stress falls on the last syllable of the verb stem in the present tense e.g., *gráf-o* 'to write' and the other being 2nd conjugation verbs or *oxytone* verbs whereby stress falls on the final syllable on the first person singular active present tense, e.g., *agapó* 'to love'. 2nd conjugation verbs can be broken down into two types. Type A 2nd conjugation verbs usually have an alternative 1st person singular in *-ao* characterized by the *-a* vowel in the active present tense, as in *agap-ó* or *agap-áo* 'to love'. Type B 2nd conjugation verbs do not have an alternative 1st person singular in *-ao,* but they have the vowels *-ei* or *-ou* in the personal endings of the active present tense, for example, the 2nd

person singular for the verb *theoro* ‘to consider’ is *theoreis* indicated by the inflectional ending *-ei* (Holton et al. 2012: 132).

Table 3: Conjugations in Modern Greek (Holton et al. 2012: 130)

Conjugation		Ending	Example
1st		-o	gráf-o ‘to write’
2nd	Type A	-ao/-o	agap-áo/agap-ó ‘to love
	Type B	-o	theor-ó ‘to consider’

2.5.2. Italian

Verbs in Italian fall into three different types of conjugations, which are all distinguished by the thematic vowel in the infinitive summarised in the table below.

Table 4: Verbal conjugations in Italian (Maiden & Robustelli 2007: 219)

Conjugation	Thematic vowel	Ending	Example
1st	-a	-re	cant-are ‘to sing’
2nd	-e	-re	perd-ere ‘to lose’
3rd	-i	-re	fin-ire ‘to finish’

2.5.3. Griko

Griko has two main conjugations, which are derived from Ancient Greek or Byzantine Greek as shown in the table below.

Table 5: Conjugations in Griko (Karanastasis 1997: 87)

Conjugation	Example
-eo/-o	gráf-o ‘to write’
-ao	agap-áo ‘to love’

3. Tense and Aspect

3.1. The present tense

The present tense is used to locate a situation or to describe events that are happening at the present time (Velupillai 2012). The present tense in Griko and Modern Greek is known as *Enestotas* (Ενεστώτας), while in Italian it is known as *Presente* (Holton et al. 2012: 120; Proudfoot & Cardo 2005: 31).

3.1.1. Modern Greek

The present tense is formed via the addition of the present tense suffixes onto the imperfective stem of the verb. The present tense suffixes change depending on the conjugation of the verb. (2nd conjugation verbs type A have an alternative suffix indicated in italics.)

Table 6: Conjugations in Modern Greek (Holton et al. 2012: 126)

Person		Conjugation					
		1st	e.g. gráfo 'to write'	2.A	e.g. agapó 'to love	2.B	e.g. theoró 'to consider'
Singular	1st	-o	gráf-o	-ó	agap-ó	-ó	theor-ó
				-áo	**agap-áo**		
	2nd	-eis	gráf-eis	-ás	agap-ás	-eís	theor-eís
	3rd	-ei	gráf-ei	-áei	agap-áei	-eí	theor-eí
				-á	**agap-á**		
Plural	1st	-oume	gráf-oume	-áme	agap-áme	-oúme	theor-oúme
				-oúme	**agap-oúme**		
	2nd	-ete	gráf-ete	**-áte**	**agap-áte**	-eíte	theor-eíte
	3rd	-oun(e)	gráf-oun(e)	-oún(e)	agap-oún(e)	-oún(e)	theor-oún(e)
				-án(e)	**agap-án(e)**		

3.1.2. Italian

The present tense is formed via the addition of the present tense suffixes onto the verb stem. The thematic vowel for the 2nd person plural changes depending on the conjugation of the verb, as shown in Table 7 in italics.

Table 7: Conjugations in Italian (Proudfoot & Cardo 2005: 32)

Person		Conjugation					
		1st *-are*	Cantare 'to sing'	2nd *-ere*	Credere 'to believe'	3rd *-ire*	Dormire 'to sleep'
Singular	1st	-o	cant-o	-o	cred-o	-o	dorm-o
	2nd	-i	cant-i	-i	cred-i	-i	dorm-i
	3rd	-a	cant-a	-e	cred-e	-e	dorm-e
Plural	1st	-iamo	cant-iamo	-iamo	cred-iamo	-iamo	dorm-iamo
	2nd	-ate	cant-*ate*	-ete	cred-*ete*	-ite	dorm-*ite*
	3rd	-ano	cant-ano	-ono	cred-ono	-ono	dorm-ono

3.1.3. Griko

In Griko, the present tense is formed via the addition of the present tense suffixes onto the imperfective stem of the verb, with the suffixes changing depending on the conjugation. For the 2nd person singular, the final *-s* is added when the following word begins with a vowel.

Table 8: Conjugations of Griko (Karanastasis 1997: 83)

Person		Declension			
		-o	gráfo 'to write'	-ao	agapáo 'to love'
Singular	1st	-o	gráf-o	-ó	agap-ó
	2nd	-i (s)	gràf-i(s)	-á(s)	agap-á(s)
	3rd	-i	gràf-i	-a	agap-á
Plural	1st	-ome	gràf-ome	-oúme	agap-oúme
	2nd	-ete	gràf-ete	-áte	agap-áte
	3rd	-oune	gràf-une	-oúne	agap-oúne

3.1.4. Comparison

Griko has retained most of the present tense suffixes from the same source as Modern Greek as seen from Table 9, which compares the present tense suffixes for Griko, Modern Greek and Italian. Table 10 compares the verb *gráfo* 'to write' and *agapáo* 'to love' for both Griko and Modern Greek, as both verbs are found in both languages.

It could be said that Griko follows the 1st conjugation Modern Greek suffixes to form the present tense for the *-o* conjugation, while for the *-ao* conjugation, Griko follows the 2nd conjugation type A suffixes, as the suffixes for both conjugations in Griko and Modern Greek appear to be the same. Interestingly, Griko uses the suffix *-oune* for the 3rd person plural which is only used in informal contexts in Modern Greek. Furthermore, for the 1st person plural, Griko uses the suffix *-ome*, which could be derived from the Ancient Greek suffix *-omen*, while Modern Greek uses the suffix *-oúme*.

Table 9: Present tense suffixes for Griko, Modern Greek and Italian (Karanastasis 1997: 81; Proudfoot & Cardo 2005: 32; Holton et al. 2012: 126)

Language		Griko		M.G			Italian		
Conjugation		-o	-ao	1st	2.A	2.B	1st	2nd	3rd
Singular	1st	-o	-ó	-o	-ó -áo	-ó	-o		
	2nd	-i (s)	-á(s)	-eis	-ás	-eís	-i		
	3rd	-i	-a	-ei	-áei -á	-eí	-e		
Plural	1st	*-ome*	-oúme	*-oume*	-áme -oúme	-oúme	-iamo		
	2nd	-ete	-áte	-ete	-áte	-eíte	-ate	-ete	-ite
	3rd	-oune	-oún(e)	-oun(e)	-oún(e) -áne	-oún(e)	-ano	-ono	

Table 10: Griko and Modern Greek present tense for the verbs *gràfo* 'to write' and *agapáo* 'to love' (Karanastasis 1997; Proudfoot & Cardo 2005; Holton et al. 2012)

Person		Conjugation			
		-o / 1st gráfo 'to write'		-ao / 2nd agapáo 'to love'	
		Griko	M.G	Griko	M.G
Singular	1st	gráf-o	gráf-o	agap-ó	agap-ó agap-áo
	2nd	gráf-i(s)	gráf-eis	agap-á(s)	agap-ás
	3rd	gráf-i	gráf-ei	agap-á	agap-áei agap-á
Plural	1st	gráf-ome	gráf-oume	agap-oúme	agap-áme agap-oúme
	2nd	gráf-ete	gráf-ete	agap-áte	agap-áte
	3rd	gráf-une	gráf-oun(e)	agap-oúne	agap-oún(e) aga-án(e)

In conclusion, Griko present tense suffixes are cognate to those of Modern Greek, as they appear to be the same for both *-o* conjugation verbs and *-ao* conjugation verbs. It has no similarities with Italian.

3.2. The imperfect

The imperfect tense is used to express an action or a state as a continuous event in progress or as an event which is repetitive or habitual in the past (Holton et al. 2012: 121). In Griko and Modern Greek, the imperfect is known as *Paratatikos* (Παρατατικός), while in Italian the imperfect is known as *l'imperfetto*. The imperfect in Griko and Modern Greek is the combination of the imperfective stem of the verb with the past, while in Italian the imperfect suffixes are added to the verb stem, which happens to be the same as the 2nd person plural indicative (Proudfoot & Cardo 2005: 32; Holton et al. 2012: 121). In Griko and Modern Greek, the imperfect presents events described by the verb as incomplete and can have two senses (Holton et al. 2012). When the imperfect is used in some past time, the action described by the verb can be interpreted as ongoing or continuing, as in example 21, with the verb *évlepe* 'see', *éklaige* 'cry' and *miloúse* 'speak' taking the 3rd person singular imperfect suffix *-e*. When used in a repetitive or habitual sense in the past, it presents the event described by the verb as a recurrent pattern, as in example 22, whereby the act of seeing Hellen happened every time 'I' was in Athens, with the frequency being supported by the word *syxná*, with the verb to *évlepa* 'see' taking the 1st person singular imperfect suffix *-a*. Note that, three verbs in the following examples take the syllabic augment *e-* which carries the stress.

(21) **Modern Greek**

évlep-e to paidi pou éklaig-e
see.IMPF.3SG the child which cry-IMPF.3SG
alla den tou miloús-e
but not him speak-IMPF.3SG
'He was watching the child crying but was not speaking to him.'
(Holton et al. 2012: 295)

(22) **Modern Greek**

otan ím-oun stin Athína évlep-a syxná tin Eléni
when be-IMPF.1SG in Athens see-IMPF.1SG often her Hellen
'When I was in Athens, I used to see Helen frequently.' (Holton et al. 2012: 295)

In Italian, the imperfect can be used to describe situations with information or events happening the same time in the past, as in example 23, whereby the act of raining and the car running happened at the same time, with both verbs, *pioveva* 'rain' and *correva* 'run' taking the 3rd person singular imperfect suffix *-va*. The Italian imperfective is also used to relate the background of a situation in which an event happened, as in example 24, whereby the auxiliary verb *habere* 'have' is in the imperfect creating the background which inflected for the 1st person singular imperfect suffix *-vo*.

(23) **Italian**

er-a mezzanotte, piove-va
be-PST.3SG midnight, rain-IMPF.3SG
e la macchina corre-va silenziosa
and the car ran-IMPF.3SG silently
'It was midnight, it was raining, and the car was running silently.'
(Proudfoot & Cardo 2005: 46)

(24) **Italian**

non ho mangiato perché non ave-vo fame
not have.PST.1SG eat.PTCP because not have-IMPF.1SG hungry
'I didn't eat because I wasn't hungry.' (Proudfoot & Cardo 2005: 46)

3.2.1. Modern Greek

The imperfect is formed via the addition of the imperfect suffixes onto the imperfective stem of the verb. Most of the 1st conjugation verbs take an augment *e-* (except the irregular verbs) which carries the stress for the 1st, 2nd and 3rd singular and 3rd person plural when the verb stem begins with a consonant, and the verb stem together with the suffix form two syllables only. For 2nd conjugation verbs, the suffixes are attached to the imperfective stem of the verb which takes the affix *-oús*, which always carries the stress of the verb. 1st conjugation verbs and 2nd conjugation verbs take the same suffixes with the difference being that 2nd conjugation verbs take the suffix *-ous* before the inflectional ending as shown in Table 11.

Table 11: Imperfect suffixes of Modern Greek verbs (Holton et al. 2012: 131)

Person		Conjugation				
		1st		2nd		
		Suffix	e.g., gráfo 'to write'	Suffix	Type A e.g., agapáo 'to love'	Type B e.g., theoró 'to consider'
Singular	1st	-a	é-graf-**a**	-oúsa	agap-**oúsa**	theor-**oúsa**
	2nd	-es	é-graf-**es**	-oúses	agap-**oúses**	theor-**oúses**
	3rd	-e	é-graf-**e**	-oúse	agap-**oúse**	theor-**oúse**
Plural	1st	-ame	gráf-**ame**	-oúsame	agap-**oúsame**	theor-**oúsame**
	2nd	-ate	gráf-**ete**	-oúsate	agap-**oúsate**	theor-**oúsate**
	3rd	-an	é-graf-**an**	-oúsan(e)	agap-**oúsan(e)**	theor-**oúsan(e)**

Interestingly, 2nd conjugation verbs of type A can also take the 1st conjugation imperfect suffixes while 2nd conjugation verbs type B cannot, as they can only take the 2nd conjugation type A imperfect suffixes as shown in Table 12.

Table 12: Modern Greek imperfective suffixes of 2nd conjugation verbs type A and type B (Holton et al. 2012: 131)

Person		2nd conjugation				
		Type A				Type B
		agapáo ‘to love’				theoró ‘to consider’
				Alternative Suffix		
Singular	1st	-oúsa	agap-oúsa	-a	agápag-a	theor-oúsa
	2nd	-oúses	agap-oúses	-es	agápag-es	theor-oúses
	3rd	-oúse	agap-oúse	-e	agápag-e	theor-oúse
Plural	1st	-oúsame	agap-oúsame	-ame	agapág-ame	theor-oúsame
	2nd	-oúsate	agap-oúsate	-ate	agapág-ate	theor-oúsate
	3rd	-oúsan	agap-oúsan(e)	-an(e)	agápag-an(e)	theor-oúsan(e)

3.2.2. Italian

The imperfect is formed via the addition of the imperfect suffixes onto the verb stem, which is identical to the stem of the verb found in the 2nd person plural indicative, for 1st (*-are*), 2nd (*-ere*) and 3rd (*-ire*) conjugation verbs as shown in Table 13. The suffixes remain the same, but the thematic vowels change depending on the conjugation as shown in Table 14.

Table 13: Formation of the imperfect in Italian (Proudfoot & Cardo 2005: 219)

Tense	Conjugation		
	1st	2nd	3rd
Infinitive	parlare ‘to speak’	credere ‘to believe’	finire ‘to finish’
2nd person plural indicative	parlate	crede	finite
Imperfect (1stSG)	parlavo	credevo	finivo

Table 14: Imperfect suffixes in Italian (Proudfoot & Cardo 2005: 219)

Person		Ending	Conjugations		
			1st parlare ‘to speak’	2nd credere ‘to believe’	3rd dormire ‘to sleep’
Singular	1st	-vo	parl-**avo**	cred-**evo**	dorm-**ivo**
	2nd	-vi	parl-**avi**	cred-**evi**	dorm-**ivi**
	3rd	-va	parl-**ava**	cred-**eva**	dorm-**iva**
Plural	1st	-vamo	parl-**avamo**	cred-**evamo**	dorm-**ivamo**
	2nd	-vate	parl-**avate**	cred-**evate**	dorm-**ivate**
	3rd	-vano	parl-**avano**	cred-**evano**	dorm-**ivano**

3.2.3. Griko

The imperfect in Griko is formed via the addition of the imperfect suffixes to the imperfective verb stem, which happens to be the same verb stem as the present (Karanastasis 1997). The table below compares the imperfect for *-o* conjugation verb, e.g., *gráfo* ‘to write’ and *-ao* conjugation verb *agapáo* ‘to love’. Both conjugations take the

same suffixes for all person with the only difference being in the 3rd person singular, whereby *-o* conjugation verbs take the suffix *-e* while *-ao* conjugation verbs take the suffix *-a*. The *-s* for both conjugations is added onto the 2nd person singular only when the following word begins with a vowel.

Table 15: Imperfect suffixes in Griko (Karanastasis 1997: 83)

Person		Suffix	Conjugation	
			-o	-ao
			e.g., gráfo 'to write'	e.g., agapáo 'to love'
Singular	1st	-a	é-graf-**a**	agápon-**a**
	2nd	-e(s)	é-graf-**e(s)**	agápon-**e(s)**
	3rd	-e	é-graf-**e**	agáp-**a**
Plural	1st	-amo	e-gráf-**amo**	agapoú-**amo**
	2nd	-ato	e-gráf-**ato**	agapoú-**ato**
	3rd	-ane	e-gráf-**ane**	agapoú-**ane**

3.2.4. Comparison

The following table compares the imperfect suffixes for Griko, Modern Greek and Italian.

Table 16: Imperfect suffix comparison (Karanastasis 1997; Proudfoot & Cardo 2005; Holton et al. 2012)

			Language				
			Griko	Modern Greek			Italian
Declension			-o / -ao	1st	2nd		1st /2nd / 3rd
						Alternative	
Person	Singular	1st	-a	-a	-oúsa	-a	-vo
		2nd	-e(s)	-es	-oúses	-es	-vi
		3rd	-e/-a	-e	-oúse	-e	-va
	Plural	1st	-amo	-ame	-oúsame	-ame	-vamo
		2nd	-ato	-ate	-oúsate	-ate	-vate
		3rd	-ane	-an(e)	-oúsan(e)	-an(e)	-vano

Based on Table 16, it is clear that the Griko imperfect suffixes are cognate with those of 1st conjugation Modern Greek verbs for both *-o* and *-ao* conjugations, as the suffixes appear the same in both languages for the 1st, 2nd and 3rd singular and 3rd person plural, despite the 3rd person singular for *-ao* conjugation verbs taking the suffix *-a*. Interestingly, the *-ao* conjugation in Griko uses the alternative suffixes used in Modern Greek for the 2nd conjugation type A, which is the same as the suffixes used by the 1st conjugation verbs.

However, for the 1st person plural in Griko, the suffix *-amo* does not appear to be the same as Modern Greek *-ame*, but more similar to Italian *-vamo,* while the 2nd person plural suffix *-ato* is not the same in either Modern Greek *-ate* or Italian *-vate.* The 3rd person plural for Griko and Modern Greek appears to be the same with the only difference being that, in Modern Greek, the suffix *-ane* is used in informal contexts.

The following table compares the verbs *gráfo* ‘to write’ and *agapáo* ‘to love’ for both Griko and Modern Greek, as the verbs are the same in both languages.

Table 17: Comparison of the verb *gráfo* ‘to write’ and *agapáo* ‘to love’ in Griko and Modern Greek (Karanastasis 1997: 83; Holton et al. 2012: 127)

Person		-o / 1st Gráfo ‘to write’		-ao / 2nd agapáo ‘to love’		
		Griko	Modern Greek	Griko	Modern Greek	
					Alternative	
Singular	1st	é-graf-a	é-graf-a	agápon-a	agápag-a	agap-oúsa
	2nd	é-graf-e(s)	é-graf-es	agápon-e(s)	agápag-es	agap-oúses
	3rd	é-graf-e	é-graf-e	agáp-a	agápag-e	agap-oúse
Plural	1st	e-gráf-amo	gráf-ame	agapoú-amo	agapág-ame	agap-oúsame
	2nd	e-gráf-ato	gráf-ete	agapoú-ato	agapág-ate	agap-oúsate
	3rd	e-gráf-ane	é-graf-an	agapoú-ane	agápag-an(e)	agap-oúsan(e)

Following from the above differences between Griko and Modern Greek, both *-o* conjugations and 1st conjugation verbs take the augment *e-*, which carries stress for the 1st, 2nd and 3rd person singular; however, for the 3rd person plural, the stress in Modern Greek is on the augment while for Griko, the stress is on the penultimate syllable. Furthermore, the augment is kept in Griko for the 1st and 2nd person plural, while for Modern Greek it is removed. Regarding the differences between the *-ao* conjugation verbs in Griko and the 2nd conjugation verbs in Modern Greek, the suffix *-ous* which is used to form the imperfect in Modern Greek is not used in Griko, as Griko uses the 1st conjugation suffixes. One similarity between these two declensions is that in both Griko and Modern Greek, the stress falls on the verb stem.

In conclusion, Griko imperfect suffixes are cognate to those of Modern Greek to form the imperfect for *-o* conjugation verbs and *-ao* conjugation verbs as they appear the same in both languages. The only difference between Griko and Modern Greek is the absence of the *-ous* suffix in Griko as Griko uses the 1st conjugation verb suffixes of Modern Greek. Griko has no similarities to Italian for this tense.

3.3. The past simple

The past simple is used to indicate that an action described by the verb took place at a specific time in the past. Griko and Modern Greek do not possess the past simple but the *aorist* (αόριστος), which blends the properties of the perfective, which is seen by the perfective stem of the verb, and the imperfect, which is indicated by the suffixes (Holton et al. 2012). The aorist is used to describe an action or state that took place and finished at some point in the past, as in example 25, whereby the act of writing started at some point in the past and finished, indicated by the past simple 1st person singular suffix *-a*.

(25) **Modern Greek**

é-graps-a	pénte	grámmata	xthés
write-PST.1SG	five	letters	yesterday

‘I wrote five letters yesterday.’ (Holton et al. 2012: 120)

Italian possess the *passato remoto* ‘simple perfect’ which is used to express the distance of past events in terms of remoteness from the present situation as shown in example 26, whereby the disappearance of the dinosaurs happened very far in the past with no relationship to the present time.

(26) **Italian**

i	dinosauri	scomparv-erono	65	milioni	d’anni	fa
the	dinosaurs	disappeared-PST.3PL	65	million	years	ago

‘Dinosaurs disappeared 65 million years ago.’ (Proudfoot & Cardo 2005: 49)

3.3.1. Modern Greek

The simple past in modern Greek is formed via the active perfective stem of the verb followed by the imperfect suffixes. All three conjugations, 1st conjugation verbs and 2nd conjugation types A and B take the same suffixes as shown in Table 18. 1st conjugation verbs whose verbal stem begins with a consonant and forms two syllables with the suffix, take an augment *-e* before the verb stem, which carries stress for the 1st, 2nd and 3rd person singular and 3rd person plural. The aorist is indicated by the presence of an *-s* onto the stem of the verb however, it can change depending on the verb.

Table 18: The aorist for 1st and 2nd conjugation verbs in Modern Greek (Holton et al. 2012: 132)

Person		Suffix	Conjugation		
			1st	2nd	
				Type A	Type B
Singular	1st	-a	égraps-**a**	agápis-**a**	theóris-**a**
	2nd	-es	égraps-**es**	agápis-**es**	theóris-**es**
	3rd	-e	égraps-**e**	agápis-**e**	theóris-**e**
Plural	1st	-ame	gráps-**ame**	agapís-**ame**	theorís-**ame**
	2nd	-ate	gráps-**ate**	agapís-**ate**	theorís-**ate**
	3rd	-an	égraps-**an**	agápis-**an**	theóris-**an(e)**

3.3.2. Italian

The *passato remoto* ‘simple perfect’ is used to express the distance of past event in terms of their separateness from the present situation. The *passato remoto* is formed by the addition of the *passato remoto* suffixes onto the verb stem as shown in Table 19.

Table 19: *Passato remoto* inflectional endings in Italian (Proudfoot & Cardo 2005: 32)

Person		Declension					
		1st	parlare 'to talk'	2nd	credere 'to believe'	3rd	dormire 'to sleep'
Singular	1st	-ai	parl-**ai**	-etti/-ei	cred-**etti**/-**ei**	-ii	dorm-**ii**
	2nd	-asti	parl-**asti**	-esti	cred-**esti**	-isti	dorm-**isti**
	3rd	-ò	parl-**ò**	-ette/-e	cred-**ette/-e**	-ì	dorm-**ì**
Plural	1st	-ammo	parl-**ammo**	-emmo	cred-**emmo**	-immo	dorm-**immo**
	2nd	-aste	parl-**aste**	-este	cred-**este**	-iste	dorm-**iste**
	3rd	-arono	parl-**arono**	-ettero/-erono	cred-**ettero**/-**erono**	-irono	dorm-**irono**

3.3.3. Griko

The aorist in Griko is formed via the active perfective stem of the verb, indicated by the presence of an *-s* on the stem, followed by the imperfect suffixes which are the same for both *-o* conjugation verbs and *-ao* conjugation verbs. Verbs in the 2nd person singular take the final *-s* when the proceeding word starts with a vowel (Karanastasis 1997: 82). For *-o* conjugation and *-ao* conjugation verbs whose verb stem begins with a consonant and the verb stem together with the suffix form 2 syllables, an augment *e-* is added before the verb stem. For *-o* conjugation verbs, the stress for the 1st, 2nd and 3rd person singular falls on the augment while for *-ao* conjugation verbs, the stress falls onto the verb stem. Furthermore, for *-o* conjugation verbs, the stress for the 1st, 2nd and 3rd person plural falls on the verb stem while for *-ao* conjugation verbs, the stress falls onto the suffix (Karanastasis 1997: 82).

Table 20: Aorist in Griko for *-o* conjugation verbs and *-ao* conjugation verbs (Karanastasis 1997: 83)

Person		Suffix	-o	-ao
			grafo 'to write'	agapao 'to love'
Singular	1st	-sa	é-graf-**sa**	e-gápi-**sa**
	2nd	-se(s)	é-graf-**se(s)**	e-gápi-**se(s)**
	3rd	-se	é-graf-**se**	e-gápi-**se**
Plural	1st	-samo	e-gráf-**samo**	e-gapi-**sámo**
	2nd	-sato	e-gráf-**sato**	e-gapi-**sáto**
	3rd	-sane	e-gráf-**sane**	e-gapi-**sáne**

3.3.4. Comparison

Table 21 compares the aorist and the *passato remoto* suffixes for Griko, Modern Greek and Italian.Griko aorist suffixes are cognate to those of Modern Greek for both *-o* and *-ao* conjugation verbs, as seen in Table 21. Furthermore, Griko forms the aorist exactly like Modern Greek, that is, using the perfective stem of the verb followed by the imperfect suffixes. Based on Table 21, the aorist suffixes for the 1st, 2nd and 3rd person singular and 3rd person plural appear to be the same; however, the 1st person plural in Griko ends in *-amo* while in Modern Greek it ends in *-ame*, and the 2nd person plural in Griko ends in *-ato* while for Modern Greek it ends in *-ate*. Lastly, the 3rd person plural in Griko ends in

-ane while in Modern Greek it ends in *-an*, with the suffix *-ane* being used in informal settings. One similarity between Griko and Modern Greek is the 1st, 2nd and 3rd person singular as the suffixes appear to be the same.

Table 21: Aorist and *passato remoto* suffix comparison for Griko, Modern Greek and Italian (Karanastasis 1997: 81; Proudfoot & Cardo 2005: 32; Holton et al. 2012: 134)

Person		Language		
		Griko	Modern Greek	Italian
Singular	1st	**-a**	**-a**	-i
	2nd	**-e(s)**	**-es**	-sti
	3rd	**-e**	**-e**	-ò/-e/-ì
Plural	1st	*-amo*	*-ame*	-mmo
	2nd	*-ato*	*-ate*	-ste
	3rd	**-ane**	**-an(e)**	-rono

Table 22 compares the aorist for Griko and Modern Greek for *-o* conjugations verbs and *-ao* conjugation verbs in Griko with 1st conjugation verbs and 2nd conjugation verbs in Modern Greek for the verb *gráfo* 'to write' the verb *agapáo* 'to love', as both verbs are the same in both languages. The primary difference between Griko and Modern Greek regarding *-o* conjugation verbs and 1st conjugation verbs is that Griko maintains the augment *e-* for all persons while it is removed for the 1st and 2nd person plural in Modern Greek, this could suggest that Modern Greek has lost its unaccented syllabic augment. Regarding the stress, it is the same for all persons except for the 3rd person plural where in Griko, the stress falls on the verb stem *e-gráfs-ane*, while for Modern Greek it falls on the augment *e-*, *é-graps-an*, due to the augment carrying the stress. Both languages have similar stress patterns as stress falls on the augment *e-* due to Griko and Modern Greek having a common ancestor, Ancient Greek. The perfective stem in Griko is distinguished by *-fs* while in Modern Greek its distinguished by *-ps*.

Table 22: Aorist comparison in Griko and Modern Greek for the verb gráfo 'to write' and *agapáo* 'to love' (Karanastasis 1997: 83; Holton et al. 2012: 134)

Person		Declensions			
		-o / 1st gráfo 'to write'		-ao / 2nd agapáo 'to love'	
		Griko	Modern Greek	Griko	Modern Greek
Singular	1st	é-grafs-a	é-graps-a	e-gápis-a	agápis-a
	2nd	é-grafs-e(s)	é-graps-es	e-gápis-e(s)	agápis-es
	3rd	é-grafs-e	é-graps-e	e-gápis-e	agápis-e
Plural	1st	e-gráfs-amo	gráps-ame	e-gapis-ámo	agapís-ame
	2nd	e-gráfs-ato	gráps-ate	e-gapis-áto	agapís-ate
	3rd	e-gráfs-ane	é-graps-an(e)	e-gapis-áne	agápis-an(e)

The primary difference between Griko and Modern Greek regarding the formation of the aorist for *-ao* conjugation verbs in Griko and 2nd conjugation verbs in Modern Greek is that in Griko the augment *e-* is added for all persons while there is no augment present in Modern Greek, as it retains the original vowel from the verb stem. Secondly, for the 1st, 2nd and 3rd person singular in both Griko and Modern Greek, the stress falls on the verb stem while for the 1st, 2nd and 3rd person plural, the stress in Griko falls on the suffix.

In conclusion, Griko aorist suffixes are cognate to those of Modern Greek aorist suffixes for both *-o* conjugation verbs and *-ao* conjugation verbs to form the aorist as they appear to the same. Furthermore, Griko, just like Modern Greek, also uses the perfective stem of the verb to form the aorist. Griko shows no similarities to Italian for the aorist.

3.4. The present perfect (*Parakeimenos* in Griko and Modern Greek and *Passato Prossimo* in Italian)

The present perfect is used to express actions or states that started in the past but continue to the present, it indicates actions or states as completed (Holton et al. 2012: 122). The present perfect in Griko and Modern Greek is known as *Parakeimenos* (Παρακειμενος) and in Italian it is known as *Passato Prossimo* 'Compound Perfect' or 'Compound Past' (Proudfoot & Cardo 2005: 32).

The present perfect in all 3 languages is formed analytically, i.e., marking tense using auxiliary verbs and participles, with the auxiliary verb in its finite form followed by the past participle of the verb. The past participle in Griko and Modern Greek is not inflected for gender, case or number, unless the alternative present perfect is used which agrees with all 3 in Modern Greek, while in Italian, verbs whose past participles uses auxiliary verb *essere*, agrees with the subject of the verb. The primary difference between Griko and Modern Greek is that Griko follows the Romance paradigm of Italian for auxiliary verb selection (Holton et al. 2012: 132).

3.4.1. Modern Greek

The present perfect is formed via the present tense of auxiliary verb *éxo* 'have' followed by the non-finite active perfective stem of the verb, which takes the suffix *-ei* as in example 27.

(27) **Modern Greek**

éxo	dés-ei
have.PRES.1SG	tie-PTCP

'I have tied.' (Holton et al. 2012: 129)

Interestingly, there is an alternative present perfect which can also be formed using auxiliary verb *éxo* 'to have' in the present tense followed by the passive perfect participle of verb which takes the suffix *-menos, -meni, -meno* as in example 28 however, it is restricted to active transitive verbs which can occur in the passive.

(28) **Modern Greek**

éxo	gram-meno	to	gramma
have.PRES.1SG	write-PTCP	the	letter

'I have written the letter.' (Holton et al. 2012: 305)

3.4.2. Italian

The present perfect is formed via the present tense of either auxiliary verb *essere* 'be' or *avere* 'have' followed by the past participle of the verb. Motion verbs use *essere* 'be', as in example 30, while all other verbs use *avere* 'have', as in example 29.

(29) **Italian** (auxiliary verb: *avere* ‘to have’)
ho cantato
have.PRES.1SG sing.PTCP
‘I sang’ (Maiden & Robustelli 2007: 32)

(30) **Italian** (auxiliary verb: *essere* ‘to be)
sono venuto/a
be.PRES.1SG come.PTCP
‘I came’ (Maiden & Robustelli 2007)

3.4.3. Griko

The present perfect in Griko is formed via the present tense of either auxiliary verb *exo* ‘have’ or *ime* ‘be’ followed by the active past participle of the verb taking the suffix *-mena,* which is derived from the Classic Greek passive past participle suffix *-menos, -meni, -meno* (Squillaci 2016). Auxiliary verb selection follows a transitive-unaccusative split, i.e., auxiliary selection is determined by the semantic class of the verb, with motion verbs using auxiliary verb *ime* ‘be’, as in example 32, and all other verbs using auxiliary verb *exo* ‘have’, as in example 31 (Squillaci 2016: 41; Schifano, Silvestri & Squillaci 2016: 76).

(31) **Griko** (auxiliary verb: *exo* ‘to have’)
éxo fá-mena
have.PRES.1SG eat-PTCP
‘I have eaten.’ (Ralli 2012: 12)

(32) **Griko** (auxiliary verb: *ime* ‘to be’)
íme artó-mena
be.PRES.1SG come-PTCP
‘I have come.’ (Ralli 2012: 12)

3.4.4. Comparison

It is clear that Griko combines both Modern Greek and Italian to form the present perfect periphrastically, that is, following the Romance model of the transitive-unaccusative split by using auxiliary verb *ime* ‘be’ with motion verbs and auxiliary verb *exo* ‘have’ with all other verbs and by using the suffix *-mena* which is derived from Classical Greek passive past participle suffix *-menos, -meni, -meno,* which is only employed in Modern Greek for active transitive verbs which can occur in the passive.

3.5. The past perfect

The past perfect is used to describe actions or states as having been completed sometime in the past, with the consequences relevant to another point in time in the past which is either specified or implied. The past perfect in Griko and Modern Greek is known as *Ipersidelikos* (Υπερσυντέλικος) while in Italian it is known as *Trapassato Prossimo* (Holton et al. 2012: 122; Proudfoot & Cardo 2005: 32). The past perfect for all three languages is formed periphrastically, i.e., marking tense using auxiliary verbs or participles, with the auxiliary verb in its finite form followed by the past participle of the verb. The past participle in Griko and Modern Greek is not inflected for gender, case or number, unless the alternative past participle is used which is inflected for all three in

Modern Greek, while in Italian, the past participles that use auxiliary verb *essere*, agrees with the subject of the verb.

3.5.1. Modern Greek

The past perfect is formed via auxiliary verb *éxo* 'to have' in the imperfect tense followed by the non-finite active perfective stem of the verb which takes the suffix *-ei* as shown below.

(33) **Modern Greek**

eíxa dés-ei
have.PST.1SG tie-PTCP
'I had tied.' (Holton et al. 2012: 132)

Interestingly, there is an alternative past perfect with a restricted use, which is formed using auxiliary verb *éxo* 'to have' in the imperfect tense followed by the passive perfect participle of verb which takes the suffix *-menos, -meni, -meno* which is inflected for gender and number based on the object of the verb as in example 34, however it is restricted to active transitive verbs which can occur in the passive.

(34) **Modern Greek**

eíxa gram-méno to grámma
have.PST.1SG write-PTCP the letter
'I had written the letter.' (Holton et al. 2012: 305)

3.5.2. Italian

The past perfect in Italian is formed via the imperfect indicative of the auxiliary verb *essere* 'to be' or *avere* 'to have' followed by the past participle of the verb. Motion verbs take auxiliary *essere* 'to be' as in example 36 while the rest of the verbs take *avere* 'to have' as in example 35 shown below.

(35) **Italian**

ave-vo cantato
have-PST.1SG sing.PTCP
'I had sung.' (Maiden & Robustelli 2007: 32)

(36) **Italian**

er-o venuto/a
be-PST.1SG come.PTCP
'I had come.' (Maiden & Robustelli 2007: 32)

3.5.3. Griko

In Griko the past perfect is formed via the auxiliary verb *éxo* 'to have' in the imperfect tense followed by the aorist active suffix *-onda* attached to the verb, as in example 37. The past perfect does not make a distinction in the semantic class of the verb as all types of verbs use auxiliary verb *éxo* 'to have' (Squillaci 2016: 41). The active participle *-onda* is derived from the classical Greek active past participle suffix *-sas, -sasa, -san* which in Medieval Greek was morphologically reduced to form *-sanda* which was further reduced to *-onda*, as shown below.

(37) **Griko**

ícha kám-onda
have.PST.1SG do-PTCP
'I had done.' (Squillaci 2016: 41)

3.5.4. Comparison

It is clear that Griko has retained the Medieval Greek construction in forming the past perfect periphrastically, that is using auxiliary verb exo 'to have' in the imperfect followed by the verb which takes the aorist active particle *-onda*. It is does not follow the Italian paradigm as Italian either selects auxiliary verb *avere* 'to have' or *essere* 'to be' depending on the semantic class of the verb. Griko has no similarities with Italian for the past perfect.

3.6. The progressive aspect

The progressive aspect implies a continuing ongoing process referring to actual events i.e. denotes that the event is ongoing (Payne 1997: 84; Velupillai 2012: 219).

3.6.1. Modern Greek

The progressive aspect in Modern Greek is expressed through the gerund, which is formed synthetically with the addition of the suffix *-ontas* onto 1st and 2nd conjugation verbs (Holton et al. 2012; Manika 2014). The gerund is used to describe action which take place at the same time as the action of the verb it modifies, with the gerund usually expressing the means or the time during which something is done or taken place, as in example 38 whereby the act of John departing and leaving a present for Anna happened at the same. The gerund has the same subject as the verb it modifies (Holton et al. 2012: 135). Only verbs with an active voice can form the progressive aspect while verbs with only a passive voice cannot (Manika 2014: 36).

(38) **Modern Greek**

feúg-ontas o Yiannis áfis-e éna dóro gia tin Ánna
leave-GER he John leave-PST.3SG one present for her Anna
'As John was leaving, he left a present for Anna.' (Holton et al. 2012: 135)

The gerund also has a perfect form which is formed via the attachment of the suffix *-ontas* onto auxiliary verb *éxo* 'to have' followed by the perfective stem of the verb which takes the suffix *-ei* (Holton et al. 2012: 135; Manika 2014: 37).

(39) **Modern Greek**

éx-ontas dés-ei
have-GER tie-PTCP
'having tied' (Holton et al. 2012: 120)

3.6.2. Italian

The progressive aspect in Italian is formed via the conjugation of the verb *stare* 'to stay' in the present tense (example 40) or in the imperfect tense (example 41) followed by the verb taking the suffix *-ando* or *-endo*, with the choice of suffix depending on the conjugation of the verb (Table 35) (Proudfoot & Cardo 2005). When *stare* is conjugated the present tense, it indicates that the action is in progress at the present time, as in

example 40, whereby the act of the boys reading is happening at the present time as we speak (Maiden & Robustelli 2007: 258):

(40) **Italian**
i ragazzi stan-no legg-endo
the boy.PL stay-PRS.3PL read-GER
'The boys are reading.' (Proudfoot & Cardo 2005: 136)

When *stare* is conjugated in the imperfect tense, it indicates that the action was in progress at a past time, as shown in example 41, whereby the act of the boys reading happened at a past time (Maiden & Robustelli 2007: 258).

(41) **Italian**
i ragazzi sta-vano legg-endo
the boy.PL stay.PST.IPFV.3PL read.GER
'The boys were reading.' (Proudfoot & Cardo 2005: 137)

Table 23: Gerund formation in Italian (Maiden & Robustelli 2007: 258)

Verb ending	Gerund	Example
-are	-ando	parlare>parlando 'to talk' > 'talking'
-ere	-endo	leggere>leggendo 'to read' > 'reading'
-ire		dormire>dormendo 'to sleep' > 'sleeping'

3.6.3. Griko

As proposed by Golovko & Panov (2013) and Squillaci (2016), the progressive aspect consists of a verb form which is preceded by the unchangeable particle *ste* that does not carry any inflection for tense, aspect, mood, person, number and gender, which is a reduced form of the grammaticalized lexical verb *steo* 'to stay' which was borrowed from the Italian verb *stare* 'to stay', followed by the verb in the appropriate tense; example 42 is the present tense while example 43 in the imperfect as shown below.

(42) **Griko**
ce sto daso, mian alipuna *ste* kulua enan alao
and in forest, a fox PROG chase.PRES.3SG one rabbit
'and a fox is chasing a rabbit in the forest' (Golovko & Panov 2013: 76)

(43) **Griko**
ce sto daso, mian alipuna *ste* kulus-a enan alao
and in forest a fox PROG chase-IPFV.3SG one rabbit
'and a fox was chasing a rabbit in the forest' (Golovko & Panov 2013: 76)

Interestingly, the progressive aspect can also be used to for future situations, as shown in example 44.

(44) **Griko**

avri *ste* pam-e totzu
tomorrow PROG go-PRES.1PL field
'Tomorrow we will go to the field.' (Golovko & Panov 2013: 76)

3.6.4. Comparison

It is clear that Griko forms the progressive aspect in a combination of Italian, by using the grammaticalized particle *ste* which was borrowed from the verb *stare* 'to stay', and a combination of Modern Greek, by the verbs being conjugated in the appropriate tense.

4. Conclusion

In conclusion Griko seems to have retained a vast majority of the Modern Greek verbal morphology as verbs appear to work in the same way as Modern Greek, despite the long contact with Italian. This is based on Griko being obviously typologically closer to Modern Greek, which is due to both languages being derived from a common ancestor. The findings in this paper suggest that the only influence of Italian regarding verbal morphology on Griko is on the present perfect, where Griko either selects auxiliary verb *éxo* 'to have' or *ime* 'to be' following the Romance model of auxiliary verb selection and on the progressive aspect, where Griko uses the grammaticalized particle *ste* 'stand', which is borrowed from Italian from the verb *stare* 'to stay'. This suggest that language contact and language change can take place, but with different degrees of intensity as in Griko, which has maintained a lot of Modern Greek elements in the language, as it is seen in the verbal morphology.

References

Douri, Angeliki & Dario De Santis. 2015. Griko and Modern Greek in Grecia Salentina: An overview. University of Salento.

Golovko, Ekaterina & Vladimir Panov. 2013. Salentino dialect, Griko and regional Italian: Linguistic diversity of Salento. *Working Papers of the Linguistics Circle* 23(1). 51–80.

Holton, David, Peter Mackridge, Irene Philippaki-Warburton & Vassilios Spyropoulos. 2012. *Greek: A comprehensive grammar*. 2nd edn. Routledge.

Karanastasis, Anastasios. 1997. *Η Γραμματική των Ελληνικών Ιδιωμάτων της Κάτω Ιταλίας [A grammar book of the Southern Italian Greek Dialects]*. 1st edn. Athens: Academy of Athens.

Maiden, Martin & Cecilia Robustelli. 2007. *A reference grammar of modern Italian*. 2nd ed. London: Hodder Arnold.

Manika, Eleni. 2014. *Το Γερούνδιο στη Νέα Ελληνική [The Gerund in Modern Greek]*. University of Ioannina MA thesis.

Payne, Thomas E. 1997. *Describing morphosyntax*. Cambridge: Cambridge University Press.

Pellegrino, Manuela. 2016. Performing Griko beyond "death." *Palaver* 5 n.s.(1). 137–162.

Proudfoot, Anna & Francesco Cardo. 2005. *Modern Italian grammar: A practical guide*. 2nd edn. Routledge.

Ralli, Angela. 2012. Verbal loanbleds in Griko and Heptanesian: A case study of contact morphology. *L'Italia dialettale: Rivista di dialettologia italiana*. Giardini Editori (73). 111–132.

Schifano, Norma, Giuseppina Silvestri & Maria Olimpia Squillaci. 2016. Some preliminary observations on auxiliary selection and participle agreement in Greko and Bovese. *Modern Greek Dialects and Linguistics Theory* 6. 166–177.
Squillaci, Maria Olimpia. 2016. *When Greek meets Romance: A morphosyntactic investigation of language contact in Aspromonte*. University of Cambdridge PhD dissertation.
Velupillai, Viveka. 2012. *An introduction to linguistic typology*. John Benjamins Publishing Company.

SOAS Working Papers in Linguistics. Volume 20 (2021): 72–94

Reverse complex predicates or something else? The case of Hindi *de* 'give' and *jaa* 'go'

Andriana Koumbarou, Shubham K. Srivastava
& Sarada B. Biswas

Abstract
"Aspectual" complex predicates (ACPs) in Hindi are formed by a sequence of two verbs that together describe a single event. The main verb is realised in root form and contributes lexical meaning, followed by a light verb which carries TAM morphology and somehow modifies the event ($V_{MAIN}V_{LIGHT}$ ordering). However, it has been noted that some verb-verb combinations allow a "reverse" order in which a light verb in root form precedes the main verb inflected for TAM ($V_{LIGHT}V_{MAIN}$ ordering). In this paper we take the light verbs *de* 'give' and *jaa* 'go' as case studies and trace their use in standardly ordered complex predicates, as well as what look like cases of reversal. We present our initial findings and argue that the unusual $V_{LIGHT}V_{MAIN}$ sequences are not simply a reordered variant of a standard complex predicate but, in fact, exhibit interpretational differences (e.g., intentionality) that stem from the initial placement and lexical semantics of directed-action *de* 'give' and directed-motion *jaa* 'go'.

Keywords: (reverse) complex predicates, light verbs, intentionality, directionality, Hindi

1. Introduction

This paper explores little discussed constructions in Hindi[1] which at least at first glance look like reordered variants of standard "aspectual" complex predicates (term introduced by Butt 1995; henceforth ACPs).[2] Standard ACPs consist of a main verb in root form which contributes lexical meaning, followed by a light verb that carries TAM morphology and provides additional information about the event (Abbi & Gopalakrishnan 1991; Butt & Ramchand 2005; Butt 1995; Butt 2010; Butt & Lahiri 2013; Poornima 2012; Hook 1973; Hook 1993). Examples of standardly ordered ACPs with the light verbs *jaa* 'go' and *de* 'give' ($V_{MAIN}V_{LIGHT}$ ordering) are given in (1).[3]

(1) **Standard ACP: $V_{MAIN}V_{LIGHT}$**

a. laṛkii seb **k^{h}aa gayii**
girl.F apple.M eat go.PFV.F.SG
'The girl ate the apple.'

[1] Hindi and Urdu are structurally very similar and in this paper we draw examples from the literature on both when discussing previous work. Data for this paper also comes from the Emille Hindi Corpus (www.emille.lancs.ac.uk - last accessed 14 February, 2021) and is cited accordingly. Unless otherwise specified, examples are constructed by us and checked with Hindi speakers in Uttar Pradesh, India.

[2] We are grateful to Professor Tista Bagchi and Professor Pradeep Kumar Das for encouraging us to pursue the topic of this paper, as well as Reeta Srivastava and Saumya Srivastava for countless hours of discussing data with us. We would also like to thank Joseph Lovestrand and Naresh Sharma for their comments which helped significantly improve the paper. Any mistakes are our own.

[3] Interlinearized examples follow the Leipzig Glossing Rules with the addition of the following abbreviations: CONJ 'conjunctive', EMPH 'emphatic' and INSTR 'instrumental'.

b. laṛkii=ne laṛke=ko (patt^{h}ar) **maar** **diyaa**
girl.F=ERG boy.M.OBL.SG=ACC/DAT stone.M hit give.PFV.M.SG
'The girl hit the boy (with a stone).'

It has been noted in the literature that a "reverse" order is also allowed, as shown in (2) where the verb that makes a light contribution precedes the lexically more dominant verb (Das 2015; Hook 1973; Poornima 2012; Poornima & Koenig 2009). In the "reverse" order the light verb is realised in root form and the main verb carries TAM morphology ($V_{LIGHT}V_{MAIN}$ ordering).

(2) **"Reverse" construction: $V_{LIGHT}V_{MAIN}$**

a. laṛkii=ne seb **jaa** **k^{h}aayaa**
girl.F=ERG apple.M go eat.PFV.M.SG
'The girl ate the apple (deliberately).'

b. laṛkii=ne laṛke=ko #(patt^{h}ar) **de** **maaraa**
girl.F=ERG boy.M.OBL.SG=ACC/DAT stone.M give hit.PFV.F.SG
'The girl hit the boy with a stone (deliberately).'

Constructions such as (2a) and (2b) have been described in previous work as "reverse" or "reordered" variants of standard complex predicates (see, for example, Das 2015; Hook 1973; Poornima 2012). The examples in (2), however, come with an intentional reading which the constructions in (1) lack. Further, while in (1b) *patt^{h}ar* 'stone' can be dropped, omitting it in (2b) where *de* 'give' is realised first would be infelicitous (as indicated with the hash symbol #). These observations call for a more careful investigation of structural and semantic aspects of the unusual $V_{LIGHT}V_{MAIN}$ sequences.

In this paper we make first steps in this direction, taking the light verbs *de* 'give' and *jaa* 'go' as case studies. We argue that the unusual $V_{LIGHT}V_{MAIN}$ ordering is pragmatically marked and propose that the inherent semantics of directed-motion *jaa* 'go' and directed-action *de* 'give' provide the base for interpretive effects such as intentionality to arise. In what follows, Section 2 gives an overview of previous work on complex predicates in Hindi and closely related Urdu. Sections 3 and 4 trace the use of *de* 'give' and *jaa* 'go' in the different orders. Finally, Section 5 draws a conclusion.

2. Background and previous work

2.1. Morphosyntactic properties

ACPs[4] in Hindi, and closely related Urdu, have received considerable attention so far (Abbi & Gopalakrishnan 1991; Butt 1995; 2010; Butt & Lahiri 2013; Butt & Ramchand 2005; Hook 1973; 1993; Singh 1998; Poornima 2012; Poornima & Koenig 2009). ACPs are formed with a main verb in bare stem form that contributes lexical meaning, followed by a light verb (or vector verb) which carries TAM morphology and contributes additional information that relates to aspect (telicity) but also other semantic notions such as volitionality, benefaction, suddenness, and forcefulness (see Section 2.2). A single verb construction is illustrated in (3) with the perfective verb *paḍ^{h}ii* 'read'. In (4) we observe a complex predicate formed with the verbal stem *paḍ^{h}* 'read' followed by the light verb

[4] Often also referred to as "compound verb" constructions (see Hook 1973; Hook 1993).

de ‘give’ carrying perfective morphology. While both (3) and (4) describe an event of reading, in (4) the light verb adds a sense of completion and emphasises the outward direction of the action,[5] i.e., the agent read the book out loud (see Section 3.1 for a detailed discussion on the semantics of *de* ‘give’).

(3) **Single verb construction**
mãĩ=ne ye kitaab **paḍʰii**
1SG=ERG this book.F read.PFV.F.SG
‘I read this book.’

(4) **Standard ACP with light verb *de* ‘give’**
mãĩ=ne ye kitaab **paḍʰ dii**
1SG=ERG this book.F read give.PFV.F.SG
‘I read this book (out loud).’

There is general agreement that historically ACPs originate from a Sanskrit structure which involved an adverbial participle (‘having X-ed’). The verb-verb sequence was reanalysed as a monoclausal structure with two co-predicating verbs (Butt 1997) in which the last verb is light and makes a weaker contribution. This happened with only a handful of verbs such as *de* ‘give’, *le* ‘take’, *ḍaal* ‘put’, *maar* ‘hit’, *jaa* ‘go’, *aa* ‘come’, *paṛ* ‘fall’, *baiṭʰ* ‘sit’, *nikal* ‘emerge’, *uṭʰ* ‘rise’. With evidence from object agreement, anaphora and control tests, Butt (1995) shows that the complex predicate construction, as in (4), has a single subject and no embedding.

Synchronically, however, some verb-verb sequences are potentially ambiguous between a complex predicate reading and an adverbial participle reading, though the different readings can be disambiguated prosodically (Butt 1997). This is shown in (5). The translation in (a) indicates the biclausal reading in which the first verb *toṛ* ‘break’ is embedded and the clause-final *ḍaalii* ‘put’ is the perfective matrix verb. In this case the subject of the embedded construction is controlled by the matrix subject. In the complex predicate reading in (b), on the other hand, the two verbs describe a single event: the main verb *toṛ* ‘break’ provides the event description, followed by the light verb *ḍalii* ‘put’. This ambiguity can be removed with the use of the conjunctive participle marker *-kar/-ke* to describe a temporal sequence of events, as shown in (6).

(5) raam=ne kachre=mẽ botal **toṛ** **ḍaalii**
Ram.M=ERG bin=in bottle.F break put.PFV.SG.F
(a) ‘Having broken the bottle, Ram put it in the bin.’
(b) ‘Ram broke the bottle in the bin.’

(6) raam=ne kachre=mẽ botal **toṛ-kar** **ḍaalii**
Ram.M=ERG bin=in bottle.F break-CONJ put.PFV.SG.F
‘Having broken the bottle, Ram put it in the bin.’

In his detailed study on Hindi verb-verb complex predicates, Hook (1973) notes that an unusual “reverse” order is also possible in which the light verb in root form precedes the

[5] Ozarkar and Ramchand (2018) note the same for the Marathi light verb ‘give’: it expresses the outward directedness of an action (i.e., away from the agent).

lexically dominating verb that provides the event description and carries TAM morphology. This is shown in (7).

(7) isii gam=mẽ govindacharya=ne apnaa sir
this.EMPH.OBL sadness=in Govindacharya.M=ERG own.M.SG head.M

diivaar=par **de maaraa**
wall=on give hit.PFV.M.SG
'In (/due to) this sadness Govindacharya hit his head on the wall.'
(Hindi Emille Corpus via Poornima 2012: 92)

Poornima (2012) shows with data from movement, adverbial modification and coordination tests that the "reverse" orderings are also monoclausal structures with a tight verb-verb constituent. The difference, she argues, between ACPs and the "reverse" ordering is in syntactic headedness (see also Poornima & Koenig 2009). The light verb modifies the meaning of the main verb in both orderings but in standard ACPs it is also the syntactic head, whereas in the reverse construction, as in (7), this role is taken by the clause-final main verb. Poornima further shows that both orders are reluctant to negation but only standard ACPs allow passivisation (though there is no explanation for this to date).

The only other study we know of that explores structural properties of the "reverse" ordering is Das (2015). Das relates possibilities for reversal to transitivity. He argues that verbs that match in terms of transitivity can reverse, as in (8) and (9), but an intransitive light verb such as *jaa* 'go' cannot be placed before a transitive main verb as it cannot support an internal argument. However, Poornima (2012) treats (10) as a case of a "reverse" complex predicate in which the intransitive light verb *jaa* 'go' precedes the transitive main verb *bech* 'sell'.

(8) a. mãĩ tʰak-kar beḍ=pe **leṭ gayaa**
1SG tired-CONJ bed=on lie.down go.PFV.M.SG

b. mãĩ tʰak-kar beḍ=pe **jaa leṭaa**
1SG tired-CONJ bed=on go lie.down.PFV.M.SG
'Being very tired, I fell on the bed.' (Das 2015: 172)

(9) a. srijan=ne gusse=mẽ gilaas **toṛ diyaa**
Srijan.M=ERG anger=in glass.M break give.PFV.M.SG

b. srijan=ne gusse=mẽ gilaas **de toṛaa**
Srijan.M=ERG anger=in glass.M give break.PFV.M.SG
'Srijan broke the glass in anger.' (Das 2015: 169)

(10) raam=ne apnaa makaan **jaa bechaa**
Ram.M=ERG own.M.SG house.M go sell.PFV.M.SG
'Ram$_i$ sold his$_i$ house.' (Hook 1975 via Poornima 2012: 117)

In previous work so far, the unusual $V_{LIGHT}V_{MAIN}$ sequences have generally been described as "reverse" or "reordered" variants of standard ACPs. However, switching the order of verbs in (10) does not easily allow a complex predicate reading as per native speaker intuitions. This is shown in (11) where a reading in which the two verbs describe a temporal sequence of events (a biclausal structure) is much preferred. Further, the example in (10) carries a sense of intentionally and resembles the English *go and* pseudo-coordination construction in (12) (see Vos 2004). In English, the verb *go* in the first conjunct does not require an actual physical movement reading but may be used to express surprise (Ross 2016).

(11) raam apnaa makaan **bech gayaa**
Ram.M own.M.SG house.M sell go.PFV.M.SG
'Having sold his$_i$ house, Ram$_i$ left.'

(12) 'He went and sold the house (despite what his relatives told him).'

These observations suggest that $V_{LIGHT}V_{MAIN}$ sequences are perhaps not best thought of as semantically equivalent, reordered variants of standard ACPs; instead, the unusual $V_{LIGHT}V_{MAIN}$ ordering may have developed independently driven by the directional semantics of verbs such as *de* 'give' and *jaa* 'go' to serve pragmatic functions (e.g., to signal that there is something unusual about an event). Before we discuss this in more detail, we provide an overview of previous work on the semantics of light verbs in the next section.

2.2. Semantics of light verbs

Generally, there is consensus that light verbs in standard ACPs lead to completive readings but also contribute some additional information which has been notoriously hard to pinpoint. Different light verbs are said to give rise to different readings to do with completion, volitionality, suddenness, benefaction, forcefulness, regret, affectedness (Abbi & Gopalakrishnan 1991; Butt 2010; Hook 1973; Hook 1991; Hook 1993; Kachru 2006; Masica 1976; Poornima 2012; Singh 1998; Singh 1994). For example, several studies explain that the light verb *de* 'give' indicates "other-benefaction" (Abbi & Gopalakrishnan 1991; Hook 1973; Kachru 2006), i.e., the agent's actions are benefiting others as indicated in (13). Poornima (2012) argues for an analysis in terms of "affectedness" rather than benefaction. She explains that in (14) *de* 'give' indicates affectedness of a non-subject referent (i.e., the ruining of the house is understood to affect others).

(13) ek kamiiz **sil-vaa do**
one shirt.F tailor-CAUS give.IMP
'Get a shirt made (for another).' (Abbi & Gopalakrishnan 1991: 692)

(14) apnii burii aadatõ=ke kaaraṇ apnaa g^{h}ar
own.F bad.F habit.F.OBL.PL=GEN.OBL because own.M.SG house.M

ujaaṛ diyaa
ruin give.PFV.M.SG
'His/her bad habits ruined his/her house.' (adapted from Poornima 2012: 208)

Butt and Geuder (2001) argue that light verbs are a special class of their own and, unlike auxiliaries, they carry lexical meaning, albeit in a weak, schematic manner. They discuss the use of the light verb *de* 'give' in combination with different main verbs and show that in addition to completive readings (i.e., the event reads as an achievement), the constructions also imply agentivity. This covers unintentional causation, as in (15) and (16), as well as (17) for which the authors argue that the light verb adds a sense of responsibility on the agent for the loss of the wallet.[6]

(15) us=ne b^{h}uul=se gilaas **toṛ diyaa**
3SG.OBL=ERG forget=INSTR glass.M break give.PFV.M.SG
'He/she broke a/the glass by accident.'
(Hook 1974: 63 via Butt & Geuder 2001: 344)

(16) b^{h}uul=se muj^{h}e apnaa sahii naam **bataa diyaa**
forget=INSTR 1SG.ACC/DAT own.M.SG true name.M tell give.PFV.M.SG
'He/she$_i$ inadvertently told me his/her$_i$ real name.'
(Hook 1974: 273 via Butt & Geuder 2001: 345)

(17) kisii=ne baṭuaa **k^{h}o diyaa**
someone.OBL=ERG wallet.M lose give.PFV.M.SG
'Someone lost a/the wallet.' (Hook 1974: 310 via Butt & Geuder 2001: 345)

Butt and Geuder recognise completion and agentivity as the meanings that are consistent with the use of the light verb *de* 'give'. They argue that other readings such as the presence of a recipient, as in (16), or forcefulness, as in (18) and (19), are dependent on the meaning of the main verb that introduces the event description. In other words, these semantic features reside in the light verb but their "activation" is dependent on the meaning of the main verb. When the feature is compatible with the meaning of the main verb, the light verb adds it or enforces it.

[6] Note that it is precisely because of this sense of responsibility that the construction in (17) with the indefinite subject *kisii=ne* 'someone' (i.e., an unknown agent) might be deemed odd when presented out of context. As per native speaker intuitions the structure in (i) with the ergative pronoun *us=ne* is a significant improvement:

(i) us=ne p^{h}ir=se baṭuaa k^{h}o diyaa
3SG.OBL=ERG again=INSTR wallet.M lose give.PFV.M.SG
'He lost the wallet again.'

(18) us=ne gusse=mẽ aa-kar kitab zamiin=par
3SG.OBL=ERG anger.OBL=in come-CONJ book.F floor=on

paṭak dii
throw give.PFV.F.SG
'In a fit of rage he/she dashed the book to the floor.' (Butt & Geuder 2001: 344)

(19) us=ne dushman=ko paanii=mẽ **ḍubaa diyaa**
3sg.OBL=ERG enemy.M/F=ACC/DAT water=in drown give.PFV.M.SG
'He/she drowned the enemy in water (forcefully).' (Butt & Geuder 2001: 347)

Structurally, Butt and Geuder argue that the light verb *de* 'give' has lost the THEME argument slot. In (20) the direct object *skuuṭar* 'scooter' is the patient of washing; it does not hold a theme relation with 'give'. They also point out that a beneficiary argument is absent; the beneficiary in (20) is realised as an optional adjunct.

(20) tum=ne (us=ke-liye) skuuṭar **dʰo dii**
2SG=ERG 3SG.OBL=for scooter.F wash give.PFV.F.SG
'You washed the scooter (for him/her).' (adapted from Butt & Geuder 2001: 348)

Light verbs, therefore, have an incomplete argument structure and do not introduce an event description. They simply modify the event description projected by the main verb and contribute additional semantic features. Butt and Geuder (2001) speculate that the only possible analysis seems to be one in which the event argument is construed as the theme. Following a neo-Davidsonian representation approach (see Parsons 1990), they propose the representation in (22) for the construction in (21). The event predicate GIVE-TYPE is interpreted in conjunction with the main event description introduced by the main verb *pʰẽk* 'throw'. The thematic roles AGENT and THEME are to be understood here in the sense of Dowty's (1991) proto-roles as verbal entailments.

(21) tum=ne kuuṛaa us=ke sir=par **pʰẽk diyaa**
2SG=ERG garbage.M 3SG.OBL=GEN.OBL head=on throw give.PFV.M.SG
'You threw the garbage on his/her head.' (Butt & Geuder 2001: 357)

(22) a. throw(e)(you, the garbage) & GOAL(e, on his/her head) & GIVE-TYPE(e)

b. GIVE-TYPE(e) = e involves the force emission/transmission pattern AGENT(e)-THEME(e) (Butt & Geuder 2001: 357)

Butt and Lahiri (2013) further elaborate on this idea by stressing that light verbs are form-identical with main verbs and are diachronically stable (but see Slade (2021) for counter-arguments). They argue for a single underlying lexical entry for both main and light uses that consists of lexical entailments and information to do with world knowledge (e.g., whether an event is (in)volitional, sudden, for the benefit of someone, etc.). Depending on whether the verb enters the syntax as a main verb or a light verb, lexical information either feeds a full event description and argument structure (main verbs) or modifies an existing event description (light verbs). Contrary to claims that light verbs are semantically bleached (Hook 1973; Hook 1993), Butt and Lahiri argue that it is not the

case that "light" meanings are derived from "full" lexical specifications. Rather, it is precisely the general nature of these verbs that allows them to double as light and main verbs.

The interpretation of the so-called "reverse" constructions is discussed to a much lesser extent in the literature. Hook (1973: 56) explains that the example in (23) "implies a suddenness not found in the unreordered sequence."

(23) kitaab zamiin=par **de paṭkii**
book.F floor=on give slam.PFV.F.SG
'He/She slammed the book to the floor.' (adapted from Hook 1973: 55)

More recently, Das (2015) argues that the ordering of verbs is driven by pragmatic factors. Similarly to Hook (1973), he explains that the "reverse" order in (24) and (25) reveals "suddenness of performing the actions" (Das 2015: 182). Other readings that Das argues to arise as a result of the unusual order are agent's intentionality and/or anger in performing an action, for which he provides the examples in (26) and (27), respectively. A further use outlined by Das is to express an uncontrolled action as in (28), i.e., to express that "this ought not to have happened" (Das 2015: 186). However, Hook (1973) and Das (2015) do not discuss in more detail the interpretive effects associated with the unusual ordering and these remain not very well understood and in need of further study.

(24) laṛke taalaab=mẽ **jaa kude**
boy.M.PL pond=in go jump.PFV.M.PL
'The boys jumped into the pond.' (Das 2015: 180)

(25) us=ne mere piiṭ^{h}=par ek mukkaa **de maaraa**
3SG.OBL=ERG 1SG.POSS.OBL back=on one punch.M give hit.PFV.M.SG
'He/She (suddenly) punched me on my back.' (Das 2015: 181)

(26) mãĩ b^{h}iiṛ=mẽ kisii tarah **jaa g^{h}usaa**
1SG crowd=in some way go enter.PFV.M.SG
'Somehow, I managed to get into the crowd.' (Das 2015: 183)

(27) kavitaa=ne saarii miṭṭ^{h}aaii kuṛedaan=mẽ **de ḍaalii**
Kavita.F=ERG all sweet.F dustbin=in give put.PFV.F.SG
'Kavita threw all the sweets in the dustbin.' (Das 2015: 183)

(28) gend naalii=mẽ **jaa luṛ^{h}kii**
ball.F drain=in go roll.PFV.F.SG
'The ball rolled into the drain.' (adapted from Das 2015: 185)

In summary, in previous work both standard ACPs and the "reverse" constructions have been shown to behave like monoclausal structures (but the different orders do show differences such as the inability of the "reverse" construction to passivise). Butt and Geuder (2001) discuss in detail light verbs in standard ACPs and argue that they carry lexical meaning which interacts with the contribution of the main verb. The so-called "reverse" constructions, however, have received significantly less attention. While some

works explain that $V_{LIGHT}V_{MAIN}$ sequences carry a sense of intentionality, anger, lack of control and/or suddenness not observed in standard ACPs (Hook 1973; Das 2015), the source of such interpretive effects has not been explored in more detail.

In this paper we argue that such extra dimensions of meaning are determined contextually but are linked to the inherent directional semantics of *de* 'give' and *jaa* 'go'. The initial placement of *de* 'give' and *jaa* 'go' draws attention to the directed-action and directed-motion aspect of an event, respectively, and leads to marked interpretations. We suggest that, rather than a reordered variant of a standard ACP, the unusual $V_{LIGHT}V_{MAIN}$ constructions may have developed independently to highlight an out-of-the-ordinary action or outcome. To understand if this is indeed the case, a closer investigation of structural and semantic aspects of the so-called "reverse" constructions is needed, as well as their use in context. If we are on the right track, this would further strengthen Butt and Lahiri's (2013) proposal for a single lexical entry that reflects the general nature of verbs that have both main and light uses. In what follows, we provide an overview of the uses of the light verbs *de* 'give' (Section 3) and *jaa* 'go' (Section 4) in the different orders.

3. The light verb *de* 'give'

We first give a brief overview of the uses of *de* 'give' in standard ACPs (Section 3.1) before discussing the "reverse" construction (Section 3.2) and associated interpretive effects (Section 3.3).

3.1. Standard ACPs with *de* 'give': an overview

In its standard clause-final ACP use, *de* 'give' is a very common light verb that surfaces with transitive and unergative verbs to describe events that involve an agent-originating outwardly directed action. This is in essence the argument that Ozarkar and Ramchand (2018) make for the Marathi light verb 'give': it expresses the outward directedness of an action (i.e., away from the agent), and this determines what main verbs it can combine with. Their argument applies well to Hindi too and is also in keeping with Butt and Geuder's (2001; 2003) proposal that the semantics of *de* interacts with the meaning of the main verb to give a sense of completion and agentivity and – when the main verb allows it – forcefulness and/or the presence of a beneficiary/recipient (see Section 2.2).

Expectedly, the light verb *de* readily combines with ditransitive main verbs such as *bʰej* 'send' (29) that describe a transfer of an entity (a theme) away from the agent and towards a recipient.

(29) is gʰaṭnaa=kii riporṭ kendra sarkaar=ko
this.OBL incident=GEN.F report.F central government.F=ACC/DAT

bʰej dii gayii hai
send give.PFV.F.SG PASS.PFV.F.SG be.PRS.3SG
'The report of this incident has been sent to the central government.'
(Emille Hindi Corpus; ehinweb191)

It also surfaces with verbs that do not express a transfer of a physical object but which nevertheless can be understood as involving the transfer or outward emission of a more abstract entity (e.g., information, sound), as in (30) and (31) with the main verbs *bataa* 'tell' and *kah* 'say', respectively. Similarly, the use of the light verb *de* with the main verb *paṛʰ* 'read' in (32) indicates that the agent read the book out loud.

(30) mãĩ=ne saaraa kissaa aapne bʰaaii=ko
1SG=ERG whole.M story.M own.M.OBL brother.M=ACC/DAT

bataa diyaa
tell give.PFV.M.SG
'I told my brother the whole story.'

(31) us=ne saaf taur=par **kah diyaa** kii ye sab
3SG.OBL=ERG clear way=on say give.PFV.M.SG that this all

bakvaas hai
nonsense be.PRS.3SG
'He said clearly that this is all nonsense.' (Emille Hindi Corpus; ehinweb1ea)

(32) mãĩ =ne kitaab **paṛʰ dii**
1SG=ERG book.F read give.PFV.F.SG
'I read the book (out loud).'

The light verb *de* 'give' combines with a wide range of transitive main verbs that describe the emission/exertion of force originating from the agent towards some other entity. This is observed in (33)-(35). In (33) and (34) the argument affected by the agent's actions carries the accusative/dative *=ko* marker.[7] In (35) it is the unmarked object *log* 'people' that is on the receiving end of the agent's actions.

(33) maalikõ=ne ek raat us=ko naukrii aur
owner.OBL.PL=ERG one night 3SG.OBL=ACC/DAT job and

gʰar=se baahar **nikaal diyaa** tʰaa
house=from outside remove/take.out give.PFV.M.SG be.PST.M.SG
'One night the owners threw her out of the job and house.'
(adapted from Emille Hindi Corpus; ehinweb141)

(34) tum=ne mujʰe dʰaratii=ka sab=se sukʰii ĩsaan
2SG=ERG 1SG.ACC/DAT earth=GEN.M all=from happy person.M

banaa diyaa
make give.PFV.M.SG
'You made me the happiest person on earth.'
(EMILLE Hindi Corpus; ehinweb006)

[7] The *=ko* marker surfaces obligatorily on indirect objects and direct objects high in animacy and specificity (Butt 1993; Bhatt & Anagnostopoulou 1996).

(35) puliskarmiyõ=ne vahãã=se hazaarõ=kii sãkʰyaa=mẽ
policeman.OBL.PL=ERG there=from thousand.PL=GEN.F number.F=in

log **bʰaag-aa diye** tʰe
people.M run-CAUS give.PFV.M.PL be.PST.M.PL
'Policemen drove away thousands of people from there.'
(adapted from Emille Hindi Corpus; ehinweb170)

However, the use of the light verb 'give' does not require the realisation of a beneficiary argument (also noted by Butt and Geuder 2001, 2003). For example, (36)-(37) could be uttered in a context in which the door was opened and the sweets were made for someone's benefit but this is by no means a requirement. The light verb 'give' does not enforce that there is a beneficiary participant; rather, it lends easily to such readings due to its directional semantics.

(36) us=ne darvaazaa **kʰol diyaa**
3SG.OBL=ERG door.M open give.PFV.M.SG
'He opened the door.'

(37) raadʰaa=ne mitʰaaii **banaa dii**
Radha.F=ERG sweet.F make give.PFV.F.SG
'Radha made sweets.'

Following this line of thought, the use of *de* 'give' with unergatives such as *ro* 'cry' in (38) and *kʰããs* 'cough' in (39) is expected as these describe outward emission events. Note that the subjects in (38)-(39) do not carry the ergative marker. Contrary to descriptive generalisations, we have found that the ergative marker *=ne* is not obligatory with the light verb *de* 'give' when combined with unergative main verbs.[8] If *=ne* is used, however, it implies that the agent has control over the action, e.g., coughing on purpose as indicated in (40).We take this to mean that purposeful and/or control over the action readings have to do with the use of the ergative marker and not with the light verb *de* 'give'.

(38) laṛkii **ro dii**
girl.F cry give.PFV.F.SG
'The girl cried.'

(39) laṛkii **kʰããs dii**
girl.F cough give.PFV.F.SG
'The girl coughed.'

(40) laṛkii=ne **kʰããs diyaa**
girl.F=ERG cough give.PFV.M.SG
'The girl coughed (on purpose).'

[8] See Butt (2017) for a detailed discussion on split-ergativity in Hindi/Urdu.

As Ozarkar and Ramchand (2018) note for Marathi, the light verb *de* ‘give’ in Hindi does not co-occur with main verbs that express an agent-oriented action. This is illustrated with the contrast between (41) and (42). In (41) the light verb *de* ‘give’ can be realised with the main verb *kʰilaa* ‘feed’ but it is ungrammatical with *kʰaa* ‘eat’ in (42). The same extends to other verbs that express an inwards action: the light verb *de* ‘give’ cannot be used with *siikʰ* ‘learn’, *pii* ‘drink’, *nahaa* ‘bathe’, *samajʰ* ‘understand’ but can be used with the causative *sikʰaa* ‘teach’, *pilaa* ‘water/give water’, *nehlaa* ‘bathe (someone)’, *samjʰaa* ‘explain’.

(41) mãĩ=ne bachche=ko anaanaas **kʰilaa** **diyaa**
1SG=ERG child.M.OBL.SG=ACC/DAT pineapple.M feed give.PFV.M.SG
‘I fed the child pineapple.’

(42) *mãĩ=ne seb **kʰaa** **diyaa**
1SG=ERG apple.M eat give.PFV.M.SG
‘I ate the apple.’

To conclude, the light verb *de* ‘give’ in standard $V_{MAIN}V_{LIGHT}$ ACPs combines with main verbs that describe agentive events to reinforce the outwards directionality of the agent’s action. The light verb *de* does not encode directly volitional/intentional readings and it does not affect argument structure; it simply adds a layer of meaning to the event predication projected by the main verb by contributing its directional semantics. In what follows, we turn to exploring the rarer $V_{LIGHT}V_{MAIN}$ sequences in which the verbal stem *de* ‘give’ is realised before a lexically more dominant verb (*de* + V sequences).

3.2. “Reverse” ordering: *de* + V sequences

Butt (1995) argues for monoeventiveness and monoclausality to be defining features of complex predication. For a *de* + V sequence to be established as a (type of) complex predicate, the two verbs in the sequence need to express a single event within a monoclausal structure. Agreement is a reliable test for monoclausality as in Hindi perfective transitive verbs show agreement with an unmarked object[9] (for other tests see Poornima & Koenig 2009; Poornima 2012). For example, in (43) agreement morphology on the finite verb *maarii* ‘hit’ indicates that *kitaab* ‘book’ is a matrix object and there is no embedding. The verbal root *de* ‘give’ does not describe an event of ‘giving’. In contrast, (44) is a biclausal structure in which the feminine *gend* ‘ball’ is part of the participial adverbial *gend de-kar* ‘having given the ball’ (see Section 2.1) and the perfective matrix verb *maaraa* ‘hit’ shows default masculine agreement.

[9] In the perfective, transitive subjects carry the ergative marker and the verb agrees with an unmarked object, as shown in (ii). If the object is marked, the perfective verb defaults to third person, singular, masculine agreement, as in (iii). For more details, see Butt (2017).

(ii) miiraa=ne taaraa=ko chʰarii maarii
Mira.F=ERG Tara.F=ACC/DAT stick.F hit.PFV.F.SG
‘Mira hit Tara with a stick.’

(iii) miiraa=ne taaraa=ko maaraa
Mira.F=ERG Tara.F=ACC/DAT hit.PFV.M.SG
‘Mira hit Tara.’

(43) miiraa=ne raam=ko kitaab **de** **maarii**
Mira.F=ERG Ram.M=ACC/DAT book.F give hit.PFV.F.SG
'Mira hit Ram with a book.'

(44) laṛkii=ne laṛke=ko gend **de-kar** **maaraa**
girl.F=ERG boy.M.SG.OBL=ACC/DAT ball.F give-CONJ hit.PFV.M.SG
'Having given the ball to the boy, the girl hit him.'

There are, however, constructions in which *de* 'give' makes (what looks like) a light contribution without following the expected agreement pattern for monoclausal structures. For example, in (45) and (46) the perfective verbs show masculine agreement indicating that the feminine *ch^{h}arii* 'stick' and *kitaab* 'book' are not matrix arguments.

(45) miiraa=ne raam=ko **ch^{h}aṛii** **de** b^{h}agaayaa
Mira.F=ERG Ram.M=ACC/DAT stick.F give chase.PFV.M.SG
'Mira chased Ram away using a stick.'

(46) miiraa=ne raam=ko **kitaab** **de** maaraa
Mira.F=ERG Ram.M=ACC/DAT book.F give hit.PFV.M.SG
'Mira hit Ram using a book.'

We speculate that in these cases *de* 'give' forms an adverbial participle that provides information on how the action described by the finite verb is accomplished. Similarly, in (47) the perfective verb *rokii* 'stop' agrees with the object of stopping *lift* 'lift', while *haat^{h} de* is a "means" participle that provides information on how the agent stopped the lift.

(47) miraa=ne lift **haat^{h}** **de** rokii
Mira.F=ERG lift.F hand.M give stop.PFV.F.SG
'Mira stopped the lift using (her) hand.'

Scrambling tests provide further evidence that we need to differentiate between the "means" participle use of *de* 'give' and "reverse" complex predication. In (48) the finite verb agrees with the object *ch^{h}aṛii* 'stick' which can move freely in the clause, as illustrated in (48a-b). The ungrammatical examples in (48c-d) show that the verb root *de* cannot be fronted away from the lexical verb *maarii* 'hit'.

(48) **"Reverse" complex predicate: *de* + V**

a. miiraa=ne raam=ko *ch^{h}aṛii* **de** **maarii**
Mira=ERG Ram=ACC/DAT stick.F give hit.PFV.F.SG
'Mira hit Ram with a stick.'

b. miiraa=ne *ch^{h}aṛii* raam=ko **de** **maarii**
c. *miiraa=ne *ch^{h}aṛii* **de** raam=ko **maarii**
d. *miiraa=ne **de** raam=ko *ch^{h}aṛii* **maarii**

When *de* forms a "means" participle, however, the situation is different. In (49) the finite verb does not show agreement with *ch^{h}aṛii* 'stick'. The adverbial participle *ch^{h}aṛii de*

moves as a unit, as shown in (49a-b), and it cannot be separated as indicated with the ungrammaticality of (49c-d).

(49) **"Means" participle: N + *de***

a. miiraa=ne raam=ko **chʰaṛii de** maaraa
Mira=ERG Ram=ACC/DAT stick.F give hit.PFV.M.SG
'Mira hit Ram using a stick.'

b. miiraa=ne **chʰaṛii de** raam=ko maaraa
c. *miiraa=ne **chʰaṛii** raam=ko **de** maaraa
d. *miiraa=ne **de** raam=ko **chʰaṛii** maaraa

We speculate that in (45-47) and (49a-b) we might be observing early stages of the development of *de* 'give' towards an instrumental postposition, though this is not listed as a common grammaticalization path of 'give' in the World Lexicon of Grammaticalization (Heine & Kuteva 2002). However, as these constructions are not the subject of this paper we leave them to one side for future research and concentrate on the structures that do show a monoclausal agreement pattern.

A further question that remains for future work is whether there are argument structure restrictions associated with the unusual *de* + V ordering. For example, in the standardly ordered ACP construction in (50a-b) the realisation of *kitaab* 'book' is not obligatory. When *de* 'give', however, precedes the lexical verb in the "reverse" construction, dropping *kitaab* is infelicitous, as shown in (51a-b).

(50) **Standard ACP: VMAINVLIGHT**

a. miiraa=ne raam=ko kitaab **maar dii**
Mira.F=ERG Ram.M=ACC/DAT book.F hit give.PFV.F.SG
'Mira hit Ram with a book.'

b. miiraa=ne raam=ko **maar diyaa**
Mira.F=ERG Ram.M=ACC/DAT hit give.PFV.M.SG
'Mira hit Ram.'

(51) **"Reverse" construction: VLIGHTVMAIN**

a. miiraa=ne raam=ko kitaab **de maarii**
Mira.F=ERG Ram.M=ACC/DAT book.F give hit.PFV.F.SG
'Mira hit Ram with a book (deliberately).'

b. #miiraa=ne raam=ko **de maaraa**
Mira.F=ERG Ram.M=ACC/DAT give hit.PFV.M.SG
'Mira hit Ram.'

In "reverse" constructions with *de* 'give' we have identified so far, there is consistently an unmarked argument, as we saw in (51a) (the exception is unergative verbs; see Section 3.3). While we point to this observation, we are aware that a much larger study of possible verb-verb combinations is needed to confirm if this holds empirically. If this is indeed the case, then that would mean that the *de* + V order requires an internal (theme) argument

slot. This contrasts with the standard ACP construction in which the clause-final light verb 'give' takes the event argument as its theme (see Section 2.2 on Butt & Geuder's (2001) proposal). We leave the question open as more work is needed to identify what selectional restrictions drive possible *de* + V combinations. We proceed to discuss interpretive effects associated with the unusual ordering.

3.3. Interpretive effects

The placement of the verb root *de* 'give' before the lexically dominant verb comes with very clear interpretive effects such as intentionality, forcefulness and suddeness which are not observed in the standard ACP ordering. Examples are given below with the transitive verbs *kʰilaa* 'feed' (52), *paṭak* 'slam' (53), *pʰẽk* 'throw' (54), *giraa* 'make fall, knock down' (55), *gʰuseṛ* 'push into' (56) and the unergative *chʰĩ̃k* 'sneeze' (57).

(52) mãã=ne bachche=ko rotii **de** **kʰilaayii**
mother.F=ERG child.M.OBL.SG=ACC/DAT bread.F give feed.PFV.F.SG
'The mother (forcefully) fed the child bread.'

(53) laṛke=ne kitaab zamiin=par **de** **paṭkii**
boy.M.OBL.SG=ERG book.F floor=on give slam.PFV.F.SG
'The boy slammed the book to the floor.'

(54) us=ne apnii maalaa jʰiil=mẽ **de** **pʰẽkii**
3SG.OBL=ERG own.F necklace.F lake=in give throw.PFV.F.SG
'He/She$_i$ threw his/her$_i$ necklace into the lake.' (adapted from Hook 1974: 34)

(55) raam=ne apnii kitaab jʰiil=mẽ **de** **giraaii**
Ram.M=ERG own.F book.F lake=in give make.fall.PFV.F.SG
'Ram$_i$ dropped (intentionally) his$_i$ book in the lake.'

(56) us=ne takiyaa=mẽ chaaku **de** **gʰuseṛaa**
3SG.OBL=ERG pillow.F =in knife.M give push.into.PFV.M.SG
'He/She pushed the knife into the pillow.'

(57) laṛkii (raam=par) **de** **chʰĩ̃kii**
girl.F Ram=on give sneeze.PFV.F.SG
'The girl sneezed (on Ram) (intentionally).'

For all constructions in (52-57) a standardly ordered ($V_{MAIN}V_{LIGHT}$) counterpart in which the main verb precedes the light verb can be constructed. However, the standard ACPs will make a more neutral assertion and will lack the intensive readings associated with the *de* + V order. For example, when the standard ACP in (58a) is used there is no indication whether the agent hit the patient with the book on purpose or by accident. The "reverse" order in (58b), however, reads as an intentional and more forceful act (examples are repeated from (50a) and (51a)).

(58) a. **Standard ACP: VMAINVLIGHT**
miiraa=ne raam=ko kitaab **maar dii**
Mira.F=ERG Ram.M=ACC/DAT book.F hit give.PFV.F.SG
'Mira hit Ram with a book.'

b. **"Reverse" construction: VLIGHTVMAIN**
miiraa=ne raam=ko kitaab **de maarii**
Mira.F=ERG Ram.M=ACC/DAT book.F give hit.PFV.F.SG
'Mira hit Ram with a book (deliberately).'

In other cases, the "reverse" order serves to emphasise the agent's unusual actions. For example, in (59) Santa's delivering of the presents before Christmas is unusual and worthy of emphasis and the order in (60) directs the hearer's attention towards Mira's out-of-the-ordinary actions.

(59) sãĩṭaa=ne krismas=se pehle saare tohfe
Santa.M=ERG Christmas=INSTR before all.PL present.M.PL

de pahũchaaye
give deliver.PFV.M.PL
'Santa delivered all presents before Christmas.'

(60) miiraa=ne daavat=ka saaraa-kaa-saaraa k^{h}aanaa akele hii
Mira.F=ERG party=GEN.M all-GEN.M-all food.M alone EMPH

de baanaayaa
give make.PFV.M.SG
'Mira made the entire food for the party alone (to prove something/out of anger).'

The interpretive effects that arise with *de* + V sequences are diverse and tricky to trace but all seem to revolve around the expression of an action that is significant in some way, e.g., it is forceful, deliberate, sudden or unexpected/unusual. While we leave the technicalities for future work, we believe that these readings are linked to the inherent directed-action nature of *de* 'give' in interaction with lexical information from the main verb and the (extra-linguistic) context. The effect of the placement of *de* 'give' before the verb that dominates the event description is one of emphasis on the "outwardly directed-action" aspect of the event, highlighting and drawing attention to the action performed by the agent. This, of course, begs for unusual circumstances and the "reverse" constructions are expectedly pragmatically marked, with what is significant/unusual about the described action being determined contextually. In the next section we turn to the light verb *jaa* 'go'.

4. The light verb *jaa* 'go'

We first give a brief overview of the uses of *jaa* 'go' as a clause-final light verb in standard ACP constructions (Section 4.1) before discussing the "reverse" *jaa* + V sequences (Section 4.2) and their interpretation (Section 4.3).

4.1. Standard ACPs with *jaa* 'go': overview

In its full lexical meaning the directed-motion verb *jaa* 'go' expresses an entity's physical movement/transfer from one location to another, as shown in (61).

(61) laṛkaa skul **gayaa**
boy.M school go.PFV.M.SG
'The boy went to school.'

As a light verb, 'go' surfaces mostly with unaccusative verbs that describe a change of state or location/position, as well as with transitive main verbs to highlight having reached an event's endpoint.[10] In its light verb use, 'go' retains the directional motion aspect of its meaning, albeit in a metaphorical sense. For example, in (62) and (63) the light verb 'go' expresses temporal motion from one state to another. 'Go' highlights the subject's transition to the state described by the main verbs *so* 'sleep' and *ṭuuṭ* 'break' from one temporal reference point to another.

(62) bachche **so** **gaye**
child.M.PL sleep go.PFV.M.PL
'The children fell asleep.'

(63) guldastaa **ṭuuṭ** **gayaa**
vase.M break go.PFV.M.SG
'The vase broke.'

Similarly, (64) shows that the light verb 'go' does not encode a movement/transfer towards a physical location as there is no clash between the meaning of the main verb *aa* 'come' and the light verb *jaa* 'go'. While the main verb *aa* 'come' encodes space-bound movement, the light verb 'go' further emphasises reaching the endpoint of the movement event. With other motion verbs, however, 'go' adds a sense of direction; in (65) it indicates movement "away" from the deictic centre.

(64) pulis **aa** **gayii**
police.F come go.PFV.F.SG
'The police have come.'

(65) kabuutar **uṛ** **gayaa**
pigeon.M fly go.PFV.M.SG
'The pigeon flew away.'

With transitive verbs, as in (66)-(68), 'go' again reinforces that the event is completed in full and exhaustively. For example, (68) implies that the tigers devoured all three cows in full (the event of eating has reached its natural endpoint; see also Singh 1994; 1998).

(66) vo saarii kitaab (ek baar=mẽ) **paṛʰ** **gayaa**
3SG whole.F book.F one time=in read go.PFV.M.SG
'He/she read the whole book (in one go).'

[10] See Butt and Ramchand (2005) who argue that light verbs lead to an achievement or accomplishment reading.

(67) mãĩ sab kuch^{h} **sik^{h} gayaa**
1SG everything learn go.PFV.SG.M
'I learned everything.'

(68) is varsh b^{h}ii ch^{h}e-janvarii=ko ye baag^{h} tiin
this.OBL year also six-January=ACC/DAT these tiger.M three

gaayõ=ko maar-kar **k^{h}aa gaye**
cow.F.OBL=ACC/DAT hit/kill-CONJ eat go.PFV.M.PL
'This year also on the sixth of January these tigers killed and ate three cows.'
(*lit.* 'having killed three cows, ate (them) up') (Emille Hindi Corpus; ehinweb147)

In summary, in its standard ACP use the light verb *jaa* 'go' retains its directional semantics and combines with both unaccusative and transitive main verbs to express temporal motion towards an event's natural endpoint, leading to readings of exhaustiveness and completion. With other motion-related verbs, however, 'go' forms a type of directional construction adding an "away" reading to a motion event. Next, we turn to discuss the more unusual ordering in which *jaa* 'go' precedes the lexically dominant verb.

4.2. "Reverse" ordering: *jaa* + V sequences

As already discussed in section 2.1, *jaa* + V sequences are in general described in previous work as "reverse", semantically equivalent variants of standard ACPs (Hook 1973; Poornima 2012; Das 2015). However, the "reversal" is by no means a productive process which casts doubts on whether *jaa* + V constructions should be thought of as "reverse" variants of standard ACPs or are a completely different beast.

First, not any standard $V_{MAIN}V_{LIGHT}$ complex predicate can "reverse" and not any $V_{LIGHT}V_{MAIN}$ sequence we have identified can "reverse back" to a standard ACP. For example, (69a) shows a *jaa* + V sequence with the animate subject *Ramesh.* Having the standard ACP order, as in (69b), where the main verb *kho* 'lose' precedes the light verb 'go', yields an unaccusative structure with the inanimate *baṭuaa* 'wallet' as the subject. The examples in (70) show that the "reverse" construction in (70a) does not have a readily available standard counterpart; the order in (70b) describes a temporal sequence of actions.

(69) a. ramesh apnaa baṭuaa **jaa k^{h}oyaa**
Ramesh.M own.M.SG wallet.M go lose.PFV.M.SG
'Ramesh$_i$ lost his$_i$ wallet.'

b. ramesh=kaa baṭuaa **k^{h}o gayaa**
Ramesh=GEN.M wallet.M lose go.PFV.M.SG
'Ramesh's wallet got lost.'

(70) a. haat^{h}ii gusse=mẽ diivaar **jaa toṛaa**
elephant.M anger=in wall.M go break.PFV.M.SG
'The elephant broke the wall in his anger.'

b. haat^{h}ii gusse=mẽ diivaar **toṛ gayaa**
elephant.M anger=in wall.M break go.PFV.M.SG
'Having broken the wall in his anger, the elephant left.'

Second, there is (at least in some cases) a certain degree of grey area as to whether *jaa* + V sequences are interpreted as describing a single event or a temporal sequence of actions. For example, Hook (1973) describes (71) as a case of a "reverse" complex predicate (though he uses the term "compound verb"). However, this example could also be understood to describe a temporal sequence in which the arrow went for a while and then it fell.

(71) mãĩ=ne apne d^{h}anush=se tiir chalaayaa
1SG=ERG own.OBL bow=INSTR arrow.M make.move.PFV.M.SG

tab vo ek miil duur **jaa giraa**
then 3SG one mile far go fall.PFV.M.SG
'When I shot the arrow from my bow it carried for a mile.' (Hook 1973: 24)

The initial placement of *jaa* 'go' often comes with a strong sense of directional motion, especially with other motion-related verbs and locative expressions. In (72)-(74), the verb root *jaa* 'go' seems to express motion in space towards the explicit locations.

(72) gend gaḍḍ^{h}e=mẽ **jaa girii**
ball.F ditch=in go fall.PFV.F.SG
'The ball fell into the ditch.' (adapted from Das 2015: 184)

(73) gend naalii=mẽ **jaa luṛ^{h}kii**
ball.F drain=in go roll.PFV.F.SG
'The ball rolled into the drain.' (adapted from Das 2015: 185)

(74) chaakuu raam=ke peṭ=mẽ **jaa g^{h}usaa**
knife.M Ram.M=GEN.M.OBL stomach=in go enter.PFV.M.SG
'The knife entered Ram's stomach.'

The ungrammaticality of (75b), compared to the standard ACP in (75a), also suggests that the early placement of *jaa* 'go' carries a sense of outbound spatial movement which clashes with the meaning of *aa* 'come'.

(75) a. pulis **aa** **gayii**
police.F come go.PFV.F.SG
'The police have come.'

b. *pulis **jaa** **aayii**
police.F go come.PFV.F.SG
'The police have come.'

With other verbs, however, it is much clearer that *jaa* 'go' may not necessarily describe physical movement. This is the case in (69a) and (76-77) where 'go' does not describe motion in space; instead, the unusual ordering of the verb root *jaa* 'go' before the lexical verb leads to a more marked interpretation (to be discussed in Section 4.3).

(76) raam=ne gusse=mẽ merii g^{h}aṛii **jaa** **toṛii**
Ram.M=ERG anger=in POSS.1SG.F watch.F.SG go break.PFV.F.SG
'Ram broke my watch in anger.'

(77) laṛkii=ne seb **jaa** **k^{h}aayaa**
girl.F=ERG apple.M go eat.PFV.M.SG
'The girl ate the apple.'

This short overview has shown that *jaa* 'go' in *jaa* + V sequences may express motion in space but, as we saw in (69a) and (76-77), need not do so. As we will see in the next section, we argue that both uses are linked to the inherent semantics of *jaa* 'go': in the first case it expresses motion in the physical sense and, in the second, in a metaphorical sense to serve pragmatic functions.

4.3. Interpretive effects

Ross (2016) shows that cross-linguistically morphemes and verbs such as 'go' that indicate a direction away from a deictic centre are often involved in a grammaticalization pattern to express mirativity[11] (see DeLancey 1997; DeLancey 2012). As already hinted in Section 2, this seems to be the case with Hindi *jaa* + V sequences, resembling the English *go and* construction said to express surprise (Vos 2004; Ross 2016). The initial placement of *jaa* 'go' in (78a), for example, leads to a marked interpretation; the example could be uttered to describe an event of falling that is perceived to be in some way surprising or unexpected (e.g., *Ram* was not careful enough and fell despite being warned). Such extra dimensions of meaning are not present with the standard ACP in (78b) where the light verb 'go' indicates the complete transition to the state described by the main verb *gir* 'fall'.

(78) a. raam gaḍḍ^{h}e=mẽ **jaa** **giraa**
Ram.M ditch=in go fall.PFV.SG.M
'Ram fell in a ditch.'

[11] The term mirativity is broadly used to describe utterances that a speaker uses to express their surprise at some unexpected state, event, or activity (see DeLancey 1997; DeLancey 2012).

b. raam gaḍḍʰe=mẽ **gir** **gayaa**
Ram.M ditch=in fall go.PFV.SG.M
'Ram fell in a ditch.'

Jaa + V sequences may also be used to express an action done with determination, as in (79) (an 'occupied' reading as opposed to 'sat on' the chair), or an action done in control as in (80) ('throwing oneself' as opposed to 'falling' accidentally). *Jaa* + V sequences may also be used to express disapproval: (81) could be uttered in a context in which the mother should not have read the letter (e.g., it was not meant for her) but she went ahead and read it anyway.

(79) raaj kursi=pe **jaa** **baiṭʰaa**
Raj.M chair=on go sit.PFV.M.SG
'Raj occupied the chair.'

(80) naukar malik=ke kadamõ=mẽ **jaa** **giraa**
servant.M master.M=GEN.OBL feet=in go fall.PFV.M.SG
'The servant threw himself at the feet of the master.' (Das 2015: 11)

(81) maa=ne apnii beṭii=kii chiṭṭʰii **jaa** **paḍʰii**
mother.F=ERG own.F daughter.F=GEN.F letter.F go read.PFV.F.SG
'The mother$_i$ read her$_i$ daughter's letter.'

The readings we have described (disapproval, surprise, deliberateness/determination, control) are diverse and context-dependent but, we believe, are linked to the semantics of directed-motion *jaa* 'go'. The verb root *jaa* 'go' in *jaa* + V sequences expresses the subject's motion towards the completion of an action (as described by the lexical verb), albeit in a metaphorical sense when no physical movement reading is present. Ross (2016) makes an intriguing argument which is relevant here: constructions such as the English *go and* involve the extension of 'go' to express deviation from an expected outcome (motion away from expectation). Along similar lines, Hindi *jaa* + V sequences could be understood as drawing attention and highlighting the subject's motion towards an outcome that is perceived to be in some way unusual or significant (as per the speaker's expectations and world knowledge).

6. Conclusion

In this paper we explored the use of directed-motion *jaa* 'go' and directed-action *de* 'give' in standardly ordered complex predicates and in the so-called "reverse" constructions. We suggested that describing *jaa* + V and *de* + V sequences as "reverse" variants of standard complex predicates might be misleading as there are significant interpretive differences between the two orderings. Given the widely observed grammaticalization of 'go' on a cross-linguistic basis to express mirative readings, it seems plausible that at least the *jaa* + V construction in Hindi has developed independently driven by the directional semantics of *jaa* 'go'.

Our central argument has been that the placement of *de* 'give' and *jaa* 'go' before the lexically more dominant verb leads to interpretive effects which can be traced back to their directional semantics. We have argued that the early placement of *de* 'give' draws

attention to the agent's actions; informally, *de* + V sequences express an agent-initiated action that is unusual or significant in some way. The early placement of *jaa* 'go' highlights the subject's (metaphorical) motion towards an unusual or in some way significant outcome. In this paper, however, we have only started scratching the surface when it comes to semantic aspects of *de* + V and *jaa* + V constructions. We hope, nonetheless, to have shown that these constructions are worth exploring as they can shed further light on verbal stem meaning and processes of complex predicate and event construal. Much more work also remains to be done on the argument and constituent structure of these constructions, as well as their use in context.

References

Abbi, Anvita & Devi Gopalakrishnan. 1991. Semantics of explicator compound verbs in South Asian languages. *Language Sciences* 13(2). 161–180.

Bhatt, Rajesh & Elena Anagnostopoulou. 1996. Object shift and specificity: Evidence from ko-phrases in Hindi. In Lise M. Dobrin, Kora Singer & Lisa McNair (eds.), *Proceedings of CLS 32*, 11–22. Chicago: Chicago Linguistic Society.

Butt, Miriam. 1993. Object specificity and agreement in Hindi/Urdu. In Katharine Beals, Gina Cooke, David Kathman, Sotaro Kita, Karl-Erik McCullough & David Testen (eds.), *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*, vol. 1, 89–103. Chicago: Chicago Linguistic Society.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu*. Cambridge: Cambridge University Press.

Butt, Miriam. 1997. Interfaces as locus of historical change. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG97 Conference*. University of California, San Diego: CSLI Publications.

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Brett Baker, Mark Harvey & Mengistu Amberber (eds.), *Complex predicates: Cross-linguistic perspectives on event structure*, 48–78. Cambridge: Cambridge University Press.

Butt, Miriam. 2017. Hindi/Urdu and related Languages. In Jessica Coon, Diane Massam & Lisa Demena Travis (eds.), *The Oxford Handbook of Ergativity*, 807–831. Oxford: Oxford University Press.

Butt, Miriam & Wilhelm Geuder. 2001. On the (semi)lexical status of light verbs. In Corver Norbert & Riemsdijk Henk van (eds.), *Semi-lexical categories: On the content of function words and the function of content words*, 323–370. Berlin: De Gruyter Mouton.

Butt, Miriam & Wilhelm Geuder. 2003. Light verbs in Urdu and grammaticalization. In Eckardt Regine, Heusinger Klaus von & Schwarze Christoph (eds.), *Words in time*, 295–350. Berlin: De Gruyter Mouton.

Butt, Miriam & Aditi Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135. 7–29.

Butt, Miriam & Gillian Ramchand. 2005. Complex aspectual structure in Hindi/Urdu. In Nomi Ertischik-Shir & Tova Rapoport (eds.), *The syntax of aspect*, 117–153. Oxford: Oxford University Press.

Das, Pradeep Kumar. 2015. Can compound verb be reversed in Hindi? An inquiry of its form and function. *Interdisciplinary Journal of Linguistics* 8. 37–60.

DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1. 33–52.

DeLancey, Scott. 2012. Still mirative after all these years. *Linguistic Typology* 16. 529–564.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. New York: Cambridge University Press.

Hook, Peter Edwin. 1973. *The compound verb in Hindi*. University of Pennsylvania PhD dissertation.

Hook, Peter Edwin. 1991. The emergence of perfective aspect in Indo-Aryan languages. In Elizabeth Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. II, 59–89. Amsterdam: John Benjamins Publishing Company.

Hook, Peter Edwin. 1993. Aspectogenesis and the compound verb in Indo-Aryan. In Mahindra K. Verma (ed.), *Complex predicates in South Asian languages*, 97–114. New Delhi: Manohar.

Kachru, Yamuna. 2006. *Hindi*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. University of Chicago Press.

Ozarkar, Renuka & Gillian Ramchand. 2018. Structure matching and structure building in Marathi complex predicates. *Journal of South Asian Linguistics* 8(1).

Parsons, Terence. 1990. *Events in the semantics of English: A study in subatomic semantics*. MIT Press.

Poornima, Shakthi. 2012. *Hindi complex predicates at the syntax-semantics interface*. University at Buffalo, State University of New York PhD dissertation. https://arts-sciences.buffalo.edu/content/dam/arts-sciences/linguistics/AlumniDissertations/Poornima%20dissertation.pdf (11 April, 2021).

Poornima, Shakthi & Jean-Pierre Koenig. 2009. Hindi aspectual complex predicates. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, 276–296. Stanford, CA: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2009/poornima-koenig.pdf (11 April, 2021).

Ross, Daniel. 2016. Going to surprise: The grammaticalization of itive as mirative. In *Cognitive Linguistics in Wroclaw Web Conference*.

Singh, Mona. 1994. *Perfectivity, definiteness, and specificity: A classification of verbal predicates in Hindi*. University of Texas at Austin PhD dissertation.

Singh, Mona. 1998. On the semantics of the perfective aspect. *Natural Language Semantics* 6(2). 171–199.

Slade, Benjamin. 2021. Development of verb-verb complexes in Indo-Aryan. In *Verb-verb complexes in Asian languages*. Oxford: Oxford University Press.

Vos, Mark de. 2004. Pseudo coordination is not subordination. In Leonie Cornips & Jenny Doetjes (eds.), *Linguistics in the Netherlands*, vol. 21, 181–192. John Benjamins Publishing Company.

SOAS Working Papers in Linguistics. Volume 20 (2021): 95–120

Temporal information in sentence-final particles: *Tse* and *keh* in Modern Shanghai Wu

Yan JIANG and Wen WANG
SOAS University of London

yj9@soas.ac.uk

Abstract
This paper scrutinizes the claim that modern Shanghainese has sentence-final particles *tse* and *keh* that have tense-marking functions. We review works by Qian (2006; 2009), Chao (1926) and Li, Thompson & Thompson (1982) and analyse Shanghainese missionary texts on the use of these SFPs. Through a functional-discoursal investigation, we identify the IN-cluster use and the END-cluster use of *tse*. We take the temporal marking function of *tse* as a consequence of its discourse function, which introduces a current reference time in the discourse. On the other hand, we take *keh* as an assertion particle, whose occasional sense of recent past comes from its confirmation of a completed event.

Keywords: tense, aspect, sentence-final particle, Shanghainese, missionary linguistics

1. Modern Shanghai Wu

Modern Shanghai Wu is a language that used to be spoken in the urban region of Shanghai and its surrounding suburban areas as recorded in written works by missionaries in the years from early 1840s to early 1950s. The language is standardly referred to as Shanghai vernacular (上海土白) in those earlier works. In contrast, contemporary Shanghai Wu denotes the Shanghai Wu used since the 1950s and is subdivided into three variants: the old (老派), the middle (中派), and the new (新派). The old variant overlaps a lot with modern Shanghai Wu and is used in the suburban areas with a declining number of speakers. The middle variant is now commonly referred to as the standard variant, used by the middle-aged and senior people in the city. The new variant, mostly used by young people in the urban area, has some distinct phonetic and lexical features but overlaps a lot with the middle variant and is often code-mixed with Mandarin.[1]

Modern Shanghai Wu turns out to be the most recorded in published form, accumulating a rich archive of biblical writings and translations, textbooks and grammar books, magazines and informal releases such as church pamphlets. Writings in modern Shanghai Wu far exceed writings published in any other modern non-Mandarin Chinese languages. However, since the mid-1950s, publishing in non-Mandarin Chinese or in non-character graphic forms other than the officially recognized romanized Mandarin system (Pinyin) was forbidden by the Chinese government in the form of statutory regulations. As a result, Shanghai Wu writings have not been in circulation in written form for a long time, with

[1] Ideas in this paper were presented at Symposium on Early Chinese Dialect Grammar at Hong Kong University of Science and Technology in June 2019 and at Online SOAS-Humboldt Univ. Wu Colloquium in June 2021. Thanks go to the organizers of the conferences, to the reviewers for helpful comments, and to the editors of *SOAS Working Papers* for their patient guidance and superb editorial work.

the recent exceptions that *Xinmin Evening Paper* launched a special weekly column in the language around 2005 called "Shanghai Language" (上海閒話). Modern Shanghai Wu scripts have become a rarity, especially because the books in the library room which used to house a copy of each and every missionary Shanghainese publication in Shanghai were burnt in the Cultural Revolution. Fortunately, due to recent efforts in digitization of rare library collections in overseas libraries, it is now possible for interested linguists to study these works first-hand, so as to learn more about the grammatical system of modern Shanghai Wu and contrast it with its contemporary variants.

2. Encoding time in modern Shanghai Wu: some recent claims

It has been argued by Qian (2004; 2006; 2009; 2010; 2011; 2014) and Zhu, Cao & Qian (2017) that while Mandarin does not have tense markers, nor does contemporary Shanghai Wu, modern Shanghai Wu, on the other hand, does have markers for tense, which can be sentence-final particles (henceforth SFP). One of them is *tse* 哉, which expresses time current to the time of speaking. Another one is *keh* 個, which marks recent past.

Qian (2006) observes that *tse* is used to indicate the state of an event or action concurrent with the temporal information denoted by the time adverbial, as shown in (1). However, in the absence of such an adverbial, *tse* points to present by default, as shown in (2)-(3). Qian (2006) does not entertain the possibility of a sentence inheriting temporal information from the context, which would have affected his claim presented here.[2]

(1) 伊話上禮拜寫個哉。

yi	wo	zong	lipa	xia	keh	tse
3SG	say	last	week	write	PST	CUR

'S/He said it had been written last week.'
(Macgowan 1862: 52, cited in Qian 2006: 948)

(2) 自來水龍頭壞脫者。[3]

zylesy	lungdeu	wa-theh	tse
running.water	tap	break-COMPL	CUR

'The water tap is broken.' (Obata 1908: 散語 70, cited in Qian 2016: 39)

(3) 表停者，勿曉得有幾點鐘者。

pioa	ding	tse,	veh	xioateh	yeh	ji-ti	tsung	tse
watch	stop	CUR,	not	know	have	how-CLF	hour	CUR

'The watch has stopped. Don't know what hour it is.'
(Obata 1908: 問答 34, cited in Qian 2016: 39)

[2] Shanghainese pinyin and glossing ours. See the appendix for the correspondence between the Shanghainese romanization adopted here and the IPA. Tones in Shanghainese are not represented here, as they tend to vary in actual phrases due to tone sandhi. All abbreviations in the glosses follow the Leipzig Glossing Rules with the addition of AM 'assertion marker', CEP 'cause and effect particle', CM 'complement marker', CP 'compound pronoun', CUR 'current state marker', HON 'honorific', MOD 'modifier marker', PM 'pause marker', RQM 'rhetorical question marker', SFP 'sentence-final particle' and SP 'status-marking particle'.

[3] 哉 and 者 are used interchangeably when used as an SFP.

According to Qian (2006), *keh* denotes recent past, as shown in (4)-(7).

(4) 儂是幾時搬到地頭來個?

nung	zy	jizy	poe-toa	diehdeu	le	keh
2SG	be	when	move-to	here	come	PST

'When did you move to here?' (Obata 1908: 問答 53, cited in Qian 2016: 39)

(5) 伊是幾時轉來個?

yi	zy	jizy	tsoele	keh
3SG	be	when	return	PST

'When did s/he come back?' (Obata 1908: 問答 59, cited in Qian 2016: 40)

(6) 我侬一直替那一家人家交易,

nguni	yehzeh	dah	na	yeh-ka	ningka	jioani,
1PL	always	with	2PL	one-CLF	household	trade

勿到別人家去個。

veh-toa	biehningka	qi-keh
not-go	others	go-PST

'We always trade with your firm alone and never have dealings with others.' (Obata 1908: 散語 49, cited in Qian 2016: 40)

(7) 公債股票咯甚賣買從前都做個,

kungtsa	kuphio	gesang	mama	zungzi	ze	tsu	keh,
bonds	shares	like.that	buy-sell	before	all	do	PST,

現在勿做者。

yize	voh	tsu	tse
now	not	do	CUR

'We/I used to buy and sell government bonds and shares, among others. Now we don't do it anymore.' (Obata 1908: 散語 72, cited in Qian 2016: 40)

Qian takes a sentence with a sentence-final tense marker to be a tensed sentence, while he takes one containing an aspectual marker to be an aspectual sentence. According to him, tense and aspectual markers occurring in one and the same sentence in modern Shanghainese can take the form of a tense-aspectual compound at the end of the sentence (Qian 2006; 2009; 2010), as shown in (8)-(10).

(8) 耶穌看見伊拉個相信哚,對癱子話:

Yasu	khouji	yila-heh	xiangxing	loa,	te	thetsy	wo
Jesus	see	3PL-POSS	belief	CEP	to	paralytic	say

"兒子呀,放心,儂個罪饒赦哉" … …

nitsy	ah,	fongxing,	nung-heh	ze	nioaseh	tse
son	PM	be.at.ease	2SG-POSS	sin	forgive	CUR

話儂個罪饒赦**拉哉**… … 。
wo nung-heh ze nioaseh la tse
say 2SG-POSS sin forgive COMPL CUR
'[and when] Jesus saw their faith, he said to the paralytic, "Take heart, my son; your sins are hereby forgiven." … to say that your sin has been forgiven' (American Presbyterian Mission Press 1895: Matthew 9:2-5, cited in Qian 2006: 950) [4]

(9) 到之末，看見房子空拉，
toa-tsy meh, khoeji vongtsy khung la,
arrive-PRF PM see house vacant DUR

打掃咾裝潢**拉哉**。
tangsoa loa tsongwong la tse
sweep and decorate COMPL CUR
'And when he returned, he found it empty, swept, and put in order.' (American Presbyterian Mission Press 1895: Matthew 12:44, cited in Qian 2006: 950) [5]

(10) 伊個人腳那能蹺**拉個**？
yi-geh-ning jiah naneng qioa la keh
3SG-that-person foot why lame DUR PST

毛病呢還是生成功**拉個**？
moabing neh hezy sangzengkung la keh
illness PM or born.with DUR PST
'Why was that person walking with difficulties? Was that an accident or was it inborn?' (Obata 1908: 散語 Chapter 7, cited in Qian 2006: 956)

In addition, a sentence may contain temporal adverbials such as time-denoting nouns or aspect-denoting adverbs. (11) and (12) are the mixed cases.

(11) 儂個令堂也已經交關難過拉哉。
nung-heh lingdong ha yijing jioakue neku la tse
2SG-POSS mother.HON also already very sad DUR CUR
'Your respected mother has already been very sad.'
(Bourgeois 1939: 72, cited in Qian 2009: 64)

[4] While Qian (2006: 950) refers to the American Bible Society (1923) when citing (8) [his example (25)], he only cites the first part of the example, which does not contain 拉 *la* , i.e. the underlined part (but *la* did appear in his example). Our view is that the perfective marker 拉 *la* should not appear in the first part because it is a speech act that takes place on the scene. On the other hand, it makes sense for 拉 *la* to appear in the second part of the example because it is an indirect speech referring to a completed event.

[5] Note that Qian (2006: 950) mistakes it (his example (26)) for a sentence from Chapter 13. Qian's literal translation is 到了以後，看見房子空著，就打掃和裝潢好了 ('He arrived and found the house vacant, so he swept and put it in order'), which is very different from ours.

(12) 現在天熱哉，要用紗祭披。

yize	thi	nieh	tse,	yoa	yung	sojiphi
now	weather	hot	CUR	should	use	thin.chasuble

‘Now it is getting hot. You should put on a thin chasuble.’
(Bourgeois 1939: 81, cited in Qian 2009: 193)

According to Qian (2006; 2009), tense/aspectual markers, especially tense-aspectual compounds, had gradually given way to the more frequent use of time adverbials as Shanghai Wu developed from its modern version to its present contemporary version.

3. Two related studies

Qian relates his characterization of Shanghainese tense to two predecessors: Chao (1926) and Li, Thompson & Thompson (1982). Chao (1926), while comparing Peking Mandarin with the Wu languages of Suzhou and Changzhou, takes the Mandarin SFP *le* (了/勒/啦/喇) to express, not the past meaning, but the present or future in two specific senses. First, it expresses inchoative meaning in indicating a new happening or registering a new realization. Second, it expresses perfect meaning, on the completion of an event or action. Chao remarks that sometimes an action can be viewed either at its initiated stage or at its stage of completion. So the inchoative reading and the perfect reading could be related. The equivalent of the Mandarin *le* in Suzhou Wu is *tse*, which is similar to *tse* in modern Shanghainese. Chao also mentions another use of *tse*, which he takes to be narrative past, as shown by (13) - (15).

Suzhou Wu[6]

(13) 商量得嘸不結果，大家就轉氣哉。

saonlian	teh	mpeh	cihkou,	daka	zeu	tsoechi	tse
discuss	COMP	not.have	result	everyone	therefore	return	PST

‘Not reaching anything in the discussion, we all went back home.’
(Chao 1926: 881; Chao 1992: 93)

(14) 一歇歇俚又出去哉。

ihshihshih	li	yeu	tshehchi	tse
a.short.while	SP	again	go.out	PST

‘After a short while, ∅ went out again.’ (Chao 1926: 881; Chao 1992: 93)

(15) 後來我就氣睏哉。

gheule	ngou	zeu	chi	khuen	tse
later	1SG	then	go	sleep	PST

‘Later, I went to bed.’ (Chao 1926: 881; Chao 1992: 93)

Chao takes the above three examples to be similar to the French *passé défini* (also called *passé simple*), as shown in (16). He further observes that in sentences where the French *imparfait* (as shown by (17)) should be used instead of *passé défini*, the equivalent sentences in Chinese should not take on *le* (for Mandarin) or *tse* (for Suzhou Wu). As far as we can see, (13) – (15) saliently point to past rather than present. In (13), the first clause

[6] The romanized pinyin system used to represent Suzhou Wu is promulgated by the Wu Language Association: http://wu-chinese.com/romanization/suzhou.html.

provides the contextual information related to a past situation, and both (14) and (15) contain aspect-denoting adverbs which point to past situations. If a sentence in French is in *imparfait*, the aspectual feature is imperfective, and a sentence in *passé défini* should carry the perfective aspectual feature. The relevant examples are given as (16) and (17). It can be inferred from the above discussion that sentences carrying *le* or *tse* carry perfective aspectual meaning.

(16) **French (*passé simple*)**

Je	remplaç-ai	le	comédien	malade
1SG	replace-PST.1SG	the	actor	sick

'I replaced the sick actor.' (Heminway 2016: 131)

(17) **French (*imparfait*)**

Je	sav-ais	qu'il	avait	raison
1SG	know-IPFV.1SG	that-he	have-IPFV.3SG	reason

'I knew he was right.' (Heminway 2016: 67)

As to the equivalent of *keh* in Suzhou Wu, represented here as *kah*, Chao (1926:45) takes it to indicate past and thinks that sentences containing it would be rendered into English in simple past. He also observes a similarity in temporal meaning between *kah* and *laizhe* (來著) in Beijing Mandarin.

Suzhou Wu

(18) 我今朝看見王先生格。

ngou	cingtsau	khoecie	Waon	siesan	kah
1SG	today	see	Wang	mister	PST

'I saw Mr. Wang today.' (Chao 1926: 909; Chao 1992: 106)

Beijing Mandarin

(19) 我今天看見王先生來著。

wo	jintian	kanjian	Wang	xiansheng	laizhe
1SG	today	see	Wang	mister	PST

'I saw Mr. Wang today.' (Chao 1926: 909; Chao 1992: 106)

Li, Thompson & Thompson (1982) provide a functional analysis of the Mandarin SFP *le*, i.e. le_2,[7] whose usage is comparable to the Shanghainese *tse*, as pointed out by Qian (2006). According to Li, Thompson & Thompson (1982: 22), le_2 is to signal a 'currently relevant state' (CRS), indicating that a state of affairs has special current relevance to some particular reference time which, in the unmarked case, is the conversational setting in which the speaker and hearer are participating as interlocutors. However, le_2 will not be used in the following contexts: in talking about simple general truths, on-going states, past and future events, and commands and requests, where no currently relevant state is involved. Nevertheless, they note one exception: le_2 can be found in written narrative about past eventuality, where it serves to bring a state of affairs into the time at which the narrative is taking place. According to them, such a use is entirely analogous to its use in conversation except that the reference point is the narrative time rather than the speech

[7] le_1 is generally characterized as a verb suffix.

time. In terms of data observation on past narrative, Chao (1926) and Li, Thompson & Thompson (1982) hold similar views.

4. Conceptual scrutiny

Qian (2006; 2009; 2014) can be said to have taken over the views of Chao (1926) and Li, Thompson & Thompson (1982) in their analyses of Chinese languages other than Shanghainese and goes a step further in claiming that in modern Shanghai Wu, *tse* and *keh* are tense markers. However, several theoretical issues need to be addressed before claims about Shanghainese can make sense against the current literature on tense and temporality.

Klein (2009) discusses six types of devices that are regularly used to encode time in language: (a) tense, (b) aspect, (c) Aktionsart, (d) temporal adverbials, (e) temporal particles, and (f) discourse principles. Chinese Han languages are known to make no use of (a) but use all of the rest. Some SFPs are likely to be examples of (e), if they have time-encoding properties. But Klein (2009) only mentions particles in Mandarin "which can [immediately] follow or [in one case] precede the verb", i.e. verb-endings *le*, *guo*, *zhe* and the pre-verbal *zai*, where the *le* is the verb-suffix le_1. Chao (1926) is a study of auxiliary words as realizations of mood and modality.[8] Auxiliary words in traditional Chinese grammar are functional words or words with no conceptual meaning, which can be either verb suffixes or SFPs, or pre-verbal functional words. Chao (1926) takes some uses of *tse* and *keh* to be clearly time-indicating. However, although translations to English and French with equivalent tense and aspectual properties are suggested, no theoretical characterization is given in Chao's paper. Li, Thompson & Thompson (1982) try to show that le_2 as an attitudinal SFP in Mandarin performs many of the functions of the perfect aspect. It can be inferred that the authors take le_2 to encode aspectual meaning. However, Qian (2006) endeavours to show that *tse* as an equivalent of le_2 in modern Shanghainese is a tense marker rather than an aspect marker, as it can co-occur with an aspect marker or aspect-denoting adverbs. Moreover, according to him, when a sentence with *tse* does not contain any aspect markers, *tse* acts as the indicator of tense, while aspectual meaning should be gleaned from the context or from relevant adverbs (Qian 2006: 947).

If *tse* is characterized as a marker for present tense, the issue is that it is not an obligatory marker for present, nor does it only encode present meaning, as it can also be used in past narrative sentences. If it is taken as indicating current state relative to another provided reference time, then it cannot be absolute tense, contrary to Qian's claims, as absolute tense must provide a clear index on the time axis. The fact that *tse* can co-occur with *keh* as shown in (1) also indicates that it is inappropriate to treat *tse* as a marker of present, since it would result in time conflicts with *keh*, which is characterized as a marker of recent past tense in Qian's works. If, on the other hand, *tse* is a relative tense marker denoting current state, then it is aspectual in nature, as relative tense is realized through aspectual features. If *tse* is to be treated as an aspect-marker of current state, in contrast with prior or posterior temporal meaning, then it is still hard to explain why it co-occurs with other aspectual markers, as shown in (8) and (9), as the resulting aspectual compound will form two unrelated aspectual meanings. In this connection, the parallelism drawn by

[8] Chao (1926) uses a more traditional notion in Chinese grammar called *kouqi* (口氣) which we propose to translate into *utterance tone*.

Qian (2006) between what he called the tense-aspect compound (or rather, aspect-tense compound) and the coalesced tense/aspect verb suffix in English is not justifiable. In the case of (8) and (9), an SFP is adjacent to an element which is either a verb-suffixing particle or another SFP, whereas in English, the tense/aspect ending is a single, portmanteau morph with a regular conjugation paradigm. Little similarity exists between the case in Shanghainese and that in English. A similar line of reasoning can be provided to cast doubt on treating *keh* as an absolute tense marker of recent past.

Li, Thompson & Thompson (1982) propose to treat *le*$_2$ as expressing perfect meaning, i.e. breaking news, furnishing information of present relevance, and asserting a current state. If the Shanghainese *tse* is identical to *tse* in Suzhou Wu and to *le*$_2$ in Beijing Mandarin, which needs to be established through data examination, this characterization does not give rise to conceptual disharmony. Yet it is necessary to examine the role of *tse* in the overall picture of tense and aspect marking in modern Shanghainese. On the other hand, the function of *keh* seems to be partially assumed by *le*$_2$ and marginally taken by *laizhe* as in (19).

Turning now to the use of *tse* in written narrative discourse of past happenings, its equivalent in Suzhou Wu is likened to *passé défini* by Chao (1926), whose meaning is distinct from *imparfait*. Its equivalent in English would be the historical present, or dramatic present, or narrative present, so should be distinct from the simple past or past progressive. This also relates to Chao's characterization of *kah* in Suzhou Wu, which he thinks expresses a meaning like the simple past in English. So *tse* and *kah* have distinct meanings and distributions in Suzhou Wu. The function of *le*$_2$ in past narrative, according to Li, Thompson & Thompson (1982), is that it brings a state of affairs into the narrated time. In other words, it highlights the time at which the narrative is taking place. All these remarks are revealing, but they are presented in the works of Chao (1926) and Li et al. (1982) with single sentences, and, in fact, about languages other than modern Shanghainese.

Qian (2006; 2009; 2010; 2014) does provide examples in modern Shanghai Wu, but they do not show a complete picture of the uses of *tse* and *keh*, because his examples are mostly single sentences taken from textbooks, which cannot lend us a good view of *tse* in narrative past, since it is not possible to know from single-sentence examples under what conditions each sentence in narrative past discourse needs to take on *tse*. Textbook examples of *keh* look very neat, often arranged to contrast with sentences containing other kinds of temporal meaning, but it is not clear to what extent *keh* is present as an SFP in discourse. Studying discourse data will not only help us solve these puzzles but will also reveal new facts. We will treat *tse* first in the next section and will get to *keh* in Section 6.

5. *Tse* in missionary texts

Missionary Shanghainese works have been consulted in order to obtain first-hand information on the use and distribution of SFPs. The scripts consulted are listed in the following table with *tse*'s occurrences.[9]

[9] The bibliographical details of works listed in Table 1 are documented in Shi (2017), with the exception of item 5, which is given in the references as Ali Paipai Feng Dao Ji (1921). Our

Table 1: *tse* in missionary texts

	File /year	Sentence (S)	Character	SFP tse 者	SFP tse 哉	tse/S Ratio	IN	END
1	亨利實錄 1856	1351	13455	72	0	0.053	45	24
2	趁早預備 1868	158	1302	0	4	0.025	4	0
3	剛担丟士 1868	221	1815	0	2	0.009	2	0
4	審判日腳 1868	92	794	0	5	0.054	4	1
5	阿里排排逢盜記 1921	2318	23515	166	1	0.072	167	16
6	十條誡總意問答 1899	238	2488	0	2	0.008	1	1
7	聖經史記 1892	6397	54226	187	32	0.034	198	24
8	耶穌言行傳 1894	5248	37726	0	222	0.042	207	15
9	使徒言行傳 1890	1882	15663	0	59	0.031	53	6
10	路加傳福音書 1886	4591	39289	0	182	0.040	174	8
11	馬太傳福音書 1895	4292	36573	0	168	0.039	156	12
	Total	26788	226846	425	677	0.041	1011	107

Scanned copies of the scripts in Table 1 were acquired from libraries or the internet, which we input into Word so that they became searchable documents. Searching *tse* either in the form of 哉 or as 者 in these files returned initial figures, from which we deleted occurrences of *tse* that are not sentence-final. Non-SFP *tse* involves the use of 者 'person' as another lexeme. *Tse* is not used as a verb suffix, even though it may happen to appear after a non-transitive verb at the end of a sentence.

5.1 Some general observations about *tse*

Judging by the ratio between the number of *tse* occurring in a script and the number of sentences contained therein (the tse/S ratio), occurrence of *tse* in most scripts is conspicuous, even though it cannot be described as frequent. Although there are variations from script to script, the longer the script, the more similar the tse/S ratios become. This conspicuous use of *tse* in modern Shanghainese contrasts sharply with its zero return from a search for the SFP use of 哉 and 者 in a database containing contemporary Shanghainese writings, which we are collecting from the Shanghainese Page (上海閒話) published weekly by Xinmin Evening Paper (新民晚報) and Shanghainese blog writings in Douban Net (豆瓣網). This result conforms to the judgment of native speakers of contemporary Shanghai Wu, who tend to think that *tse* is not in use in the middle and new variants of the contemporary version, although it is still in use in the old variant in some areas. On the other hand, searching in the missionary Shanghainese database for *le* (了) only yielded one return, which is a verb suffix, not an SFP.[10] But searching for *le* (了) in our database of contemporary Shanghai Wu returns numerous occurrences, both as verb suffixes and as SFPs. These findings provide support to the feeling shared by most language users of Shanghainese that its contemporary

references also include items 1 and 7 in Table 1 as we will discuss them in detail. The searchable database we constructed out of the digitized image files accounts for more than fifty percent of known Shanghainese missionary scripts.

[10] The example is "後首亞當又養了一个兒子", from 《聖經史記》Shengjing Shiji (Old Testament History).

version has lost much of its linguistic uniqueness due to the infiltration of Mandarin as the standard Han language.

5.2 Two uses of the SFP *tse*

We discuss the use of *tse* with the findings from 亨利實錄 *Hengli shilu* (The history of Little Henry and his bearer) as a sample, generalizing to the other texts where applicable. From the text, 72 occurrences of the sentence-final *tse* written as 者 are found. No occurrence of *tse* written as 哉 is found. A sample paragraph and its gloss are given as (20):

(20) a. 有一个小囝,
yeu yeh-geh xioanoe
there.be one-CLF child
'There was a child.'

b. 名頭叫亨利,
mingdeu jioa Hengli
name call Henry
'His name is called Henry.'

c. 養拉亭南怕戶蕩,
yang la Dingnoepho wudong
be.born at Dinapore place
'He was born at the town of Dinapore.'

d. 就是印度國裏。
zieuzy Yingdu koh li
that.is India country in
'That is in the country of India.'

e. 伊个爺,是一个英咭唎个武官,
yi-heh ya, zy yeh-geh Yingjili-heh wukuoe
3SG-POSS dad be one-CLF England-POSS military.officer
'His dad was an English military officer.'

f. 亨利養得勿多幾時,
Hengli yang teh voh-tu-jizy
Henry born COMP not-much-time
'Not long after Henry's birth.'

g. 伊个爺去攻打炮台哱,
yi-heh ya, qi kungtang phoade loa
3SG-POSS dad go attack fort CEP
'As his father was sent to attack a fort,'

h. 撥拉對敵殺脫个,
pehla tedieh sah-theh keh
pass enemy kill-COMPL PST
'was killed by the enemy.'

i. 亨利勿曾滿歲末,
Hengli vohzeng moe se meh
Henry not.yet reach year PM
'When Henry was not yet one year old,'

j. 伊个娘又死者,
yi-heh niang yeu xi tse
3SG-POSS mother also die CUR
'his mother also died.'

k. 蓋哱亨利拉小个辰光,
kehloa Hengli la xioa-heh zengkuong
that.is.why Henry at little-POSS time
'That's why when Henry was little,'

l. 就無得子爺娘哱,
zieu meh-teh-tsy yaniang loa
therefore not-have-PFV father.mother CEP
'so he lost both his father and his mother.'

m. 最苦惱只那。
tse khunoa tseh-na
most miserable CUR-SFP
'He was very miserable.' (Hengli Shilu 1856)

Tse in (20j) appears in the middle of a paragraph, presenting a situation or a state, which may be initiated earlier but has current relevance to the narration, followed by a cohesive word *kehloa* (蓋哱) 'that's why', which introduces discourse continuity. So current relevance is found to be the temporal contribution *tse* makes. To be more specific, we propose that *tse* as an SFP not only presents a proposition as a situation or the beginning of a situation when involving an inchoative verb, but also introduces a reference time. This latter function is similar to a time adverbial or a subordinate clause, which can also introduce reference time,[11] irrespective of whether there are other expressions that introduce reference time in the same sentence or whether the reference time can be inferred from the context, as is common in Chinese.

Two other discourse segments are presented below as (21) and (22). In (21), there are three occurrences of *tse*: (21f), (21h), and (21j). (21f) with its use of *tse* introduces a new situation which leads to a consequence described by (21g)-(21h), while (21h) with its own

[11] Cf. Hofmann (Hofmann 1993, Chapter 7).

use of *tse* serves to present the result. (21j) with the third *tse* is a further result in response to a situation introduced by (21i). (22h) serves the same function as (21i).

(21) a. 亨利望進來,
Hengli mang jingle
Henry look inside
'Henry looked in.'

b. 看見子畫哱要想走進去,
khoeji-tsy wo loa yoa-xiang tseu jingqi
see-COMPL painting CEP want-wish go inside
'He saw the picture and wished to enter the room.'

c. 但是勿搭蒲師一淘末,
dezy veh tah Busy yehdoa meh
but without with Boosy together PM
'but without the company of his carer Boosy,'

d. 勿敢進去,
veh koe jingqi
not dare enter
'[he] did not dare to go in'

e. 歇子一歇,
xieh-tsy yehxieh
pause-COMPL a.while
'moments later'

f. 搭子姑娘兩个熟子點者,
tahtsy kuniang liang-keh zoh-tsy ti tse
with lady two-CLF familiar-COMPL bit CUR
'getting more familiar with the lady'

g. 就是蒲師拉門外頭,
zieuzy Busy la meng-ngadeu
even.though Boosy at door-outside
'even though Boosy is not with him in the room'

h. 伊也敢進去白相者。
yi ha koe jingqi behxiang tse
1SG also dare enter play CUR
'he also dared to go in and play.'

i. 後首末一眼勿怕哱,
heuseu meh yehnge veh pho loa
later PM not.least not afraid CEP
'Later, he was not afraid at all.'

j. 常常走進去者。
zangzang tseu jingqi tse
often go enter CUR
‘[He] would often go into the [lady’s] room.’

k. 搭之姑娘兩个,
tahtsy kuniang liang-keh
with lady two-CL
‘together with the lady,’

l. 親熱晔快活得極。
qingnieh loa khuaweh-teh djieh
intimate CEP happy-CM extreme
‘He never was more happy than when he was with this lady.’
(Hengli Shilu 1856)

(22) a. 難末對亨利話,
nemeh te Hengli wo
therefore to Henry say
‘Therefore, [she] said to Henry:’

b. 第个神勿能彀救自家,
dieh-keh zeng veh nengkeu jieu zyka
this-CL god not can save self
‘This god cannot save himself.’

c. 那能救別人呢,
naneng jieu biehning neh
how save other Q
‘How can he save others?’

d. 叫伊勿能彀响,
jioa yi veh nengkeu xiang
call him not able.to talk
‘If you talk to him, he is unable to respond.’

e. 又勿能彀動。
yeu veh nengkeu dung
also not able.to move
‘Nor can he take action.’

f. 伊歇辰光,
yi-xieh zengkuong
that-CL time
‘During that period of time,’

g. 亨利聽得子姑娘个講究末，
Hengli thing-tehtsy kuniang-keh kongjieu meh
Henry hear-COMPL lady-POSS argument PM
'Henry listened to the lady's arguments,'

h. 救相信子第个道理者。
zieu xiangxing tsy dieh-keh doali tse
so believe COMPL this-CL doctrine CUR
'so he was convinced by her arguments.' (Hengli Shilu 1856)

We went through all the 1100-odd sentences containing *tse* from the eleven works listed in Table 1 and found that they basically fall into two types, as have been exemplified by the discourse segments (20), (21) and (22): those that present a situation with immediate follow-ups, and those that appear at the end of a concluding sentence.

5.3 Sentence cluster and the discourse function of *tse*

To sharpen the findings reported in 5.2, we propose to use a descriptive concept, sentence cluster (句群), for our current analysis. Sentence cluster, as proposed by Wu & Tian (2000), can be defined as (23):

(23) Sentence cluster: two or more sentences forming a group of sentences that are related to one another with cohesive devices and coherent content structure.

A sentence cluster may contain complex and compound sentences as its members. In addition to using logical connectives and sentential adverbials as cohesive devices, it also makes use of SFPs, discourse deixis and implicit reasoning to establish coherence. Within a sentence cluster, we can observe rhetorical relations as expounded in Rhetorical Structure Theory (RST) (Mann & Thompson 1988). We can also observe structural properties that are often identified in Chinese discourse, such as topic chain, paratactic construction, and multiple use of zero anaphora across discourse. All these are devices that can tie a sentence cluster together.

We can now conclude that *tse* has two uses: first, it is used within a sentence cluster to bring out a situation, to be followed by some closely related consequential extension. Second, it is used at the end of a sentence cluster to present a conclusion, a summary remark, or an end to a stretch of description.

In Table 1, the IN column records the occurrences of *tse* within a sentence cluster, while the END column records the occurrences of *tse* at the end of a cluster. It can be observed that the IN-cluster use far exceeds the END-cluster use. The IN-cluster use of *tse* mainly marks a change of state. The END-cluster use indicates the speaker's total contribution up to the point, especially when nothing follows the cluster.

Overall, both uses of *tse* contribute to the discourse organization by introducing a reference time which is concurrent to the speech time. The speech time of a sentence is by default set to be present, but in a stretch of discourse presented in narrative past the speech time can also be shifted to past in order to achieve a vivid narrative effect: the past portrayed as present, or historical present. Thus, *tse* can also relate an event to the

concurrent time of speech which is set in the past. It can be observed that the SFP *tse* only appears in some sentences, not in many others, as not every sentence needs to emphasize the current time.

Although we think *tse* resembles *le₂* in terms of some discourse functions, the two expressions are by no means equivalent, as they each belong to a different language. A full understanding of the use of *tse* is dependent on the description of the other SFPs in modern Shanghainese, such as the alleged recent past marker *keh*, among others. The findings on the use of *tse* cover both narrative use and conversational use, as they both occur in the scripts we consulted. Biblical texts appear to be narrative. Yet they contain many cases of direct speech. In both narrative and conversational texts, the temporal contribution of *tse* is constant, which is the introduction of current time, leaving the exact choice of tense information to other factors in the discourse.

6. *Keh* in missionary texts

6.1 The multiple functions of SFPs

The above study on *tse* reveals a common feature of SFPs, that they often involve several senses or functions. In an earlier paper, Qian (1996) gives the following description of *leh* 了 in contemporary Shanghai Wu, noting that its older form is *tse* 哉:

(24) *leh* 了 (old form *tse* 哉)[12]
- (a) narrating past event
- (b) narrating the process of an event
- (c) narrating future event
- (d) stating result of a hypothesis or condition
- (e) stating result of a reasoning
- (f) stating result in contrast to the usual expectation
- (g) clarifying
- (h) persuading, consulting
- (i) reporting current state

While *leh* in contemporary Shanghai Wu may no longer be the same as *tse* in modern Shanghai Wu,[13] it should also be acknowledged that discourse analyses conducted by different researchers may split out different types of discourse meaning, depending on how much one takes the contextual information into consideration. Since we only wish to give candid data description in this paper, we will not explore the possibility of rationalizing the strands of meaning of an SFP within a given theoretical framework. However, data description may tell us to what extent each sense of an SFP is present in discourse. In the case of *tse*, we have shown through the discussion above that its temporal meaning is always available, which is accompanied by its other, IN-cluster or END-cluster discoursal functions.

[12] 了 [舊：哉]：(1) 敘事過去 (2) 敘事過程 (3) 敘事將來 (4) 假設、條件的結果(5) 直推的結果 (6) 轉折的結果 (7) 申明，表白 (8) 勸聽，商量 (9) 現狀 (Qian 1996).

[13] As one evidence, Qian (2006) has argued that the temporal meaning of *tse* has been weakened, if not completely lost, in contemporary Shanghai Wu.

6.2 The senses of *keh*

Qian (1996; 1997) gives the following description of *keh* in contemporary Shanghai Wu.

(25) keh 個 [14]
- (a) asserting, affirming
- (b) clarifying
- (c) reminding, warning
- (d) forbidding
- (e) stating the result of concession
- (f) reporting a recent past event

(25) relates back to the cursory discussion of *keh* as encoding past time in modern Shanghai Wu in Sections 2 and 3, and to examples (4)-(7) and (10). Our question here is: how exactly is the sense of (25f) expressed in modern Shanghai Wu? How does (25f) stand in relation to the other senses? Does it co-occur with some other senses, or is it used to the exclusion of the others? The answers are no to be found in Qian's works, as his discussion of the use of (25f) is supported by a neat set of well-chosen single-sentence examples, as shown in Section 2, that would not let one see the use of *keh* in real discourse.

We look at our missionary database to see what functions *keh* can take on. *Keh*'s occurrences in the database are listed in Table 2. (*，＋*。means *keh*-sentences ending with a comma plus those with a full stop.) The scripts are all in traditional characters. Yet they use either 個 or 个 for *keh* but not both. With the only exception of 聖經史記 (Shengjing Shiji 1892), which uses both 個 and 个 for the SFP *keh*. It is also this work that uses both 哉 and 者 to represent the SFP *tse*.

Table 2: Distribution of *keh*

	File /year	Sentence	Character	SFP keh 個 (*，＋*。)	SFP keh 个 (*，＋*。)
1	亨利實錄 1856	1351	13455	0	74+45
2	趁早預備 1868	158	1302	5	0
3	剛担丟士 1868	221	1815	9+3	0
4	審判日腳 1868	92	794	5	
5	阿里排排逢盜記 1921	2318	23515		70+117
6	十條誡總意問答 1899	238	2488	3+4	
7	聖經史記 1892	6397	54226	41+5	144+25
8	耶穌言行傳 1894	5248	37726	0	293+106
9	使徒言行傳 1890	1882	15663	104+8	
10	路加傳福音書 1886	4591	39289		197+154
11	馬太傳福音書 1895	4292	36573	0	232+144
	Total	26788	226846		

[14] 個：(1) 確實，肯定 (2) 申明，表白 (3) 提醒，警告 (4) 禁止 (5) 讓步

An examination of the sentences containing the SFP *keh* from 亨利實錄 *Hengli shilu* (The history of Little Henry and his bearer) shows a function of *keh* which is not related to the temporal meaning of recent past event:

(26) 第个話頭末，**全拉聖書上个**。

dieh-geh	wodeu	meh,	djyuoe	la	sengsy	zong	keh
this-CL	topic	PM	all	at	sacred.book	on	AM

‘This topic can all be found in the sacred book.’ (Hengli Shilu 1856)

(27) 又造个可怕个地獄，勿肯懊悔罪孽个人，**應該住拉化个**。

yeu	zoa	keh	khupho-keh	dinioh,	
also	build	CL	horrible-MOD	hell	
veh	kheng	oahue	zenieh	keh	ning,
not	willing	repent	sin	MOD	person
yingke	zy	lawo	keh		
must	live	there	AM		

‘And the dreadful hell was also built. Those who are unwilling to repent for their sins should live there.’ (Hengli Shilu 1856)

(28) 比方儂个娘，

pifong	nung-keh	niang,
for.example	2SG-POSS	mother

叫儂到伊房裏去，

jioa	nung	toa	yi	vong-li	qi,
ask	2SG	to	3SG	room-in	go

做一樣事體，

tsu	yehyang	zythi,
do	one-CL	thing

儂倒勿肯做，

nung	toa	veh	kheng	tsu,
2SG	however	not	willing	do

娘豈勿要動氣个否。

niang	qiveh	yoa	dungqi	keh	feu
mother	RQM	will	feel.angry	AM	not

亨利話，是也，**終要動氣个**。

Hengli	wo,	zy	ya,	tsung	yoa	dungqi	keh
Henry	say	be	PM,	finally	will	feel-angry	AM

‘“If your mama were to desire you to come into her room to do something for her, and you were to refuse, would she not be displeased with you?” Henry said, “Yes, she would be angry in the end.”’ (Hengli Shilu 1856)

(29) 姑娘又話,
kuniang yeu wo,
lady again say

儂或是叫蒲師來對儂扇扇,
nung wehzy jioa Busy le te nung soe-soe,
2SG perhaps ask Boosy come for 2SG fan-fan

或是抬轎,
wehzy de djioa
perhaps carry wagon

伊倒勿肯做,
yi toa veh kheng tsu,
3SG however not willing do

或是叫伊做第樣,
wehzy jioa yi tsu dieh-yang,
perhaps ask 3SG do this-type

伊倒做子伊樣,
yi toa tsu-tsy yi-yang,
3SG however do-COMPL that-type

或是叫伊抱儂到伊塊去,
wehzy jioa yi boa nung toa yi-khue qi
perhaps ask 3SG carry 2SG to that-place go

伊到抱子到別塊去,
yi toa boa-tsy toa bieh-khue qi
3SG however carry-ASP to still.other-place go

伊豈勿差否。
yi qiveh ca feu
3SG RQM unsatisfactory not

亨利話, **是算差个**。
Hengli wo, zy soe ca keh.
Henry say, be count.as unsatisfactory AM
'The Lady continued, "Or, if you ask Boosy to fan you, or to carry you in your palanquin, and Boosy refuses. Or you ask him to do this, and he does something quite different; or if you desire him to carry you one way, and he carries you another: would he not be unsatisfactory?" Henry said, "He would not be satisfactory, to be sure."' (Hengli Shilu 1856)

(30) 蓋啤我看起來，**第種人勿好个**。

kehloa	ngu	khoe-qile,	dieh-tsung	ning	veh	hoa	keh
that.is.why	1SG	see-up	this-sort	person	not	good	AM

‘That is why I do not see these people as being virtuous.’ (Hengli Shilu 1856)

(26) reports a lasting state which is not specifically related to past or present. (27) is a modal assertion stating a subjective requirement. (28) is about a conditional consequence. (29) involves the emphatic assertion construction “*zy … keh*”. (30) is a judgment. The common factor that can be observed is the assertion function of *keh*.

But the following sentences containing *keh* do report past events. (31) is about some past event that recurred but is not necessarily recent. (32) is a past event followed by other narrative content so can be considered as recent past. (33) is recent habitual past state. (34) is an assertion about some recent past action. In (35), the factual conditional states the relationship between a possible action performed in the recent past and its consequence.

(31) 除子蒲師末，

zytsy	Busy	meh
except	Boosy	PM

亨利無得別人當心伊，

Hengli	meh-the	biehning	tongxing	yi
Henry	not-possess	other.person	care.for	3SG

格啤有个人曉得

kehloa	yeuheh	ning	xioateh
therefore	some.PL	person	know

蒲師待亨利實蓋能个軟心腸啤，

Busy	de	Hengli	zehkehnengkeh	nyuoe	xingtsang	loa
Boosy	treat	Henry	so.very	tender	hearted	SFP

稱讚伊个好處，

cengtsoe	yi	keh	hoacy
praise	3SG	MOD	virtue

三不時，人末送點物事拉伊个。

sepehzy	**ning-meh**	**sung**	**ti**	**mehzy**	**la**	**yi**	**keh**
occasionally	person-PM	give	some	thing	to	3SG	PST

‘Except for Boosy, Henry had no one else to take care of him. That is why some people noticed the very kindness of Boosy to the child, and he got presents from many people.’ (Hengli Shilu 1856)

(32) 亨利病重格辰光,

Hengli	bing	zung	keh	zengkuong
Henry	illness	worsen	MOD	time

蒲師常庄望伊哞,

Busy	zangtsong	mang	yi	loa
Boosy	often	inspect	3SG	SFP

勿肯到牀上去睏,

veh	kheng	doa	zanglang	qi	kung
not	willing	to	bed.on	go	sleep

直奉承到病好子哞放心个。

zeh	vengzeng	doa	bing	hoa	tsy	loa	fongxing	keh
till	look.after	to	illness	good	ASP	PM	be.at.ease	PST

'When Henry was seriously ill, Boosy often inspected his situation and would not go to sleep. Only until Henry recovered did he stop worrying.' (Hengli Shilu 1856)

(33) 自伊自家對自家話,

zyyi	zyka	te	zyka	wo
CP.3SG[15]	self	to	self	say

伊拉平素日腳,

yila	**bingsu**	**niehjiah**
3PL	ordinary	time

窮苦來非凡个,

djiongkhu	**le**	**five**	**keh**
poor	duration	extraordinary	PST

現在竟然要用傢生來量金子,

yize	jingzoe	yoa	yung	kasang	le	liang	jingzi
now	even	need	use	utensil	for	weigh	gold

伊拉个暴富,

yila-keh	boafu
3PL-MOD	become.suddenly.rich

[15] CP denotes a compound pronoun in an obsolescent pronoun system in Shanghai Wu (a similar system still exists in some Wu languages now). A compound pronoun is the usual pronoun prefixed by *zy* 'self' yet has no reflexive meaning. It arguably can be used in an emphatic way but not necessarily, as some speakers prefer to stick to compound pronouns.

不言可知个者。
peh-yi-khu-tsy keh tse
can.tell.without.saying PST CUR
'He told himself, they were normally extremely poor. Now they even need to use utensils to measure gold. Their sudden rise to richness is self-evident.'
(Ali Paipai Feng Dao Ji 1921)

(34) 阿里排排勿等伊个女人个說話話完,
Ah-li-ba-ba veh teng yi-keh nyuning keh sohwo wo woe
Ali.Baba not wait 3SG-CL woman MOD words say finish

就阻擋自伊哚話,
zieu zudang zyyi loa wo
then stop CP.3SG PM say

自儂勿必緊嚇得,
zynung veh pieh jinghah tah
CP.1SG not must feel-afraid SFP

我一定勿至於做盜賊个。
ngu yehding veh tsyyu tsu doazeh keh
1SG certainly not end.up act gangster PST
'Without waiting for her to finish, Ali Baba interrupted and said, "Don't you be afraid! There is no way I became a robber."' (Ali Paipai Feng Dao Ji 1921)

(35) 自儂已經得著拉个金子銀子,
zynung yijing teh-tseh-la-keh jingzi ningzi
CP.2SG already get-COMPL-COMPL-PST gold silver

我也勿要者。
ngu ha veh yoa tse
1SG also not want CUR

自儂若使敢有一句說話哄騙自我个,
zynung zohsy koe yeu yeh-jyu sohwo hungphi zyngu keh
CP.2SG if dare have one-CL word cheat CP-1SG PST

我就要到官府宕去告自儂,
ngu zie yoa doa koefoadang qi koa zynung
1SG then will to official.place go report CP.2SG

話自儂是窩賊。
wo zynung zy wotsang.
say CP.2SG be help.hide.loot
"You have got your gold and silver, which I don't want. If you dared to cheat me on anything, I will turn you in at the government office for abetting the gangsters in keeping the loot." (Ali Paipai Feng Dao Ji 1921)

Overall, *keh* is not necessarily present in all statements about recent past. In fact, its presence is rather limited. In such cases, it involves the reporting of a completed event, which is a form of non-emphatic assertion. So the temporal meaning of recent past is more of an inferred sense rather than an encoded sense for *keh*. It is parasitic on the assertive use of *keh*.

7. **Concluding remarks**

To conclude, SFPs in Shanghainese as exhibited by *tse* and *keh* do not encode absolute tense information themselves. *Tse*'s function is to convey current relevance of a sentence it is attached to, and *keh*'s function is to mark an assertion.

Appendix

Shanghainese alphabetic system and its correspondence with the IPA phonemes (Tang 2000).

上海話拼音系統與國際音標對應表

（1）consonants 聲母（27）

Shanghainese Pinyin	IPA	Example
p	[p]	爸爸 papa 5 21 'father'
ph	[p']	泡泡 poapoa 5 21 'bubble'
b	[b]	婆婆 bubu 13 'grandma'
m	[m]	媽媽 mama 5 21 'mother'
f	[f]	發福 fah foh 3 4 'stout'
v	[v]	俯伏 vuvoh 1 3 'prostrate'
t	[t]	爹爹 tiatia 5 21 'dad'
th	[t']	太太 thatha 3 4/5 21 'wife'
d	[d]	洞洞 dungdung 1 3 'hole'
n	[n]	囡囡 noenoe 5 21 'baby'
l	[l]	玲瓏 linglung 1 3 'dainty'
ts	[ʦ]	珠珠 tsytsy 5 21 'bead'
c	[ʦ']	拆穿 cahcoe 3 4 'expose'
s	[s]	伸縮 sengsoh 5 21 'flexible'
z	[z]	暫時 zezy 1 3 'tentative'
j	[ʨ]	姐姐 jiajia 3 4 'sister'
q	[ʨ']	親戚 qingqieh 5 2 'relatives'
dj	[ʥ]	齊全 djidjyuoe 1 3 'complete'
ni	[ȵ]	肉牛 niohnieu 1 3 'meat cattle'

ny	[ȵ]	女人 nyuning 1 3 ‘woman’
x	[ɕ]	休息 xieuxieh 5 2 ‘rest’
zi	[ʑ]	謝謝 ziazia 1 3 ‘thank’
k	[k]	哥哥 kuku 5 21 ‘brother’
kh	[k‘]	刻苦 khehkhu 3 4 ‘diligent’
g	[g]	共用 gungyung 1 3 ‘shared’
ng	[ŋ]	硬臥 ngangngu 1 3 ‘hard berth’
h	[h]	好貨 hoahu 3 4 ‘good stuff’
h	[ɦ]	鞋盒 hahah 1 3 ‘shoebox’

(2) vowels 韵母（43）

Shanghainese pinyin	IPA	Example
y	[ɿ]	支持 tsyzy 5 21
a	[A]	喇叭 laba 1 4
o	[o]	火車 hoco 3 4
oa	[ɔ]	報告 poakoa 3 4
eu	[ɣ]	後頭 heudeu 1 3
e	[E]	來三 lese 1 4
oe	[ø]	半盤 boeboe 3 4
i	[i]	飛機 fiji 5 21
ia	[ia]	爺爺 yaya 1 3
ioa	[iɔ]	小巧 xioaqioa 3 4
ieu	[iɣ]	九流 jieulieu 3 4
ie	[iE]	機械 jiye 5 21
u	[u]	姑父 kufu 5 21
ua	[ua]	娃娃 wawa 1 3
ue	[uE]	關懷 kuewe 5 21
uoe	[uø]	管理 kueoli 3 4
yu	[y]	語句 nyujyu 1 4
yuoe	[yø]	圓圈 yuoeqyuoe 1 4
el	[əl]	反而 fe’el 3 4
m	[m̩]	姆媽 mma 5 12
n	[n̩]	∅奶 ngna 5 21
ng	[ŋ̍]	魚頭 ngdeu 1 3
ang	[ã]	冷場 langzang 1 3
ong	[ɑ̃]	剛剛 gonggong 5 21
eng	[əŋ]	餛飩 wengdeng 1 3
ung	[oŋ]	籠統 lungthung 1 4
ah	[aʔ]	百搭 pahtah 3 4
oh	[oʔ]	北角 Pohkoh 3 4
eh	[əʔ]	勒勒 lehleh 1 3
iang	[iã]	想像 xiangxiang 3 4

iong	[iɑ̃]	手旺 seu yong 3 13
ing	[iŋ]	精靈 jingling 5 21
iung	[ioŋ]	汹涌 xiungyung 5 21
iah	[iaʔ]	吃藥 qieh yah 5 12
ioh	[ioʔ]	汰浴 dayoh 1 3
ieh	[iəʔ]	業績 niehjieh 1 4
uang	[uã]	光火 kuanghu 5 21
uong	[uɑ̃]	狂妄 guongwong 1 3
ueng	[uəŋ]	困難 khuengne 3 4
uah	[uaʔ]	挖苦 wahkhu 3 4
ueh	[uəʔ]	骨頭 kuehdeu 3 4
yun	[yn]	均匀 jyunyun 5 21
yueh	[yəʔ]	月亮 yuehliang 1 13

References

Ali Paipai Feng Dao Ji. 1921. *Ali Paipai Feng Dao Ji 阿里排排逢盜記 [Ali-Baba et les quarante voleurs]*. Chang-Hai: Imprimarie de l'Orphelinat de T'ou-se-we, Ze-Ka-Wei.

American Bible Society. 1923. *Shanghai Tubai Xinyue Quanshu 上海土白新約全書 [The New Testament, Shanghai vernacular translated version]*. Shanghai: American Bible Society.

American Presbyterian Mission Press. 1895. *Gospel of St. Matthew, Shanghai vernacular translated version*. Shanghai: American Presbyterian Mission Press.

Bourgeois, Albert 蒲君南. 1939. *Leçons sur le Dialecte de Chang-hai: Cours Moyen*. Shanghai: Imprimerie de T'ou-sè-wè.

Chao, Yuen Ren 趙元任. 1926. Beijing, Suzhou, Changzhou yuzhuci de yanjiu 北京、蘇州、常州語助詞的研究 [A study of auxiliary words in the dialects of Beijing, Suzhou and Changzhou]. *Tsinghua Xuebao 清華學報 [Journal of Tsinghua University]* 3(2). 865-917. Also published as Chao (1992).

Chao, Yuen Ren 趙元任. 1992. Beijing, Suzhou, Changzhou yuzhuci de yanjiu 北京、蘇州、常州語助詞的研究 [A study of auxiliary words in the dialects of Beijing, Suzhou and Changzhou]. *Fangyan 方言 [Dialect]* 2. 85–111.

Heminway, Annie. 2016. *Complete French grammar*. 3rd edn. New York: McGraw-Hill Education.

Hengli Shilu. 1856. *Hengli Shilu 亨利實錄 (上海土白) [The history of Little Henry and his bearer (in Shanghai vernacular)]*. 上洋 趙吟松 刊刻印訂.

Hofmann, Thomas R. 1993. *Realms of meaning: An introduction to semantics*. London: Longman.

Klein, Wolfgang. 2009. How time is encoded. In Wolfgang Klein & Ping Li (eds.), *The expression of time*, 39–81. Berlin: Mouton de Gruyter.

Li, Charles N., Sandra A. Thompson & R. McMillan Thompson. 1982. The discourse motivation for the perfect aspect: The Mandarin particle *LE*. In Paul J. Hopper (ed.), *Tense-aspect: Between semantics & pragmatics*, vol. 1, 19–44. Amsterdam: John Benjamins Publishing Company.

Macgowan, John 麥高溫. 1862. *A collection of phrases in the Shanghai dialect systematically arranged*. Shanghai: Presbytarian Mission Press.

Mann, William C. & Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse* 8(3). 243–281.

Obata, Masabumi 御幡雅文. 1908. *Huyu Bianshang 滬語便商　一名上海語 [The Shanghai dialect guide for visiting traders]*. 東京: 御幡雅文.

Qian, Nairong 錢乃榮. 1996. Shanghai fangyan de yuqi zhuci 上海方言的語氣助詞 [Mood particles in Shanghai dialect]. *Yuyan Yanjiu 語言研究 [Language Study]* 1. 34–47.

Qian, Nairong 錢乃榮. 1997. *Shanghaihua Yufa 上海話語法 [A grammar of Shanghainese]*. Shanghai People's Press.

Qian, Nairong 錢乃榮. 2004. Yige yufa cengci yanbian de shili – Shanghai fangyan 160 nian zhong xianzai wancheng shitai de xiaoshi guocheng 一個語法層次演變的實例 —— 上海方言 160 年中現在完成時態的消失過程，[A case study of grammatical change related to stratification]. *Zhongguo Yuwen 中國語文 [Chinese Language]* 3. 232–240.

Qian, Nairong 錢乃榮. 2006. Wuyu zhong shi ti jiehe de fuhe shiti 吳語中時體結合的復合時體 [Compound tense and aspect in Wu dialects]. In Dah-an Ho, H. Samuel Cheung, Wuyun Pan & Fuxiang Wu (eds.), *Shanggao Shuichang: Ting Pang-hsin Qizhi shouqing Lunwenji 山高水長：丁邦新先生七秩壽慶論文集 [Linguistic Studies in Chinese and Neighboring Languages: Festschrift in Honor of Professor Pang-hsin Ting on His 70th Birthday]* (Language and Linguistics Monograph Series No. W-6), 945–965. http://www.ling.sinica.edu.tw/LL/zh/monographs (28 August, 2021).

Qian, Nairong 錢乃榮. 2009. Shanghai fangyan de shitai jiqi liubian 上海方言的時態及其流變 [Tense and aspect in Shanghai dialect and their evolution]. *Dongfang Yuyanxue 東方語言學 [Eastern Linguistics]* 5. Also published as Qian (2016: 57-67).

Qian, Nairong 錢乃榮. 2010. Cong《Huyu Bianshang》suo jian de Shanghai hua shita i 從《滬語便商》所見的上海話時態 [Tenses and aspects? Old Shanghainese as found in the Shanghai Dialect Guide for Visiting Traders]. *Journal of Chinese Linguistics* (Diachronic Change and Language Contact — Dialects in South East China) 44–57.

Qian, Nairong 錢乃榮. 2011. Lun xiandai hanyu zhong de shitai 論現代漢語中的時態 [On tense and aspect in modern Chinese]. *Hanzangyu Xuebao 漢藏語學報 [Bulletin of Sino-Tibetan Languages]* 5. Also published as Qian (2016: 68-82).

Qian, Nairong 錢乃榮. 2014. *Xifang Chuanjiaoshi Shanghai Fangyan Zhuzuo Yanjiu: 1847-1950 Nian de Shanghaihua 西方傳教士上海方言著作研究：1847-1950 年的上海話 [A Study of Western Missionary Works on Shanghai Dialect: Shanghainese between 1847 and 1950]*. Shanghai University Press.

Qian, Nairong 錢乃榮. 2016. *Qian Nairong Yuyanxue Lunwenji 錢乃榮語言學論文集 [Linguistic Works by Qian Nairong]*. Shanghai University Press.

Shengjing Shiji. 1892. *Shengjing Shiji 聖經史記 (1-5) [Old Testament History]*. Vol. Vol. 1: 1886, Vol. 2: 1892, Vol. 3-4: 1889, Vol. 5: 1890. The American Presbyterian Mission Press.

Shi, Ye 施曄. 2017. Jindai chuanjiaoshi suo zhuan Shanghai fangyan wenxian kaoshu 近代傳教士所撰上海方言文獻考述 [A Study on Shanghai Dialect Documents Written by the Western Missionaries in Modern China]. *道風:基督教文化評論 [Logos & Pneuma: Chinese Journal of Theology]* 46. 115–156.

Tang, Zhixiang 湯志祥. 2000. *Shiyong Shanghai Hua 實用上海話 [Practical Shanghainese]*. Shanghai Education Press.

Wu, Weizhang 吳為章 & Xiaolin 田小琳 Tian. 2000. *Hanyu Juqun 漢語句群 [Chinese Sentence Clusters]*. Beijing: Commercial Press.

Zhu, Zhenmiao 朱貞淼, Weifeng Cao 曹偉鋒, & Nairong Qian 錢乃榮. 2017. Shanghai Fengxian qu Zhuanghang zhen fangyan de shitai jiqi yufahua guocheng. 上海奉賢區莊行鎮方言的時態及其語法化過程. [The tense and its grammaticalization process in the dialect of Zhuanghang Town, Fengxian District, Shanghai]. *Studies in Language and Linguistics 語言研究* 37(1). 49–53.

SOAS Working Papers in Linguistics. Volume 20 (2021): 121–141

Koronashokku: Loanwords in Japanese and the Covid-19 pandemic

Tom O'Neill

665766@soas.ac.uk

Abstract
Foreign loanwords have come to form a substantial portion of the contemporary Japanese vocabulary. While many studies have been undertaken on language borrowing, the phenomenon of global events, disasters and socio-economic movements precipitating their inception is underexplored. The purpose of this paper is to build on the prior research on well-established loanwords by examining novel usage in the specific context of the Covid-19 pandemic. Data was obtained using tweeted replies to coronavirus-related articles in order to identify which novel loanwords have come into use due to Covid-19. Subsequently, to determine comprehension rates and attitudes, a survey was carried out using 121 participants drawn from members of the Japanese public. It found a low comprehension rate of the novel loanwords, particularly amongst participants over 60 years of age. Drawing from these results, the utility of novel loanwords related to Covid-19 is called into question.

Keywords: Linguistic borrowing, English loanwords, Japanese language, comprehension, Twitter, Covid-19, language policy

1. Introduction

Throughout the 20th and 21st centuries, Japanese identity has developed a strong association with disaster. From the bombings of Hiroshima and Nagasaki in 1945, to the Fukushima Daiichi nuclear disaster in 2011, unprecedented crises profoundly influenced the Japanese nation. This is particularly evident in the arena of language as discussion of that which is wholly novel necessitates new vocabulary. In times of crisis, it is not only important to describe and report on events, but also to convey vital information to those who are affected.[1]

Perhaps the most common way for a language to create vocabulary is to borrow from a source language to produce a so-called "loanword". Indeed, this process can be triggered by disaster: one of the higher frequency Japanese loanwords in English — tsunami — is the result of linguistic borrowing (Liberman 2011). However, in more recent times, the flow of loanword creation tends to run in the other direction. Japanese words of English origin such as *oiru shokku* オイルショック (oil shock) and *baburu keizai* バブル経済 (bubble economy) came to be well known in the 1970s and 1980s respectively as the result of economic phenomena, while *raifurain* ライフライン (lifeline - essential utilities or critical infrastructure) experienced use in the advent of the Kobe earthquake (Ōnishi &

[1] I would like to thank my dissertation supervisor Dr Julia Sallabank as well as the other academic staff at SOAS from both the linguistics and Japanese departments including Dr Akiko Furukawa, Dr Seiko Harumi, Dr Hitoshi Shiraki and Dr Barbara Pizziconi for their support and advice during the writing of this paper. I would also like to show my appreciation for those people who took the time to complete my questionnaire and supply feedback, your kind cooperation made this research possible.

Kajiki 1995). Evidently, disaster, emergency, and socio-economic shocks have already been key factors in the creation of loanword vocabulary.

Japan now faces the same crisis that has engulfed the rest of the world: the Covid-19 pandemic. In much the same way that those in English-speaking domains have had to incorporate new coinages (social distancing), the revival of antiquated words (furlough), and scientific jargon (flattening the curve), the Japanese have likewise had to adapt to an influx of new terminology. Due to the singular international nature of the crisis, these terms are often foreign in origin. As such, the novel virus gives rise to novel loanwords.

The origin and derivation of loanwords (from here LWs) or *gairaigo* have attracted a wealth of academic attention (Loveday 1986; Stanlaw 2004), yet the role of major events in their creation and proliferation is underexplored. In addition, while much research has gone into investigating the comprehension of and attitudes towards high-frequency LWs in Japanese (see Ishino 1983; Ishiwata 1989; Shibata 1993; Honna 1995; NHK 2001; NINJAL 2006), LWs of novel conception are less well explored.

This study is an investigation into the phenomena of loanwords born of crisis and focusses upon the language surrounding the Covid-19 pandemic. I aim to clarify the position of novel loanwords in the Japanese language and their utility in public life whilst also shedding light on the historical interrelation of crises and linguistic borrowing.

2. What is a loanword?

Japanese, largely due to contact with various other languages, has an ever-expanding vocabulary. Foreign LWs have become so prevalent in modern Japanese that, at the turn of the century, they were estimated to constitute around 10% of the language (Hogan 2003). While LWs are common in most of the world's languages, Japanese is often held up as a notable case due to the high frequency of foreign word borrowing.

Originally, the Japanese language came from a variety of sources, and its vocabulary can be categorised into three different groups according to their source; these are:

- *wago* 和語 (words of native Japanese origin)
- *kango* 漢語 (words of Chinese origin)
- *gairaigo* 外来語 (words of foreign, non-Chinese origin)

This paper is concerned with the third of these word types: *gairaigo* (Umesao, Kindaichi & Hinohara 1989). *Gairaigo* is orthographically represented using *katakana* (one of two phonetic scripts in Japanese), though it can also be denoted using *rōmaji* (Roman letters) and is even occasionally given its own *kanji* (Chinese characters).[2]

In modern times, *gairaigo* has been dominated by English, yet this was not always the case. Indeed, several other European languages have claimed greater shares of Japanese foreign vocabulary than English in the past (Yazaki 1964). From the Spanish and Portuguese religious terminology that entered the archipelago in the Middle Ages, to the Dutch language of commerce in the 17th century that still remains in use today (kōhī 珈

[2] This is usually reserved for much older loanwords such as *tabako* 煙草 (tobacco).

琲 (coffee) and bīru ビール (beer)), borrowing patterns are not random and the distribution of LWs in the Japanese lexicon is often a reflection of Japan's relationship with other nations at different points in time. However, in the late 19th century, several other languages eventually came to supersede Dutch as the language of diplomacy and foreign relations. Chief among these was English.

When the Meiji restoration eventually brought an end to Japan's renowned period of *Sakoku* (isolationist foreign policy) in 1868, the new government also opened the doors to Western influence. While French and German were also studied at this time, borrowing from English outstripped its competitors in the post-war era, largely due to the occupation of Japan by the United States. A study of 90 Japanese magazines by the National Language Research Institute (NLRI) in 1956 revealed that around 80% of total LWs were derived from English (NLRI 1964), while more recent studies have put the proportion as high as 94.1% (Tomoda 1999). The dominance of English in borrowing is in no way unique to Japanese, yet it stands out due to its sheer ubiquity.

3. *Gairaigo* in the media

Although Japan has never gone so far as to recognise English as an official language, the presence of the world's foremost lingua franca candidate is certainly visible in the country's media culture. Certain fields such as pop music, advertising and product design provide particularly fertile soil for English to take root.

The prevalence of English and other *gairaigo* in Japanese advertising is well-established, though the reasons for this are manifold and still debated. In his seminal 1989 work *Symbolic Values of Foreign Language Use*, Harald Haarmann points to the visual power of *katakana*, a view also espoused by Rebuck (2002) who emphasises that because *gairaigo* is written using *katakana*, it draws the attention of Japanese readers who are more used to seeing the higher frequency scripts of *hiragana* and *kanji*.

Furthermore, English can carry an air of prestige (Stanlaw 1987), and so its use in advertising has the effect of giving products an exotic or sophisticated image. In her survey of loanwords in advertising, Takashi (1991) categorized LWs by function and found that "words used for special effect" were the most common. She concluded that the "special effect" of *gairaigo* was to bestow products with a fresh mood as she observed a greater proportion of LWs in advertisements for modern products than traditional ones (Takashi 1991). This aligns with Haarmann (1986) who claimed that the role of *gairaigo* in a copywriter's arsenal is actually more about conveying the desired imagery than it is about communicating information.

However, it is not only commercial texts that utilise loanwords. Along with advertising, Haarmann (1989: 65) also lists the "use of English in Japanese mass media" among his "domains of English in modern Japanese society". While one might assume that the news would wish to be as comprehensible as possible and might therefore avoid excessive use of LWs, many critics feel that the proportion of *gairaigo* in the news media is still too great (Mizutani 2003), and there is even a perception that the media tries to achieve a higher register by using esoteric loanwords. As a result, the level of *gairaigo* usage in the news has come under heavy criticism from commentators (Ōno, Morimoto & Suzuki 2001; Yoshimi 2018).

In fact, the Japanese public themselves have often voiced their discontent regarding the frequent use of *gairaigo* in the media. This made the headlines in 2013 when a 71-year-old man sought compensation from the national broadcaster NHK for the emotional distress caused by the inordinate use of *gairaigo* on the air. He claimed that his inability to comprehend adopted words—such as those based on "risk," and "trouble"— led to "mental distress" as he could not understand the content of programs (Osaki 2013).
Such outrage at foreign terms has occasionally been echoed by politicians, the very people who are themselves often accused of *gairaigo* abuse. Indeed, many were quick to point out this ironic fact when Abe Shinzo, the recently retired Prime Minister of Japan renowned for his nationalistic views, claimed that to "create a 'beautiful Japan' we must first remember how wonderful our nation is," and that "we will start a new, future-oriented *purojekuto* (project) aimed at strategically promoting the new Japanese *kantorii aidentiti* (country identity)" (Otake 2007). Abe's unconscious choice of wording demonstrates a simple fact: that *gairaigo* has become so firmly embedded in higher register Japanese such as political rhetoric, that any ideologically based hesitancy to use it is superseded.

4. Previous research on *gairaigo*

4.1. Categorisation

When attempting to categorise loanwords, researchers of *gairaigo* have developed various groupings. For example, Honna (1995) based his seven divisions of *gairaigo* on borrowing patterns and the processes through which *gairaigo* is derived.[3] However, in these complex systems of categorisation, several issues emerge such as overlapping categories, homophones, and changes in meaning over time. A functional categorisation of loanwords is provided by Myers-Scotton (2006), who created a dichotomy called "core/cultural" which is based on both borrowing type and function. Cultural borrowing refers to LWs that fill a gap in the lexicon as they denote objects or concepts that are novel to the culture; these are equivalent to what Tomoda (2005) called lexical-gap fillers. Core borrowing, on the other hand, indicates LWs that duplicate a word already in existence in the recipient language but are often used for special effect. To take a commonly used example in Japanese, the word *shoppu* ショップ 'shop', would fall into the category of core borrowing as it already has a *kango* Japanese equivalent: *mise* 店. Conversely, the word for convenience store (コンビニ, *konbini*), had no direct native Japanese equivalent at the time of its conception and is therefore an example of cultural borrowing.

The "core/cultural" divide is perhaps the most objective means of classifying LWs. Its concrete foundation on the pre-existence of a referent in Japanese culture is less given to ambiguity than the overlapping categories of borrowing types. My decision to use this method of categorisation in the following research was also based on the fact that this paper aims not to identify how novel loanwords are derived, but rather what effect they have (see Section 9).

[3] Semantic shift and narrowing, Japanese English, combinations of Japanese words and English loanwords, contractions, acronyms, abbreviations of compound words, and word play.

4.2. Comprehension

Much of the practical discussion of *gairaigo* has focused on recognition and understanding of LW terms as well as generational differences in comprehension (Tomari 1985; Shibata 1993; NHK 2001; NINJAL 2006; Horikawa 2012).

Many investigations and surveys on the question of comprehension were conducted in the 20th century. For example, the national broadcaster NHK carried out a survey in order to assess recognition and understanding of *gairaigo* by the Japanese public using LWs that were commonly found in the media. While the average rate of recognition stood at 77%, the comprehension rate was significantly lower at 50%. This was because several loanwords were subject to misunderstanding. For example, *disukaunto* ディスカウント (discount) was misconstrued as meaning 'very cheap' (Ishino, Maruta & Tsuchiya 1988). A 1995 survey conducted by Ōnishi & Kajiki using 15 LWs found a lower rate of recognition at 59%, and comprehension at just 36%.

However, the 'comprehension rates' described above, were derived using the original English definition of the LWs as a yardstick for measuring understanding. This method does not consider the common phenomenon of semantic shift that often occurs when LWs are adopted. This means that comprehension may have been higher if they had measured it within a Japanese context (Tomoda 2005). However, even when asked to self-report on their comprehension of *gairaigo* more recently, 28.1% of the Japanese republic reported trouble understanding LWs broadcast on TV (Yamashita & Katō 2000).

When making comparisons in LW comprehension across age groups, stark differences come to light. For example, a 1988 NHK survey revealed a 35% difference in comprehension between those aged 25-29 (63%) and those aged 60 years and over (28%) (Ishino, Maruta & Tsuchiya 1988). Indeed, it seems that the older the generation, the more they struggle with comprehension, as revealed by Loveday (1996), who found the comprehension of LWs in those aged 70–79 to be just 9.2%.

In 2006, the National Institute for Japanese Language and Linguistics gathered data from the Japanese public (NINJAL 2006). Horikawa (2012) tested the overall comprehension of 52 of the more high-frequency LWs and came up with an overall comprehension figure of 63.5%. However, participants aged 60 years and older were found to have a comprehension rate of just of 43.9% (Horikawa 2012). She found that rates between different LWs varied greatly, ranging from 6.3% to 92.6% comprehension.

In her breakdown of NINJAL's results, Horikawa also analysed the types of borrowing using Myers-Scotton's (2006) core/cultural dichotomy. The difference in comprehension between the age ≥ 60 group and the overall group was largest for cultural borrowings, but it was not as great for core borrowings. Horikawa explains this by referring to the more complex process that is involved in learning vocabulary derived from cultural borrowing; a Japanese native speaker must first "learn the content of the referent itself before they are able to connect the meaning and the word" (Horikawa 2012: 61). These findings suggest that the participants over 60 years of age struggled to understand the LWs derived via cultural borrowing as they lacked the prior knowledge of the objects or concepts to which the LWs refer.

4.3. Attitudes

English can function as a prestige and status marker (Haarmann 1989), yet this is at odds with the highly monocultural and monolingual character of Japanese society (see Heinrich 2012). Such a juxtaposition brings about polarised opinions and commentators tend to be divided into two camps: those who accept the LW influx as a welcome signal of positive change and those who view the LWs as a malignant force that corrupts Japan's language and erodes its culture.

Arguments in support of *gairaigo* usage often focus on the way LWs can enrich the Japanese lexicon and provide additional nuance (Ishino 1983; Kajima 1994; Bordilovskaya 2012). Other commentators have connected the influx of *gairaigo* with internationalisation, cosmopolitanism, and increased English proficiency (Ogaeri 1960; Ishiwata 1989; Honna 1995). Some in the academic world, however, view *gairaigo* more pragmatically: as an inevitable product of globalisation which should be welcomed if Japan is to progress as a nation (Shibata 1993).

There are also many condemnatory stances towards *gairaigo*. Indeed, there is a view that the overuse of LWs tarnishes the Japanese language, or as Loveday (1996: 208) puts it "(the opinion that) the current extent of Western borrowings is leading to language 'decline' and is taken as a sign that the Japanese have lost faith in their own linguistic creativity". This interpretation is manifest in the idea that giving in to Western influence is liable to lead to an erosion of culture that invites confusion and exhibits shallowness (Ishii 1998; Ōno, Morimoto & Suzuki 2001).

Opponents of *gairaigo* usage have tended to focus on lack of comprehension and the social division this could cause. Similarly, the overuse of *gairaigo* by government bodies has even led to calls from scholars for the adoption of an exclusionist policy to LWs modelled on the French approach (Mizutani & Ōno 1995). These grievances are commonly aired by language purists such as the aforementioned NHK viewer and those who submit *tousho* (letters to the editor) that newspapers receive in abundance, berating them for their excessive quantity of *gairaigo* (Yahagi 2013; Yoshimi 2018).

To explain the polarity of views evidenced above, Irwin (2011: 199–200), proposed that the Japanese populace are party to a so-called "love/hate relationship" with LWs. This relationship manifests itself in the societal belief that *gairaigo* is simultaneously an indispensable tool for creating a more advanced, democratic society, at the same time as being a linguistically imperialistic or even colonialist threat to Japanese culture and tradition (Irwin 2011: 200).

4.4. Summary

While the research on loanword types, comprehension, and attitudes may seem comprehensive, the never-ceasing production of new loanwords means that there will always be space for studies that focus on newly derived vocabulary. This led me to believe that the trends revealed in previous research, such as a low comprehension rates in older generations and polarised attitudes, may also be prevalent, or even more extreme, in Covid-19 related LWs.

Whilst reading prior studies on LWs, I noticed a gap in research related to demographic variables other than age. Bearing this in mind, I decided to include the variable of location in my own research in an attempt to highlight further demographic differences in comprehension and attitudes towards loanwords. Furthermore, I found that very few attitude studies directly compared attitudes towards loanwords and their native Japanese equivalents. This led me to believe that a study which did so could reveal insights into the utility and value of loanwords.

5. Background to the research

Covid-19 related loanwords began emerging very early in the pandemic. On the 25th of March 2020, Tokyo Governor Koike Yuriko held a press conference to address a rapid increase in Covid-19 infections in the capital. She warned residents to stay at home so as not to cause an *ōbāshūto* オーバーシュート (overshoot), an English LW taken to mean a sudden explosive rise in cases. While not out of character for the former TV newsreader, Koike's repeated use of LWs such as *rokkudaun* ロックダウン (lockdown) and *kurasutā* クラスター (cluster), sparked arguments over the suitability of this terminology in such a deeply important broadcast, when universal comprehension was so vital (Brasor 2020).

On one side of the argument were opponents to *gairaigo* usage such as then Defence Minister Kono Taro, who tweeted his concern that discussion of the coronavirus crisis involved foreign LWs that are largely incomprehensible to the average Japanese citizen (Yamashita 2020). Kono insisted that public officials should find more easily understood native Japanese terms instead.

The translator Reizei Akihiko, on the other hand, believed that the use of Japanese equivalents to Koike's LWs would actually cause more confusion. In his essay in defence of Koike, he theorised that her use of the LWs allowed her to better convey a "sense of crisis" (Reizei 2020). This, he argued, is because a Japanese equivalent in current use such as *shūdan kansen* 集団感染 'cluster of infections', is already associated with the yearly seasonal flu, and thus may inappropriately encourage a false sense of normality.

This is not the first time that a crisis put *gairaigo* usage in the spotlight. In the wake of the Great Kansai Earthquake of 1995, the news media came under fire for its use of loanwords such as *infura* インフラ (infrastructure) and *raifurain* ライフライン (lifeline utilities, i.e. water, electricity, gas) in resources aimed at survivors (Tomoda 2005). This criticism may well have been justified, as illustrated by Ōnishi & Kajiki (1995) who found that *infura* インフラ was understood by just 34% of those who recognised it and that 43% of participants believed it to have the same meaning as the loanword *infure* インフレ (inflation). It is reasonable to say that a lack of comprehension of such words in a time of crisis could lead to increased anxiety rather than clarity and assurance.

Two days after the governor's appearance on television, a Tokyo Shimbun reporter decided to put Koike's loanwords to the test and took to the streets of the capital to ask young people if they understood the governor's message. The results were similar to those found by Ōnishi & Kajiki in the aftermath of the 1995 earthquake. One high school-aged respondent believed that "lockdown" was a synonym for the virus, while another 19-year-old interviewee answered that Koike's phrasing meant he felt the announcement had no connection to his own life (Brasor 2020).

In the following sections, I will aim to build on the informal research performed by journalists that has already been undertaken on this topic. The aim is to explore which LWs have entered the Japanese vocabulary due to the Covid-19 pandemic, to categorise them and then to ascertain the levels of comprehension and attitudes towards them in order to determine their appropriateness for usage in the media and reporting. There are three primary research questions explored in the two case studies described in the following sections:

1) Which novel loanwords have emerged as the result of the Covid-19 crisis?
2) Are these novel loanwords comprehensible to Japanese speakers?
3) What attitudes are associated with these novel loanwords?

6. Case study: part one

6.1. Sourcing and identifying novel loanwords

In order to answer the first of my research questions and obtain data for the others, I decided to use Twitter data. Having reached 4.9 million registered users as of July 2020, Twitter was the second-most engaged-with social media platform in Japan after Line (Statista, 2020) and is commonly used for the proliferation and discussion of news items. Twitter could therefore guarantee data that involved authentic public discussion of Covid-19 in the form of user replies to news articles.

The final data used for this study was comprised of the user replies to 100 tweets from 5 of Japan's largest news sources. User replies were chosen as opposed to tweets from the news sources themselves so as to study the language of the Japanese public, rather than that of the media. Historical tweets were retrieved from Twitter's API through a process called scraping. Tweepy, an open-source Python package, was utilised to refine the tweets through the use of parameters. I selected articles containing the keyword *koronauirusu* コロナウイルス tweeted by the newspapers Asahi Shimbun (@asahi), Mainichi Shimbun (@mainichi), Sankei Shimbun (@Sankei_news) and the Nikkei Shimbun (@nikkei), as well as the national broadcaster NHK (@nhk_news) from March the 22nd - 29th, the week that saw Japan's first rapid spike in Covid-19 cases ("Tokyo governor urges" 2020).[4]

I then used a combination of UniDic (Den et al. 2010) and Mecab (Kudou 2013) to modify the data. These are corpus analysis tools designed to prepare Japanese, which lacks spaces, for Natural Language Processing. UniDic, facilitates tokenisation by splitting Japanese text into individual units. These units are divided into words and non-word units (letters, punctuation marks and symbols). The word units are then assigned several descriptive fields by the programme which includes the lexeme, written form, part of speech and word type. I utilised the "part of speech" description in order to separate word units from non-word units and discard the latter from the study, then used the "word type" description to sort the words by their origin: Japanese, Chinese or *gairaigo*.

Once I had isolated the *gairaigo*. the next step was to identify and extract "novel loanwords" (henceforth NLWs) related to coronavirus. Firstly, to ascertain words with a

[4] I also considered The Yomiuri Shimbun but chose to exclude it as its online presence is minimal and does not generate enough user interaction to be useful to this study.

greater relevance to the discussion of coronavirus I excluded proper nouns such as country, place, and company names. I also excluded the word *koronauirusu* コロナウイルス (coronavirus) itself.

Secondly, as a means of addressing the question of novel usage I removed the LWs that are already a regular part of the Japanese lexicon. To standardise this classification, I only used LWs without an entry in the *Sanseidō kokugo jiten*, a general-purpose Japanese dictionary (Kenbo et al. 2013).

Some of the most frequently mentioned LWs in the data including *risuku* リスク (risk – 20 times), *panikku* パニック (panic – 15 times) and *piiku* ピーク (peak – 3 times), may have relevance in the context of the Covid-19 pandemic, but they could not be included as examples of novel usage as they are already well-established LWs. This process of refinement left the list of loanwords seen below in the results section (Table 1) which were used in the second part of the study.

Table 1: Individual novel loanwords ordered by frequency

Transcription	Katakana	English translation	Count	Borrowing type
Rokkudaun	ロックダウン	Lockdown	34	Core
Kurasutā (kurasuta)	クラスター（クラスタ）	Cluster	30 (33)	Core
Ōbāshūto	オーバーシュート	Overshoot	25	Core
Pandemikku	パンデミック	Pandemic	11	Core
Ebidensu	エビデンス	Evidence	10	Core
Feikunyūsu	フェイクニュース	fake news	4	Core/Cultural
Fēsushiirudo	フェースシールド	face shield	3	Cultural
Masugomi	マスゴミ	mass comms + rubbish (portmanteau)	2	Cultural
Koronashokku	コロナショック	coronavirus shock	2	Cultural
Koronapanikku	コロナパニック	coronavirus panic	2	Cultural
Autobureiku	アウトブレイク	Outbreak	2	Core
Fakutochekku	ファクトチェック	fact check	2	Core
Chainauirusu	チャイナウイルス	China virus	2	Cultural
Suteihōmu	ステイホーム	stay home	2	Core
Heitosupiichi	ヘイトスピーチ	hate speech	2	Core/cultural
Afutākorona	アフターコロナ	post-coronavirus	1	Cultural
Conpuraiansu	コンプライアンス	compliance	1	Core
Monitaringu	モニタリング	monitoring	1	Core
Orinpikkufāsuto	オリンピックファースト	Olympics first	1	Cultural
Japanfāsuto	ジャパンファースト	Japan first	1	Core/Cultural
Medianoriterashii	メディアリテラシー	media literacy	1	Cultural
Total words: 21			142	

6.2. Results

The words were categorised using Myers-Scotton's (2006) "core/cultural" dichotomy which is based on function and the absence or presence of a Japanese equivalent (henceforth JE). In addition, a third category — core/cultural borrowing — suggested by Horikawa (2012) combines the first two:

1) **Core borrowing:** LWs that have JEs already in existence.
2) **Cultural borrowing:** LWs that fill gaps in the Japanese lexicon as they refer to objects or concepts that are new to Japanese culture.
3) **Core/cultural borrowing:** LWs that could refer to multiple objects or concepts, some of which existed in Japanese culture before and some of which did not.

Core borrowing was responsible for around half of word types (11) but as much as 85.2% of the total NLW token. Cultural borrowing, however, provided a little under 10% with core/cultural borrowing making up just 4.9%. Horikawa's (2012) results for high-frequency English LWs also followed this pattern, with core borrowing tokens outweighing cultural borrowing tokens 2 to 1. Takashi's (1992) study on LWs in advertising found similar results with 45% of LW tokens classified as 'special-effect givers' (core borrowing), while only 16% served the function of filling a lexical gap (cultural borrowing).

The majority of NLWs have low frequencies and are unlikely to become candidates for entering the Japanese vocabulary. Furthermore, if we pick out several of the examples of NLWs, many appear to be portmanteau words, such as *koronashokku* (corona + shock) and *masugomi* (mass communication + rubbish), made from two other loanwords and intended as a fleeting piece of wordplay to satirise a specific cultural event.

However, at the top of the table we do see several NLWs with high token counts. It is these which account for the dominance of core borrowing. This shows that much of the borrowing is occurring for, in the words of Takashi (1992), "special effect". The nature of this special effect will be tested and discussed in the attitudes section of the second part of this case study, however the prevalence of core borrowing (the use of words which have existing JEs) does already call into question the utility of these words at the level of mass communication when weighed against difficulties in comprehension.

7. Case study: part two

7.1. Methodology

In order to answer the second and third of this paper's research questions, I decided to use a questionnaire that collected a combination of qualitative and quantitative data. This allowed me to gauge the comprehension rates and attitudes towards the NLWs related to Covid-19.

This study follows a number of others examining comprehension of LWs in Japanese, specifically Ōnishi & Kajiki (1995) and NINJAL (2006) amongst others. The study will test the hypothesis that Japanese speakers have difficulty understanding LWs and that older generations and those residing outside of Tokyo have more trouble than the general population with comprehension.

The questionnaire was made up of six sections and directed at the native Japanese speaking population. It was therefore presented entirely in Japanese. It was created using Google Forms' web-based survey function and distributed online through social media, forums, and personal contacts. Responses were recorded for two weeks from August 4 to August 18, 2020.

It must be acknowledged here that the questionnaire relied on participants to self-report on their own comprehension of loanwords. This is potentially problematic due to well-known issues with self-report studies such as exaggeration and social desirability (Northrup 1996).[5] In future surveys, techniques such as true-or-false or multiple-choice questions and presenting the NLWs in context could be used to improve validity and create a more realistic test of the respondents' comprehension.

7.2. Participants

Of the 121 total participants 119 chose to disclose their gender, resulting in 41 men and 78 women. In terms of age groupings, 64 were 18-29, 19 were 30-39, 5 were 40-49, 16 were 50-59 and 17 were aged 60 and above. All participants were native speakers of Japanese and living in Japan; 25 were Tokyo residents and 96 lived elsewhere.

7.3. Questionnaire

The questionnaire was divided into 7 sections. The first section was the informed consent form, the second asked participants their basic demographic information: sex, age, residence. The next five sections asked participants to self-report on their comprehension of a NLW both before the Covid-19 outbreak and in the present as seen in Figure 1.

This page was presented to each participant five times, addressing a different NLW each time. The target vocabularies were chosen from the five most frequently occurring NLWs as per the results from the first part of this case study, these were as follows:

1) *pandemikku*　パンデミック (pandemic)
2) *rokkudaun* ロックダウン (lockdown)
3) *kurasutā* クラスター (cluster)
4) *ōbāshūto* オーバーシュート (overshoot)
5) *ebidensu* エビデンス (evidence)

Using the quantitative data gathered from these questions, I was able to calculate comprehension rates for each of the NLWs separately and take an average for all five. I could then compare the comprehension rates for the participants overall with those in the "age ≥ 60" category and those who were Tokyo residents.

[5] Due to similar concerns about reliability, the results produced by asking participants to recall their comprehension of loanwords 6 months prior to when the questionnaire was distributed were omitted.

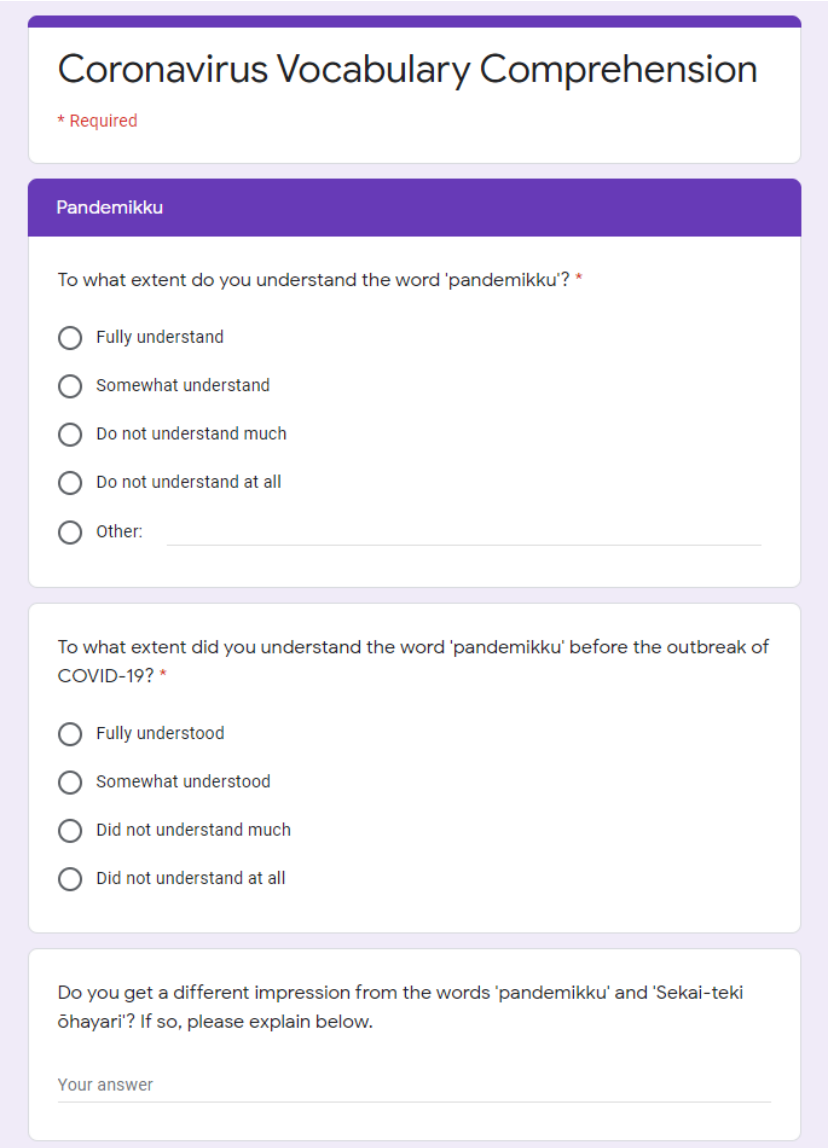


Figure 1. Example page from the questionnaire (English translation)

Table 2: Native Japanese equivalents for NLWs

Novel loanword	Native Japanese equivalent
pandemikku パンデミック	*sekaiteki dairyūkou* 世界的大流行
rokkudaun ロックダウン	*toshi heisa* 都市封鎖
kurasutā クラスター	*shūdan kansen* 集団感染
ōbāshūto オーバーシュート	*kansen bakuhatsu* 感染爆発
ebidensu エビデンス	*konkyo* 根拠

After self-reporting on their comprehension of each NLW, the participants were then asked to compare the NLWs with their Japanese equivalents. The JEs to each NLW seen in Table 2 were selected based on the suggestions of critic Reizei (2020). These open-ended questions were optional, so answers were not required for participants' responses to be recorded.

7.4. Comprehension results

The overall average comprehension rate for the five NLWs was 35.7%, while the average comprehension rate for participants over 60 years old was far lower at 17.6%. Tokyo residents, however, showed a distinct advantage in comprehension with an average rate of 46%, 10.3% above the overall comprehension rate. The average comprehension rates found in this study were far lower than those found by Horikawa in her study of high-frequency LWs using data from NINJAL's 2006 survey (Horikawa 2012: 59). However, both studies also showed an average difference between the overall group and over-60s of almost 20%.

Comprehension also varied greatly depending upon the word. While *rokkudaun* (lockdown) was understood by just over half of the participants and 29.4% of those aged 60 or over, *ōbāshūto* (overshoot) was understood by only 10.7% of all participants and was not understood by anyone aged 60 or above. The comprehension rates were generally better for NLWs that appeared more frequently in the Twitter data, though *ōbāshūto* was a marked outlier.

If we compare the NLWs in Table 3 with high-frequency loanwords that were found to have similar comprehension rates in Horikawa's investigation (Table 4) we notice an interesting trend. The majority of high-frequency loanwords with lower comprehension rates in Horikawa's study were related to computers or the internet. It is therefore possible to make an analogy between the comprehension of NLWs related to the coronavirus pandemic in 2020 and high frequency LWs related to computers and the internet in 2006 when NINJAL collected the data for Horikawa's study.

Table 3: Comprehension rates for NLWs

LW (Romanised)	Katakana	English translation	Frequency count in data	Borrowing type	Total comp rate (%)	60+ comp rate (%)	Tokyo resident comp rate (%)
rokkudaun	ロックダウン	lockdown	34	Core	52.1	29.4	68.0
kurasutā	クラスター	cluster	33	Core	48.8	35.3	52.0
ebidensu	エビデンス	evidence	11	Core	38.8	5.9	60.0
pandemikku	パンデミック	pandemic	10	Core	28.1	17.6	40.0
ōbāshūto	オーバーシュート	overshoot	30	Core	10.7	0.0	12.0
Average					**35.7**	**17.6**	**46**
Average found by Horikawa (2012)					**63.5**	**43.9**	**N/A**

Table 4: High-frequency LWs with an overall comprehension rate of under 50% as recorded in the *Language Attitude and Pervasion Survey* (NINJAL 2006) adapted from (Horikawa 2012)

Loanword Romanised	Katakana	English translation	Borrowing type	Overall comp rate (%)	age ≥ 60 comp rate (%)
Fōramu	フォーラム	Forum	Core	46.8	31.6
Dētābēsu	データーベース	Database	Cultural	45.6	21.5
Gurōbaru	グローバル	Global	Core	41.3	18.7
Tsūru	ツール	Tool	Core/Cult	40.9	18.7
Purobaidā	プロバイダー	Provider	Cultural	40.6	15.2
Daunrōdo	ダウンロード	Download	Cultural	40.6	8.2
Rinku	リンク	Link	Core/Cult	38.5	10.4
Saito	サイト	Site	Cultural	34.4	7.8
Inishiachibu	イニシアチブ	Initiative	Core	27.4	15.0
Kontentsu	コンテンツ	Contents	Core/Cult	23.0	8.8

7.5. Attitudes results

The participants were also asked to compare the NLWs with their Japanese equivalents as listed in Section 7.3. Of the 605 possible responses to the open-ended questions, I received 59 responses. Of these responses, I found that on 29 occasions the respondent received the same impression from the NLW and JE, on 18 occasions they felt the words were synonymic, while the remaining 12 responses were neutral (Figure 2).

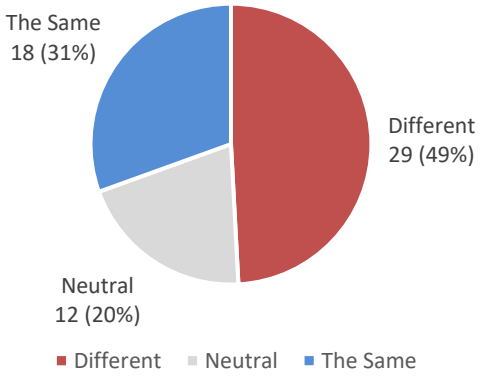


Figure 2. Impression of the NLWs vs. JEs

7.5.1. Comprehensibility

Of the 7 responses that referred to comprehensibility, 6 respondents said that they found the JE easier to understand than the NLW. They reported the following on this matter:

> "I don't understand the nuance of *'ebidensu'*, if they have the same meaning, I don't get why the Japanese government use that word instead of '*konkyo'*."
> (Male aged 18-29)

In contrast to this, only one response claimed that the NLW was easier to understand. This provides further evidence to support the qualitative data in this study as well as the plentiful previous research (see Ishino & Yasuhira 1991; Ōnishi & Kajiki 1995; Yamashita & Katō 2000), which concludes that comprehension problems relating to *gairaigo* are common.

7.5.2. Severity

13 participants commented on their perceived sense of severity regarding the pairs of words. 8 claimed that the NLW sounded more severe; of these, 5 referred to the word *pandemikku*. Most participants gave similar reasons for this:

> "I feel that *pandemikku* has more of a nuance that infection is spreading quickly."
> (Male aged 18-29)

> "With *pandemikku* I feel a greater sense of danger."
> (Female aged 50-59)

Here we can see the practical benefits of using the NLWs that Reizei (2020) referred to in his essay in defence of Governor Koike's use of English LWs. The participants quoted above attest to a sense of crisis being conveyed by the NLWs. However, 5 participants claimed that two of the NLWs (*kurasutā* and *ōbāshūto*) actually had a softer impression and that the JEs sounded more severe:

> "I feel more of a sense of warning from the kanji version (than *ōbāshūto*)."
> (Female aged 50-59)

> "With *kurasutā* I have an image of a smaller scale infection cluster."
> (Male aged 18-29)

Given the feedback from respondents, I believe there are potentially two reasons for this. Firstly, because Japanese people can read and abstract meaning from the *kanji* that are used to write the JEs, they get a more immediate sense of danger, they do not even need to fully comprehend the word as a whole. Secondly, because *kurasutā and ōbāshūto* can have other potential applications or referents (see below), their meanings are broader and so their efficacy at conveying a sense of danger is diluted.

7.5.3. Broadness/narrowness

On the topic of broadness there was an even split with 3 participants believing that the NLW carries a broader definition and 3 believing that the JE did. 2 participants believed that *pandemikku* has a much narrower meaning than its JE:

> "*Pandemikku* refers to an infectious disease but *sekaiteki dairyūkou* can possibly refer to other things like the spread of a song or trend."
> (Female aged 30-39)

Here we see what Ishino (1983) and Bordilovskaya (2012) embraced as the enriching effect that LWs can have upon the Japanese vocabulary. By using the NLW, one is able to achieve additional nuance in their speech and point at more specific referents allowing for clearer use of language. However, much like the responses regarding severity, I found that other participants actually believed the NLWs words *kurasutā* and *ōbāshūto* to be less effective in the discussion of Covid-19:

> "I have sometimes heard *ōbāshūto,* for example in economics TV programmes (…) in most cases I think it is used to mean 'going too far'."
> (Male aged 40-49)

This is because both have an additional domain of use, separate to their JEs. For *ōbāshūto* this is economics and, as I was informed by one participant, *kurasutā* can also be used as *otaku* (nerd) terminology to mean a gathering of a certain type of people (usually a fandom).

7.6. Discussion

The overall comprehension rate of 35.7% provides an answer to the second research question in that it proves that the NLWs are largely incomprehensible to, or at least not fully understandable by, the average Japanese citizen. Overall, the results of this study agree with then Defence Minister Kono Taro in his concern that discussion of the coronavirus crisis involved an excess of foreign LWs which precipitates misunderstanding (Yamashita 2020). Meanwhile, the concern about low levels of comprehension of LWs in the older generations that prompted the research of Ishino, Maruta & Tsuchiya (1988) and Loveday (1996), was proven to be relevant in the current age due to the extremely low average comprehension rate of 17.6% amongst those aged 60 and above in this research.

The results also suggest that age is not the only factor to have an influence on NLW comprehension. Tokyo residents considerably outperformed the average, perhaps as the result of better English language education and exposure to a more cosmopolitan life. This feeds into the idea championed by opponents of *gairaigo* such as Mizutani & Ōno (1995), that *gairaigo* could fuel social division.

It also seems there will be great variance in longevity as vocabulary. An overall comprehension rate of 52.1% suggest that a word like *"rokkudaun"* has potential sticking power, whereas "*ōbāshūto*" is unlikely to outlast the Covid-19 pandemic with a comprehension rate of just 10.7%. Even Tokyo residents seem to have forgotten Governor

Koike's speech, with a comprehension rate of 12.0% regarding "*ōbāshūto*" by August 2020. Indeed, the short lifespan of some LWs is recognised as a part of the naturalisation process of borrowing (Matsuda 1986; Tomoda 1999).

It was difficult to answer the third research question based on the results of the open-ended attitudes questions. While they did show that these NLWs were more difficult to comprehend than their JEs, the results regarding severity and broadness varied between NLWs. This means that Reizei's (2020) defence of LW usage as a means of highlighting a "sense of crisis" could not be applied to every NLW. A larger data set would help to establish trends and provide a more definitive answer in future research. What the responses did teach us, however, is that even LWs specifically coined for usage in relation to Covid-19 may have other conflicting domains of use which make them just as ambiguous as their often polysemous JEs.

8. *Afutākorona*: implications

The case study illustrated that not only people over 60 years old and non-Tokyo residents, but the Japanese population as a whole may have problems in comprehending NLWs. It seems that NLWs related to Covid-19 may share similarly low comprehension rates with those which were introduced as computer and internet-related terms over a decade ago (Horikawa 2012). However, the role of loanwords may be a more urgent issue now, amid a global pandemic, than it was 10 or 15 years ago. A lack of knowledge regarding the language of public health and safety could cause serious issues. If Japanese officials continue to brand safety campaigns (see Figure 3) and litter political speeches with loanwords, it could have a dangerous exclusionary effect on those with lower English levels and in older generations, the group most vulnerable to Covid-19 (CDC 2020).

Figure 3: A chef walks past a KFC restaurant in Kamakura where Colonel Sanders in a mask and samurai armour tells him to "*sutei hōmu*" (stay home) ("New virus cases" 2020).

As evidenced by the words found on Twitter, there is some degree of uptake and usage of the NLWs used by public officials in Japan and, while an optimist would say that this is evidence of people engaging with *gairaigo*, the reality is that Twitter users are a narrow

sample, not representative of the Japanese populace, and less likely to contain those in older generations.

Furthermore, while some commentators (Ishino 1983; Bordilovskaya 2012) might insist on the additional nuance that can be achieved through the use of *gairaigo*, the attitudes gleaned from the questionnaire showed mixed responses. The impressions associated with NLWs varied depending upon the item and therefore it strikes me that it is extremely important to consider when the use of a certain LW term is appropriate.

9. Limitations and suggestions for future research

The main limitation of this study was the sample sizes. Due to time and resource limitations, I was only able to recruit 121 participants for the questionnaire. To improve the validity of my findings, a greater sample size is required. Similarly, I would like to gather more qualitative data in order to draw stronger conclusions and provide more detailed insights into the relationship between Japanese society and novel loanwords. Further examination of these attitudes may reveal deeper understanding of the reason they are created and used.

While this study does go some way towards helping understand why NLWs are used (i.e. "special effect") and what impact they have, the issue of how NLWs are formed is not addressed. A recommendation for future research would be a project that scrapes tweets over a longer period to chart where NLWs come from and how their usage changes over time, I believe this would give a clearer idea of how crises and events influence the creation of NLWs.

Another limitation was the age of the sources I used for comparison. For example, Horikawa's (2012) study used data from 2006, yet it is likely that comprehension rates have changed since. For future research, I would suggest gathering a fresh set of data regarding high-frequency LWs with which to compare the NLWs.

10. Conclusion

In conclusion, crises, especially those on a global scale, can be a significant event in Japanese language change. However, this is not necessarily a change that affects the entire population equally, or fairly. This is an issue for a large proportion of the Japanese population who are less able to adapt to novel loanwords quickly. If they fully understand the effect these NLWs have on the public and recognise that they can be inaccessible to many Japanese people, I believe that public officials in Japan might begin to monitor their use of NLWs going forward.

However, this phenomenon is not unique to Japan. With the pandemic still raging over one year later, this is a time for all nations to reassess the accessibility of their own language and governments may need engage in similar forms of language planning. While language planning on this scale has proven difficult in the past (see Fishman 1983), perhaps the unprecedented nature of the times we live in may be a catalyst for change.

References

Bordilovskaya, Anna. 2012. A study of loan colour terms collocation in modern Japanese. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 34, 1362–1367.

Brasor, Philip. 2020. COVID-19 spurs debate over loan words. *Japan Times*. www.japantimes.co.jp/news/2020/04/11/national/media-national/covid-19-spurs-debate-loan-words/ (20 June, 2020).

CDC. 2020. Older Adults. *Centers for Disease Control and Prevention*. www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html (25 September, 2020).

Den, Y., A. Yamada, H. Ogura, H. Koiso & T. Ogiso. 2010. *UniDic*. http://www.tokuteicorpus.jp/dist/ (15 June, 2020).

Fishman, Joshua A. 1983. Modeling rationales in corpus planning. In Juan Cobarrubias & Joshua A. Fishman (eds.), *Progress in Language Planning: International Perspectives*. Berlin: Mouton Publishers.

Haarmann, Harald. 1986. Verbal strategies in Japanese fashion magazines - a study in impersonal bilingualism and ethnosymbolism. *International Journal of the Sociology of Language* 58. 107–121.

Haarmann, Harald. 1989. *Symbolic values of foreign language use: From the Japanese case to a general sociolinguistic perspective*. Berlin: Mouton de Gruyter.

Heinrich, Patrick. 2012. *The making of monolingual Japan: Language ideology and Japanese modernity*. Bristol: Multilingual Matters.

Hogan, Jackie. 2003. The social significance of English usage in Japan. *Japanese Studies* 23(1). 43–58.

Honna, Nobuyuki. 1995. English in Japanese society: Language within language. *Journal of Multilingual and Multicultural Development* 16(1 & 2). 45–62.

Horikawa, Naoko. 2012. *English loan words in Japanese: Exploring comprehension and register*. Portland State University MA thesis. https://pdxscholar.library.pdx.edu/open_access_etds/913/ (18 August, 2020).

Irwin, Mark. 2011. *Loanwords in Japanese*. Amsterdam: John Benjamins.

Ishii, S. 1998. Nihonjin to eigo no kankei. In Y. Tsuda (ed.), *Nihonjin to eigo*. Kyoto: Kokusai-nihonbunka-kenkyūsentā.

Ishino, H., M. Maruta & T. Tsuchiya. 1988. Gairaigo wa dō ukeirareteiruka. *Hōsōkenkyū-to-Chōsa* July. 24–39.

Ishino, H. & M. Yasuhira. 1991. Kokusaikajidai no nihongo 1. *Hōsōkenkyū-to-Chōsa* August. 2–11.

Ishino, Hiroshi. 1983. *Gendai-gairaigokō*. Tokyo: Taishūkan.

Ishiwata, T. 1989. Loanwords in historical perspective. *Japan Echo* 16. 17–21.

Kajima, S. 1994. *Katakana-eigo no hanashi*. Tokyo: Nan'undō.

Kenbo, Hidetoshi, Takashi Ichikawa, Yoshifumi Hida, Makoto Yamazaki, Hiroaki Iima & Takehiro Shioda (eds.). 2013. *Sanseidō kokugo jiten 三省堂国語辞典*. 7th edn. Sanseidō.

Kudou, Taku. 2013. *Mecab*. github.com/taku910/mecab.

Liberman, Mark. 2011. Antedating tsunami. Weblog. *Language Log*. https://languagelog.ldc.upenn.edu/nll/?p=3028 (21 June, 2021).

Loveday, Leo J. 1986. Japanese sociolinguistics: An introductory survey. *Journal of Pragmatics* 10. 287–326.

Loveday, Leo J. 1996. *Language contact in Japan*. Oxford: Clarendon Press.

Matsuda, Yutaka. 1986. Cross-over languages: Japanese and English, part 2. *Kwansei Gakuin* 35(12). 47–80.
Mizutani, O. 2003. Nihongo no kokusaika to gairaigo. *Nihongogaku* 22(7). 4–8.
Mizutani, O. & S Ōno. 1995. Tairitsu-tōron: Ra-nuki to katakanago. *Yomiuri* 14–15.
Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism*. Malden, MA: Blackwell Publishers.
"New virus cases." 2020. New virus cases in Tokyo drop to 5; none reported in Osaka. *Japan Today*. japantoday.com/category/national/update1-new-virus-cases-in-tokyo-drop-to-5-none-reported-in-osaka (25 September, 2020).
NHK. 2001. *Broadcasting Culture Research Institute [Nippon Hōsō Kyōkai]*. www.nhk.or.jp/bunken (15 September, 2020).
NINJAL. 2006. Gairaigo teichakudo chousa. http://www.ninjal.ac.jp/gairaigo/Yoron/index.html (18 August, 2020).
NLRI. 1964. *Vocabulary and Chinese characters in ninety magazines of today*. Tokyo: Shūeishuppan.
Ogaeri, Y. 1960. Kango ni kawaru katakanago. *Gengoseikatsu* 103. 26–32.
Ōnishi, K. & M. Kajiki. 1995. Katakana-gairaigo 'raifurain' wa wakarimasuka. *Hōsōkenkyū- to-Chōsa* September. 14–25.
Ōno, S., T. Morimoto & T. Suzuki. 2001. *Nihon, nihongo, nihonjin*. Tokyo: Shinchōsha.
Osaki, Tomohiro. 2013. Gifu man, 71, sues NHK for distress over its excess use of foreign words. *Japan Times*. www.japantimes.co.jp/news/2013/06/27/national/crime-legal/gifu-man-71-sues-nhk-for-distress-over-its-excess-use-of-foreign-words/ (30 August, 2021).
Otake, Tomoko. 2007. Japanese: A language in a state of flux: "Torrential" import of foreign words threatens the basis of communication. *Japan Times*. https://www.japantimes.co.jp/life/2007/09/23/general/japanese-a-language-in-a-state-of-flux/ (18 August, 2020).
Rebuck, Mark. 2002. The function of English loanwords in Japanese. *NUCB Journal of Language Culture and Communication* 4(1). 53–64.
Reizei, Akihiko. 2020. Kiki-kan no hasshin ga katakana ni naru nihongo no fushigi. *News Week Japan*. https://www.newsweekjapan.jp/reizei/2020/03/post-1155.php (13 August, 2020).
Shibata, T. 1993. Gairaigo wa nihongo o midasuka. In Kawadeshobō (ed.), *Kotobadokuhon gairaigo*, 7–15. Tokyo: Kawadeshobō-shinsha.
Stanlaw, James. 1987. Japanese and English: Borrowing and contact. *World Englishes* 6(2). 93–109.
Stanlaw, James. 2004. *Japanese English: Language and Culture Contact*. Hong Kong: Hong Kong University Press.
Takashi, Kyoko. 1991. A sociolinguistic analysis of English borrowings in Japanese advertising texts. *World Englishes* 9(3). 327–341.
Takashi, Kyoko. 1992. Language and desired identity in contemporary Japan. *Journal of Asia Pacific Communication* 3(1). 133–144.
"Tokyo governor urges." 2020. Tokyo governor urges people to stay indoors over weekend as virus cases spike. *Japan Times*. https://www.japantimes.co.jp/news/2020/03/25/national/science-health/tokyo-logs-40-coronavirus-cases/ (30 August, 2021).
Tomari, J. 1985. Hanran suru katakanago. *Gengo* 14(9). 29–33.

Tomoda, Takako. 1999. The impact of loanwords on modern Japanese. *Japan Forum* 11(2). 231–253.

Tomoda, Takako. 2005. *The loanword (Gairaigo) influx into the Japanese language: Contemporary perceptions and responses*. University of New South Wales PhD dissertation.

Umesao, T., H. Kindaichi & S. Hinohara. 1989. Japanese as a vehicle for science. *Japan Echo* 16. 43–45.

Yahagi, Kunihiko. 2013. 「外国語使いすぎ」で NHK 提訴ーカタカナ語について考える [NHK sued for "overuse of foreign languages" - thinking about Katakana Kunihiko Yahagi]. *Yahoo! JAPAN*. https://news.yahoo.co.jp/byline/yahagikunihiko/20130627-00026016?fbclid=IwAR1ao78_cVt8pB9zpOOJKEFCOIFDiUdT9nC6vxuetk0LQdW1WtUV1b7LfRw (30 August, 2021).

Yamashita, Ryuichi. 2020. 河野氏「なんでカタカナ？」「クラスター」連発に苦言 [Kono's complaint "cluster" "Why Katakana?"]. *Asahi Shimbun*. https://www.asahi.com/articles/ASN3T660VN3TUTFK00N.html (30 August, 2021).

Yamashita, Y. & M. Katō. 2000. Korekarano nihongo ni eikyō o ataerumono wa? *Hōsōkenkyū-to-Chōsa* Sept. 48–67.

Yazaki, G. 1964. *Nihon no gairaigo*. Tokyo: Iwanamishoten.

Yoshimi, Mitsuo. 2018. 多すぎる英語のカタカナ表記 [Too much English katakana]. *Mainichi Shimbun*. https://mainichi.jp/articles/20180307/ddm/005/070/007000c (30 August, 2021).

SOAS Working Papers in Linguistics. Volume 20 (2021): 142–164

While we are asleep: Master/Mentor-Apprentice language learning initiative in the Ryukyus

Miho Zlazli

601339@soas.ac.uk

Abstract
I currently conduct participatory action research with new speakers of Ryukyuan languages by running a project called MAI-Ryukyus, which is designed based on Hinton's Master/Mentor-Apprentice Language Learning Program, to explore both emotional and cognitive aspects of learning one's own Indigenous ancestral tongue. The findings of the PhD research will be discussed in a future article. In this paper, I introduce my research design, discuss the issues identified in current language revitalisation efforts through ethnographic observation, and conclude with future directions.

Keywords: Indigenous transformative paradigm, language revitalisation efforts, Master/Mentor-Apprentice Language Learning Program, new speakers, Ryukyuan languages.

1. Imaginary paradise in Japan

Ryukyuan languages[1] are spoken in the Ryukyu[2] Islands which spread across six hundred miles in the north-western Pacific of South Japan (Kan 2011) (Figure 1). While the Ryukyu Islands have been a popular domestic destination for Japanese tourists where local people are assimilated to Japanese *to the right degree* with a touch of exoticism (Tada 2015), historical contexts of the Ryukyus and their current political affairs ("Tai Chugoku" 2021) arising from the US-Japan Status of Forces Agreement (MOFA 1960) are less known among them.[3]

For example, Irisuna Island was used as a symbol of imaginary southern paradise in a popular TV drama series called *Churasan*[4] (NHK 2001) (Figure 2) in the context of a so-called "Okinawa Boom" that has flooded Japanese popular culture and mass media since the mid-1980s or 1990s (Ina 2010; Murray 2017). However, it is hardly known (except among residents of the adjacent Tonaki Island) that Irisuna Island has been heavily used as a US Rifle Range following WWII (Figure 3). For the residents of Tonaki Island, Irisuna Island is an irreplaceable place that has sacred groves (see Figure 5, *utaki*).

Here, a repercussion of Japanese imperialism is identifiable. Inoue (2012) illustrated that pre-WWII colonial discourses were pervasive to the extent that they permeated into children's literature of the time. Kawamura (1992 as cited in Inoue 2012) exemplified a

[1] Ryukyuan languages belong to the Japonic language family alongside Japanese language (Pellard 2015).
[2] The name *Ryukyu* (琉球: Lewchew) was given by the Xuande Emperor of the Ming dynasty in 1430 (Lim 2016). I use the Japanese reading *Ryukyu* for the time being, following the current academic convention.
[3]I would like to thank community people for their insights, my PhD supervisors and colleagues for their support and guidance, and CHASE AHRC Studentship and *Gesellschaft für bedrohte Sprachen* project for funding my research.
[4] *Churasan* means 'pure' in Okinawa language.

novel about a Japanese boy's adventures on an exotic South Sea Island with the colonialist dichotomy of *civilization* versus *barbarism*. A similar configuration is still observable among contemporary tourists or immigrants who visit or relocate to the Ryukyu Islands (Sudo 2016).

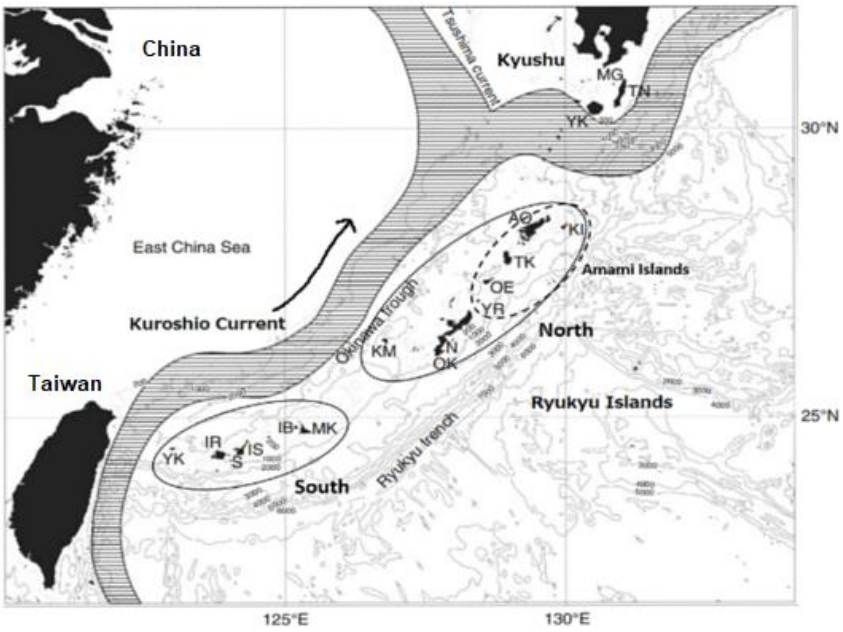


Figure 1. The Ryukyu Islands (adapted from Kan 2011).

Figure 2. A distant view of Irisuna Island that was used in the opening of the drama *Churasan* (adapted from Tonaki jima nikki 2013).

Figure 3. A US helicopter shooting rockets into Irisuna Island (adapted from 1st MAW Marines 2021). A link to the original video is available from the reference list.

2. The purpose of the research

The Indigenous territory of the Ryukyus is politically divided into Kagoshima and Okinawa Prefectures, which correspond to the geographic areas subdued by the Satsuma Clan from Japan in the early 17th century (the Amami Islands) and annexed to Japan in the late 19th century (the rest of the Ryukyus), respectively. On the other hand, the same territory is linguistically classified into the Northern and Southern regions, where Amami, Kunigami and Okinawa languages, and Miyako, Yaeyama and Yonaguni languages[5] are spoken respectively (Heinrich, Miyara and Shimoji 2015). Each language is spoken in a region of the same name except the Kunigami language which is spoken across a part of the Amami Islands and the northern end of Okinawa Prefecture. The narrow (Okinawa region) and broad (Okinawa Prefecture) definitions of Okinawa often create confusion between different contexts, such as identity and political discussions.

Following the assimilation policy imposed by the Meiji government of Imperial Japan as part of the Rich Nation, Strong Army campaign in the first half of the 20th century (Samuels 1994; Kondo 2014) and the subsequent internalised assimilation under the post-war US occupation (Masiko 2014), the Ryukyuan people have given up transmitting their mother tongue to younger generations. My impression is that, most traditional speakers are over eighty at the time of writing. According to Yokoyama (Yamada et al. 2018; Yokoyama and Kagomiya 2019), language comprehension seems maintained among people in their 40s and older but rapidly reduces among people under 40.

Despite the current language revitalisation efforts in the Ryukyus, new speakers[6] have not increased effectively. Prior to the fieldwork, I assumed that one of the main causes was that the efforts mainly focused on the cognitive aspect of language learning. However, as Swain (2013) emphasised, emotional and cognitive aspects of language learning are inseparable. Another cause seemed to be that they mainly targeted children at school (Okinawa Prefecture 2020). However, as Fishman (1991) identified, intergenerational language transmission mainly occurs in an immersive environment at home and in the community. In order to recreate such an environment, we need to fill the gap of *missing generations* between children and traditional speakers (Hinton and Meek 2018).

In the following subsections, I introduce my positionality and research questions, and discuss the implications of the research.

2.1. The author's positionality

I am an Indigenous[7] Ryukyuan researcher originally from the Okinawa region, which has been the political centre of both the former Ryukyu Kingdom and the current Okinawa

[5] The Ryukyuan languages have a few hundred 'regional lects' in total that correspond to their traditional community units (see Figure 5). I use 'regional lect' as a substitute for 'dialect' for the time being to eliminate the negative connotation associated with the term as a dialect of Japanese language (Clarke 2015). I use the Japanese readings for the region and language names for the same reason as in Footnote 2 and to avoid issues of power relations within Ryukyuan society by applying a reading in a specific regional lect, among others.

[6] Here, 'new speakers' means people who had little home or community exposure to the target endangered language but have acquired it through language revitalisation efforts (O'Rourke, Pujolar and Ramallo 2015).

[7] When the word 'Indigenous' is capitalised, it does not simply indicate that Indigenous peoples have unique history and relationship with their lands, territories and resources, but also that they are marginalised

Prefecture of Japan. I had an assimilated Japanese identity and a monolithic view on diverse Ryukyuan communities until I started my PhD research. I first joined the SOAS community as a master's student to get qualified as a Japanese language teacher to help heritage Japanese speakers, including my son. However, when I was reminded by linguists that I had my own ancestral tongue and it was endangered, I realised that anxiety and fear that I constantly had were deeply connected to the issue (Zlazli, forthcoming). This experience led me to the current research on language revitalisation of Ryukyuan languages to resolve emotional insecurity of the Ryukyuan people and help them recover agency in their life.

> "Indigenous language revitalizes us, not the other way around. If we take care of our language, it will take care of us. This is our wellbeing." (Galla and Goodwill 2017)

2.2. Research questions

To fill the gap of missing generations between children and traditional speakers, I explore:

(i) What motivates adult new speakers to speak Ryukyuan languages?
(ii) How can they effectively acquire Ryukyuan languages without compromising the diversity of Ryukyuan languages?

2.3. Implications of the research

The main stakeholders of current language revitalisation efforts, to my knowledge, are grassroots traditional speakers or communities (Anderson 2014; Hammine 2020b) and the Okinawa Prefecture, which does not have jurisdiction over the Amami Islands (Ishihara et al. 2019; Okinawa Prefecture 2020). In limited regions, grassroots new speakers (Sakihara and Oyakawa 2021; Okinawa Hands-On NPO, n.d.; Zlazli 2021) and grassroots researchers (Tohyama 2019; Matsuda and van der Lubbe 2020; Port Language Revitalization Project 2020; Yokoyama 2021) also actively engage in local language revitalisation efforts.

Given that language revitalisation is a newly emerged discipline that requires more empirical research and theorisation, collaboration between Indigenous peoples and academic researchers from a range of related disciplines is indispensable. However, community-researcher and interdisciplinary collaborations are not fully established in the Ryukyuan context due to mismatch between "well-defined short-term goals" by different stakeholders and "ill-defined long-term goals" as holistic language revitalisation (Madsen 2021).

Given that new speakers are the essential actors to maintain the language use in society (Hammine 2020a; Zlazli 2021), reframing language revitalisation efforts from their perspectives may have a potential to achieve: (i) the necessary collaboration among stakeholders with agreed long-term goals, and (ii) consensus of "ideological clarification" on what language revitalisation means (Kroskrity 2009; Heinrich 2011; Heinrich and Ishihara 2017).

---

in the mainstream of society (Johnson et al. 2007; UNDESA 2008). This notion has been developed in the historical context of settler colonialism as an international political category to seek for a higher authority beyond a nation to address their human rights issues (Merlan 2009).

3. The research design

Guba and Lincoln (Guba, Lincoln and Lynham 2017) identified that any researchers, whether knowingly or unknowingly, have a certain philosophical assumption on their axiology (what they value), ontology (how they view the nature of a reality or realities), epistemology (how they engage with the reality/realities), and methodology and methods (how they conduct research on the reality/realities). Therefore, it is crucial to examine whether the philosophical assumption that underpins the research paradigm (Figure 4) is coherent with the nature of the research questions. Otherwise, we could face a risk of having unreliable findings or an unethical research process.

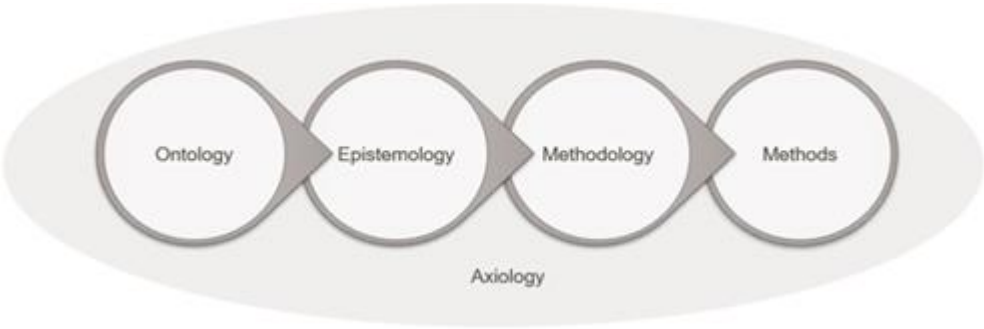


Figure 4. Elements of research paradigm.

3.1. Indigenous transformative paradigm

In the current research, I employ an Indigenous transformative paradigm based on Mertens' transformative approach (Widianingsih and Mertens 2019) and Wilson's (2008: 62) Indigenous research paradigm. It consists of the following elements.

Axiology: *Relational accountability* (Wilson 2008: 97) is of the utmost importance between: (i) stakeholders of language revitalisation efforts to prevent potential harm to those who are vulnerable in power relations,[8] and (ii) living people, the Indigenous land and sea, and ancestors and spiritual beings to appreciate the Indigenous knowledge system. An Indigenous knowledge system can be understood as a holistic system of knowledge that has developed over many generations through a complex fabric of practices and understandings in the corresponding Indigenous communities (Howden 2001).

Ontology: Voices of Indigenous peoples are marginalised in mainstream society (Johnson et al. 2007; UNDESA 2008), such as those of the Ryukyuan people in Japan. Layers of similar power relations are also observable between Okinawa and other Ryukyuan regions and between larger and smaller communities in the respective regions. As is the case with other Indigenous peoples (e.g., Henry and Pene 2001; Chilisa et al. 2016), the Ryukyuan people also have close-knit ties of kinship and relations (i) within their traditional community units that are uniquely situated on their Indigenous land and sea (Figure 5) and (ii) with their ancestors and spiritual beings to form a unique cosmological construction (Abe 2016).

[8] A same person can be both dominant and vulnerable in different aspects of power relations, e.g., Indigenous Okinawans are dominant to people from other regions of the Ryukyus, and non-Indigenous researchers can be vulnerable if they are women or in the early stage of their career.

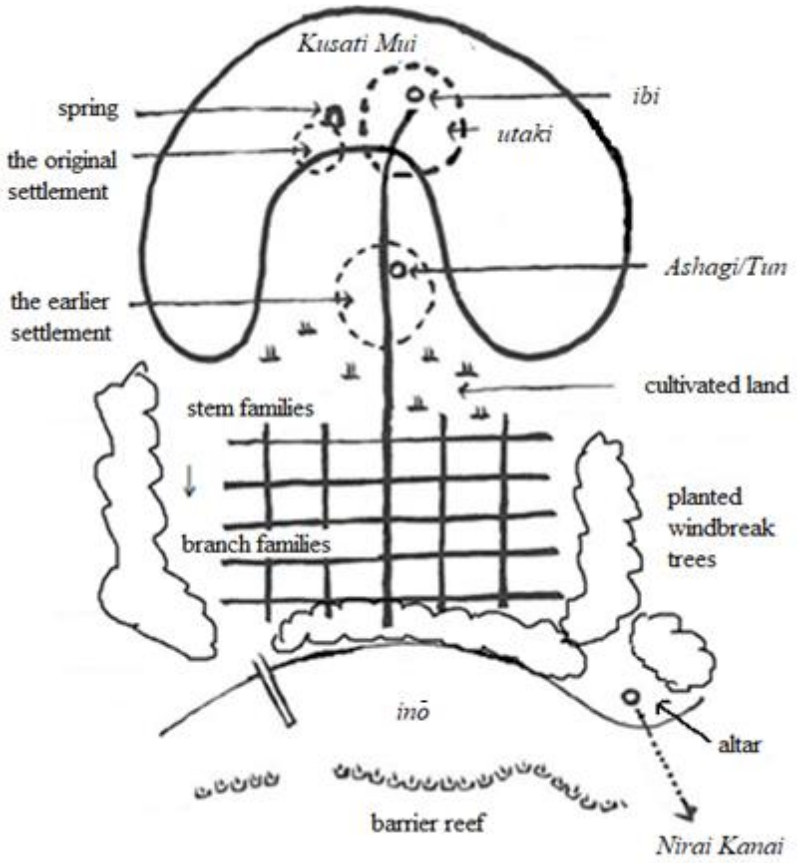


Figure 5. A traditional community unit in the Ryukyus (Adapted from Architectural Institute of Japan [1989] in Tomigusuku City 2020). *Kusati Mui*: Cuddling Forest, *utaki*: sacred grove, *ibi*: sanctuary, *Ashagi/Tun*: Worship House, *Nirai Kanai*: the Everlasting World, *inō*: coral reef lagoon.[9]

Epistemology: Researchers have the responsibility of *Primum non nocere* [First, do no harm] to the Indigenous communities through rigorous iterative reflection on their own positionalities (Manohar, Bhole and Arora 2017) (Table 1). Therefore, they are expected to identify marginalised voices of stakeholders, which are often invisible[10] to those in a dominant position and incorporate their voices into the language revitalisation efforts and future language planning. They are also expected to be aware of how they are positioned in the Indigenous knowledge system.

[9] The Cuddling Forest surrounds the community like Mother to protect it from typhoons in summer and north winds in winter alongside planted windbreak trees. The forest has a sacred grove (which has a sanctuary where their guardian deity descends and only priestesses are allowed to enter) and burial site where their ancestors rest. They believe the existence of the Everlasting World far off the coast to the south-east. It is the origin of all life, and life is eternal there. The forest, the coral reef lagoon, and the cultivated land provide resources to the community.

[10] Gohard-Radenkovic (2012) describes the state as *zone blanche*, which will be hidden or masked by *untold* or *overtold* stories. Indigenous people also feel difficult to speak their mind freely within their communities due to their close-knit complex relations.

Table 1: A checklist for researcher's positionality (partially adapted from Lin 2015; Manohar, Bhole and Arora 2017).

1. What kinds of interest motivate you to do research?
2. What kind of knowledge will you produce?
   2.1 What is the possible impact of the knowledge, and for whom?
   2.2 How will data be collected and disseminated?
   2.3 Who has the agency on the research process?
3. Is there any value- or interest-free research? Why/Why not?
4. What is your relationship with other stakeholders?
   4.1 How do you view yourself (e.g., political allegiance, religious faith, gender, sexuality, geographical location, race, culture, ethnicity, social class, age, linguistic tradition, personal experience)?
   4.2 How do you position other stakeholders, and vice versa?
   4.3 What possible impact does your personal position have on other stakeholders and the research context and process?
5. How are you positioned in the Indigenous knowledge system?

Methodology and Methods: With the abovementioned philosophical assumption, I conduct collaborative autoethnography (Chang, Ngunjiri and Hernandez 2016: 17) with stakeholders of language revitalisation to unearth marginalised voices among them[8] and incorporate the voices into future action plans to eliminate power disparities. The research process is expected to agree with Indigenous ways of living to keep harmony with both Indigenous peoples and their ancestors and spiritual beings.

3.2. MAI-Ryukyus project

Based on the research questions and the Indigenous transformative paradigm, I designed a participatory action research project MAI-Ryukyus (Figure 6) with reference to Hinton's Master/Mentor-Apprentice Language Learning Program (Hinton et al. 2018). MAI stands for Master/Mentor-Apprentice Initiative. In this project, adult new speakers (equivalent to "Apprentice" in Hinton's program) are widely recruited online. They are mainly self-identified Ryukyuans including myself, but other people are also welcome if they are interested in the Ryukyuan people, not only their languages, to prevent cultural appropriation. They will take initiative to spend time with traditional speakers ('Master/Mentor') in daily life settings or online to elicit an immersive environment of their target Ryukyuan language to acquire it. They have access to a peer support network like *Thirdspace* (Soja 1996) to explore their new language practice and pluralistic identity (Ting-Toomey 2015). They will also receive specialist support as required (e.g., introductory sessions to practice Master/Mentor-Apprentice interactions and learn how to use existing linguistic resources). New speakers are also encouraged to take initiative in wider language revitalisation efforts as they come up with new ideas.

I have been documenting the trajectories of new speaker participants' language acquisition[11] since the summer of 2019. As positive and negative outcomes surface one after another, we (new speakers and I) analyse the causes and explore better practices

[11] Austin and Sallabank (2018; Austin 2020) argue the importance of documenting the process of language revitalisation efforts, e.g., decision-making, events, success, and failure, so that other people can learn from the experience.

through iterative trials and errors. I plan to theorise effective approaches to language learning in the Master/Mentor-Apprentice approach at the end of the PhD project.

The findings of this research will be discussed in a future article. In this paper, I discuss issues identified in the current language revitalisation efforts through ethnographic observation and conclude with future directions.

Figure 6. MAI-Ryukyus project.

4. Current issues

The most prominent issue identified in the current language revitalisation efforts during my fieldwork was a controversial attempt to promote language revitalisation while waiting for the completion of cultural assimilation because most efforts have not addressed the ongoing impact of internalised assimilation on traditional speakers and new speakers of Ryukyuan languages, which has been the very reason of their language shift.

Paradigmatic issues were also identified on a personal level. Most stakeholders, including myself, seem to remain on the subjective perspective (Figure 7), with which we are prone to react without considering relations that we are embedded in. It has been causing deadlocks that prevent us from collaboration among stakeholders. I argue that we need more awareness-building dialogues among ourselves based on active listening of other parties and sharing honest opinions (Hạnh 2013) to gain the relative perspective underpinned by careful examination of our positionalities (Table 1). From there, we can explore the transformative perspective to comply with relational accountability in the Indigenous transformative paradigm. Tolerant dialogues over an extended period of time might be needed because paradigm shift is often out of one's comfort zone (Marcum 2013).

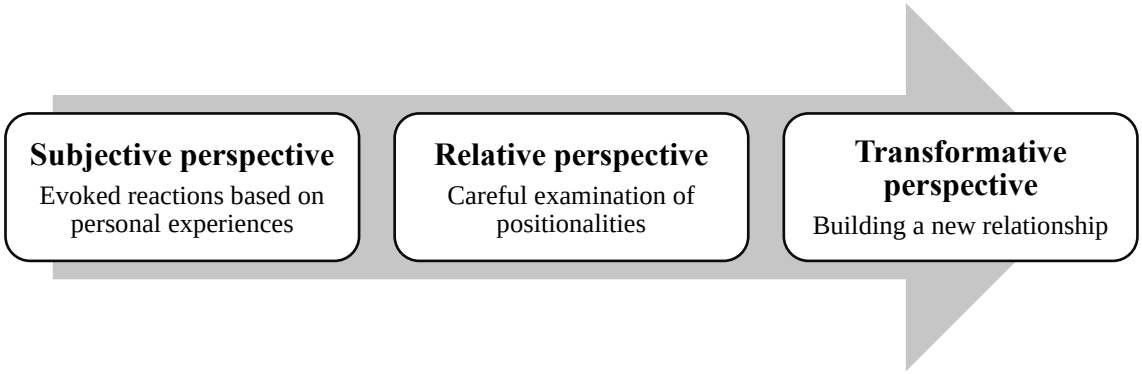


Figure 7. A required paradigm shift on a personal level.

4.1. Unconscious bias in Japan

In order to discuss the internalised assimilation of the Ryukyuan people, we first need to address their Indigenous rights (UN 2017).[12] However, such discussions tend to be avoided as being politically sensitive among many researchers who work with the Ryukyuan people in Japan (personal communications, 2019-2020):

> Researcher 1 (Japanese):
> "As a team, we conduct both language documentation and language revitalisation … When I interact with community people, I often call their language as 'dialect' because they call their own language as a 'dialect' [of Japanese]. As a linguist, I also use the term 'dialect' as a dialect of a specific Ryukyuan language. I try not to get involved in the political discussion [of the term 'dialect'] because it is a sensitive problem. I respect the decision of the community people."
>
> Researcher 2 (Japanese):
> "I wondered if you [the author] were Japanese or not because your family name was foreign, and you wore a [Muslim] headscarf. … Isn't it quicker to just begin [teaching the community people how to learn and document their Ryukyuan languages] rather than having disputes over political issues?"
>
> Researcher 3 (Ryukyuan):
> "When I was a graduate student, I got hurt when a person who studied abroad criticized me for not calling our ancestral tongue as 'language' and criticized that I was poisoned by the way of thinking in the mainland Japan. Now I try to express the complexity of my suffering somehow, thinking that it's okay to have a marginalised anguish because research could be a means of artistic expressions like poetry."

Researcher 1 had a positivist assumption that he could be politically neutral (Rubin and Rubin 2011: 16) in his relationship with the community people, while Researcher 2 ideologically contested if I (Indigenous Ryukyuan) would satisfy her homogeneous

[12] The Ryukyuan people satisfy the definition of Indigenous peoples (UNDESA 2008: 8), but the Japanese government does not recognise them to be Indigenous people (Japanese Language Division of the Agency for Cultural Affairs, personal communication, March 2019).

expectation of being Japanese (Sugimoto, 2019). Such practices as a person in a dominant position both as a researcher and a Japanese person may potentially prolong the assimilative impact on the Ryukyuan people, but both of them were not aware of it.[13] On the other hand, Researcher 3 expressed her struggles with mismatch between being Indigenous and her familiar research practices in Japan.

Imposing assimilative pressure on Indigenous persons supresses their ancestral tongue under the state of *diglossia* (Maher 2019), which is counter-effective to language revitalisation (A in Figure 8). As the notion of "Indigenous peoples" has been developed as an international political category to seek for a higher authority beyond nations to claim their neglected collective rights (Merlan 2009), the knowledge system of Indigenous communities should be acknowledged as being well beyond the system within the assimilative national ideology (Sugimoto, 2019). In this way, non-hierarchical translanguaging (Li Wei 2018) and negotiation of pluralistic identity (Ting-Toomey 2015) become possible, which works in favour with language revitalisation (Zlazli, forthcoming) (B in Figure 8).

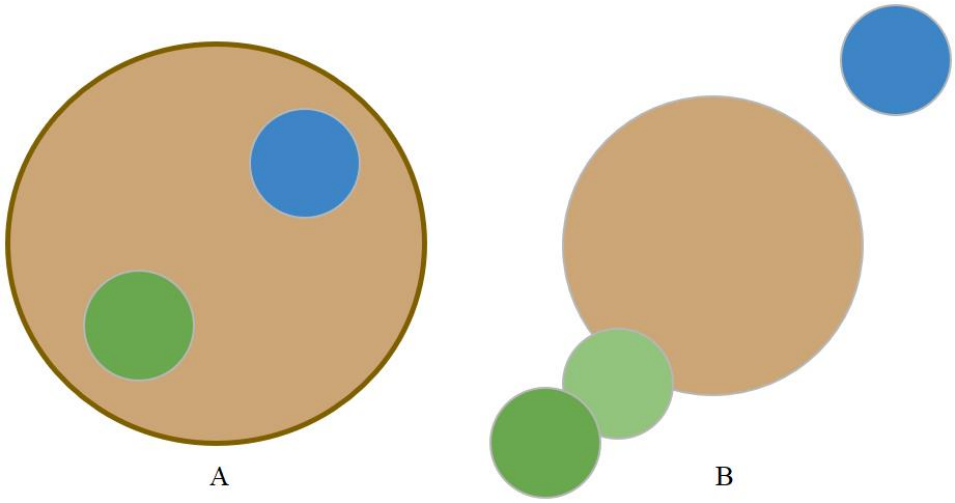


Figure 8. Perspectives within (A) and beyond (B) the homogeneous national ideology in Japan. A: Under an exclusive and assimilative ideology (Sugimoto 2019), the use of minority languages including Indigenous languages will be supressed under the state of *diglossia* (Maher 2019). B: By acknowledging that the knowledge system of the minority-language-speaking communities is well beyond the system within the assimilative ideology, non-hierarchical translanguaging (Li Wei 2018) and negotiation of pluralistic identity (Ting-Toomey 2015) become possible.

A similar power relation was observed with the author's positionality. Prior to the fieldwork, I conceived an idea that I was an Indigenous Ryukyuan person from the literature review during my MPhil period. Carrying the attitude that we are the *same* Ryukyuan people as a dominant Okinawan person exerted unexpected assimilation

[13] They have significantly contributed to the communities who they have worked with for more than a decade (Shimoji 2020). I have not had opportunities to conduct ethnographic research within their communities due to the pandemic of COVID-19. Therefore, the discussion is limited to my impression as an Indigenous person and based on feedback from collaborative autoethnography with Ryukyuan people from several other regions.

pressure on Ryukyuan people from other regions. An Indigenous researcher from another Ryukyuan region once warned me that I was potentially breaching research ethics, but it took her more than a year to speak it up to me. I realised how difficult it could be for marginalised stakeholders to disclose their concerns and for researchers to identify their own unconscious bias. In order to create a safe space where community people feel more comfortable to express their voices, researchers should carefully reflect on their own positionality (Table 1).[14]

In contrast, Indigenous new speakers who were motivated to learn Ryukyuan languages *for themselves* beyond communicative needs in the workplace or care homes *for traditional speakers* had a rediscovered Ryukyuan identity or a regional identity distinct from other regions (interview data, 2020-2021):

> New Speaker 1 (Yaeyama language, early-20s):
> "I used to want to leave the small islands where I grew up to see the world. Then, I had frequent opportunities to look at my home islands from outside through such as internship and volunteering in Southeast Asian countries and the Loochoo[15] Identity Summit in Hawai'i, which made me realise the significance of my own roots. In parallel, I was seriously concerned when I learned at university that our history and culture were about to be eliminated by colonialism.[16] These motivated me to learn Yaeyama language, especially my family's regional lect."[5]

> New Speaker 2 (Miyako language, early-20s):
> "I used to try to reduce my accent in Japanese to sound like a Tokyoite when I was in secondary school in Miyako. Now, I read history at university in Tokyo. At the university library, I found books authored by Ryukyuan scholars. I read them through and got intrigued. I also had opportunities to meet researchers including an anthropologist who conducted fieldwork in my own community. While I had to stay on my home island during the last academic year due to the pandemic, I had a chance to experience a compilation work of community people's life, rituals, and memories as a member of our community's 300th anniversary committee. I realise that our traditional practices are on the verge of disappearance due to marginalisation caused by layers of power relations in Miyako, the Ryukyus, and Japan. Alongside our culture and history, our regional lect[5] in Miyako is also precious to me now."

[14] Madoka Hammine (personal communication, August 2021) observed the importance of researcher-community collaboration in revitalisation efforts in other Indigenous languages communities, in which being aware of one's own positionality was a prerequisite.
[15] Another reading for 'Ryukyu' chosen by some Ryukyuan people.
[16] Ono (2015) argues that Okinawan protesters have criticised the current Japanese political system as domestic colonialism in search of a breakthrough for political reform.

New Speaker 3 (Kunigami language, late-20s):
"Although I couldn't clearly grasp the meaning and I was not particularly interested in the language, I'd had opportunities to hear the Northern Okinawan language[17] since my childhood such as in conversations between relatives. I loved local events, place names, and Okinawan culture in general. Later in my life, I was inspired by a novel written by Shun Medoruma, in which he incorporated the Northern Okinawan language to characters' speech lines. I felt a strong attachment to the Northern Okinawan lines while I wasn't really moved by materials written in Shuri/Naha Okinawan language.[18] It urged me to relearn the language I'd known since I was young."

New Speaker 4 (Okinawa language, mid-40s):
"I was brought up in the Okinawan diaspora in Osaka, Japan. I had many classmates with Okinawan roots. We actively performed Okinawan traditional performances at festivals held by a neighbourhood association. We even had our original design of *Paarankuu*[19] [Okinawan hand-held drum] cut out of *Wadaiko* [Japanese drum] because we didn't have easy access to original ones from Okinawa at that time. So, it's kind of iconic to our diaspora.

When I moved to a different place for university, I felt different from others. I felt culturally appropriated when *Okinawa Boom* hit mainland Japan. I also had inner conflicts, feeling powerless about Japanese people's apathy to the way contradictions in Okinawa were broadcast in Japan (such as a campaign against the US Futenma base following the 1995 Okinawa rape incident by US servicemen; a new US Henoko base construction under the name of US Futenma base removal; and the 2004 US military helicopter crash at Okinawa International University).

I was inspired by authors and activists who addressed these issues, and I realised that I should also be the one to change the situation, not someone else. Later in my life, I started performance arts alongside work to explore what I can do, and I am further inspired by other performance artists from across the world. I want to speak Okinawan because it's also my language."

4.2. Ryukyuan identity

Due to the assimilative ideology justified by social evolutionist discourses developed in the context of Japanese imperialism (Meyer 2007; Heinrich 2012; Inoue 2012), the Ryukyuan people have had emotional insecurity with their own identity (Zlazli, forthcoming). They also lack a unified Ryukyuan identity and attribute themselves to layers of smaller social groups with corresponding vernacular language practices. There is a sense of rivalry among groups, on which political division into Kagoshima and Okinawa Prefectures also casts a shadow (Kiyama 2008). Due to layers of complex power

[17] "The Northern Okinawan language" indicates Kunigami language.
[18] Shuri/Naha lect is a regional lect spoken in the political centre of the Ryukyus. Shuri lect was also a sociolect spoken by aristocrats and royals in the former Ryukyu Kingdom.
[19] The origin of *Paarankuu* is said to be Octagonal Drum (八角鼓) which was brought from China (Lim 2016; *Ryukyu Shimpo* 2003) during the Ryukyu Kingdom era.

relations, the Ryukyuan people struggle to come to a consensus on being a unified Indigenous people who are entitled to claim their Indigenous rights (UN 2017).[20]

The Japanese government uses a rhetoric that they cannot force the Ryukyuan people to be Indigenous while they are voluntarily assimilated to Japan without acknowledging their historical responsibility.[21] This attitude also coincides with Researcher 1's attitude (see Section 4.1). While we are asleep, we are missing the opportunity to seek a higher authority *beyond the nation* to claim our rights that we have been fighting for *within the nation*'s framework (Inoue 2004; Nakashima 2010) or to convince external researchers to support our Indigenous rights.

The fact that we lack representative bodies of the Ryukyuan people also contributes to Okinocentrism. People from Okinawa region generally use "Okinawa" and "the Ryukyus" interchangeably and often forget the existence of the Amami Islands in Kagoshima Prefecture, while Ryukyuan people from other regions clearly distinguish Okinawa from the other Ryukyuan regions. For example, Okinawa independence movements under the name of "the Ryukyus" (Ginoza 2015) are often criticised for their self-serving agendas that do not represent voices of other Ryukyuan regions (community people, personal communications, 2019-2021). In terms of Ryukyuan languages, some activists have attempted sole promotion of Okinawa language over other Ryukyuan languages and standardisation of Okinawa language based on Shuri/Naha lect[18] for the sake of economic efficiency (Sato 2020). However, as seen in the interviews of new speakers (see Section 4.1), the diversity of the Ryukyuan languages should not be actively compromised. Otherwise, such a language policy will impact on people from other Ryukyuan regions as double assimilation to Japan and Okinawa.

In addition to the diversity within the Ryukyus, we also have diverse Ryukyuan diasporas across the world[22] (Kondo 2014; Maeda 2014; Yomitan Village History Editing Room 2021) and mix-roots people who self-identify as Ryukyuan. Along with the fact that the Ryukyuan knowledge system is gradually disappearing due to the ongoing assimilation to Japan or mainstream societies where Ryukyuan diasporas are embedded in, negotiating contingent pluralistic identity may help the Ryukyuan people to realise a new efflorescent Ryukyuan society (Roche, Maruyama, and Virdi Kroik 2018; Zlazli, forthcoming), which may provide opportunities to create new social domains for using Ryukyuan languages.

[20] Vasiliki Vita (personal communication, July 2021) argues that groups of people who have certain similarities do not necessarily need to be grouped under a unified identity, which could be an internalised ideology that should be unlearned because it relates back to a vicious circle of an oppression setting, such as the Japanese state promoting the Ryukyuan people's voluntary assimilation to Japan. I agree with her point because in fact the Ryukyuan people, who I call so collectively, have never been unified except being ruled under the former Ryukyu Kingdom for different periods of time at slightly different times. I have had a decentralised vision of the unified Ryukyuan people's network (B in Figure 9) (Section 5), but I should explore what community people potentially wish for their future.

[21] Choi (2003) argues that Japan's impunity for negating its historical imperialism is supported by Western industrialised nations for the current economic benefits.

[22] In statistical data, they are often classified as Japanese or citizens of Okinawa or Kagoshima Prefecture because mass emigration began following the annexation to Japan or the abolition of feudal domains and establishment of prefectures.

4.3. Linguistic fieldwork and language revitalisation

Most linguists who work with the Ryukyuan people were mainly trained in Japanese research institutions which provide only linguistic fieldwork trainings (NINJAL 2016; ILCAA 2021) without giving considerations on researcher's positionality (Table 1) or Indigenous research paradigm (Wilson 2008).

In traditional academic discourses in Japan, a social evolutionist belief was pervasive among researchers who worked with Indigenous peoples (Ikeda 2020; Rots 2019). While the use of Ryukyuan languages across social domains had rapidly declined under the policy of progressive assimilation, self-serving linguists extensively collected linguistic data across the whole of the Ryukyu Islands, with their own version of justification in terms of "community causes", i.e., creation of "comprehensive" linguistics resources for future generations of descendants of the current speakers who might be ready to utilise them in some unspecified way (Karimata 2003). Their agenda also brought with it a *fly-in-fly-out* method and attitude (Austin 2013; Hokuto City Library 1998) which has made community people feel that their knowledge was being culturally appropriated and led them to be wary of researchers (community people, personal communications, 2019 to 2021).

Leonard (2018) argues the importance of decolonization in language documentation, and Austin (2020) and Bowern (2011) emphasise careful consideration on designing a language documentation project to meet the community's expectations. Now, the global knowledge in this area is accumulating (e.g., Cruz and Woodbury 2014; Fitzgerald 2018; 2020; Genee and Junker 2018). However, along with researchers from the Global South (Sanders 2020), I argue that many researchers in Japan (including myself until recently) struggle to keep updated with the latest global discussions, partially due to their negative attitude towards the hegemony of English across the Global North (Ning 1997; Macedo, Dendrinos and Gounari 2016).

In contrast, they did not question the notion of selecting Japanese over other languages that minority language communities comprehend in the creation of bilingual resources, whether they be languages spoken in Japan or elsewhere.[23] This could also be the case with bilingual resources of Ryukyuan languages for non-Japanese speaking Ryukyuan people, e.g., Ryukyuan diasporas.

There are researchers in the early stages of their careers who have concerns over traditional practice, but they struggle to make themselves heard in the mainstream of practice due to rigid, gendered, hierarchical human resource management systems in Japanese society (Froese, Sekiguchi and Maharjan 2018).

4.4. Emotional resistance

I wrote that a paradigm shift is often out of one's comfort zone (see Section 4). Even if a person rationally understands its importance, implementing it could be emotionally

[23] Researchers at ILCAA (Tokyo University of Foreign Studies) are typical for adopting this view, e.g., they published a bilingual dictionary of the Chinese minority language Eynu with glosses only in Japanese and a large two-volume Mon-Japanese dictionary written entirely in Mon script (with no phonetic representation) so that only people literate in Mon AND Japanese would be able to use the work (Peter Austin, personal communication, April 2021).

challenging if the person struggles to overcome an associated psychological pain. Therefore, we need tolerant dialogues to emancipate psychological pain or trauma of stakeholders alongside those of traditional and new speakers of Ryukyuan languages.

For example, I often struggle to admit my positionality in a complex social network of stakeholders. I once overreacted to an external researcher who had a longstanding researcher-community collaboration in one Ryukyuan region when she called me an outsider to the community. While I believe that she should not have called the Indigenous person as an outsider on behalf of the community people, my reaction was mainly triggered by my family's psychological trauma of ostracism as minority Christians from our local community (Zlazli, forthcoming).

Researcher 2 (see Section 4.1) also struggles to start examining her own positionality because she fears that her decadelong hard work in the field (e.g., investing a significant amount of her own money, effort and time, building the researcher-community collaboration from the scratch, her passion devoted in the creation of a reference grammar which did not exist before, and so on) might be cancelled to be null by "beautifully theorised arguments" by a person who does not know the trajectory of her work once she admits that she is an external researcher.

On the other hand, I also witnessed some community people whose psychological trauma was topped up with abhorrence against researchers that was induced by sociologists' criticism on traditional practices of linguistic fieldwork mentioned above in Japan. When disseminating research findings, we need careful consideration of the topic, its depth, and the pitch so that the discourse can promote prospective community-researcher collaborations.

5. Future directions

On completion of my PhD research, I plan to establish a non-governmental organisation based on the MAI-Ryukyus project to provide long-term coherent language planning support, including promotion of researcher-community collaboration and creation of Massive Open Online Courses (MOOCs) on Indigenous knowledge system with a multidisciplinary team.[24] The organisation might also function as a representative body of the Ryukyuan people in future (B in Figure 9).[20] As the Māori people have established their own research paradigm (Bishop 1998; *Rangahau*, n.d.), I am also interested in exploring a Ryukyuan research paradigm to develop a more suitable research methodology for the Ryukyuan people based on the Indigenous transformative paradigm (Section 3.1).

While descriptive linguists collect and elicit empirical linguistic data to create comprehensive reference grammars of Ryukyuan languages *as a system* (Dixon 2010; Aikhenvald 2014) and build centralised comprehensive digital dictionaries (Carlino and Shimoji 2021), new speakers can also contribute to language documentation by accumulating discourses of Ryukyuan languages *as a practice* (Austin and Sallabank 2018; Heinrich 2018).

[24] The MOOCs on Indigenous knowledge system can be incorporated into the existing framework of *Shimakutuba Kentei* [Community language examinations] provided by the Shimakutuba Fukyu Center (2018), Okinawa Prefecture.

If community-researcher collaboration is carefully designed based on the Indigenous transformative paradigm, such a symbiotic collaboration may bring a bottom-up transformative experience to both parties who will coevolve for mutual benefit (Viswanathan et al. 2021) to create comprehensive linguistic resources for future generations. Such collaboration may also advocate for Indigenous peoples to become Peoples, lessen the current dichotomous *insider/outsider* conflicts to embrace *stakeholder spectrum with different positionalities* who are embedded in a flexible distributed network in language revitalisation (C in Figure 9) while maintaining the cosmos of traditional community units (A in Figure 9). Each stakeholder from a different background will bring a unique perspective to language revitalisation efforts in a specific community, and the efforts will provide a unique insight back to the stakeholder who can reflect it to their relationship with their community of origin.

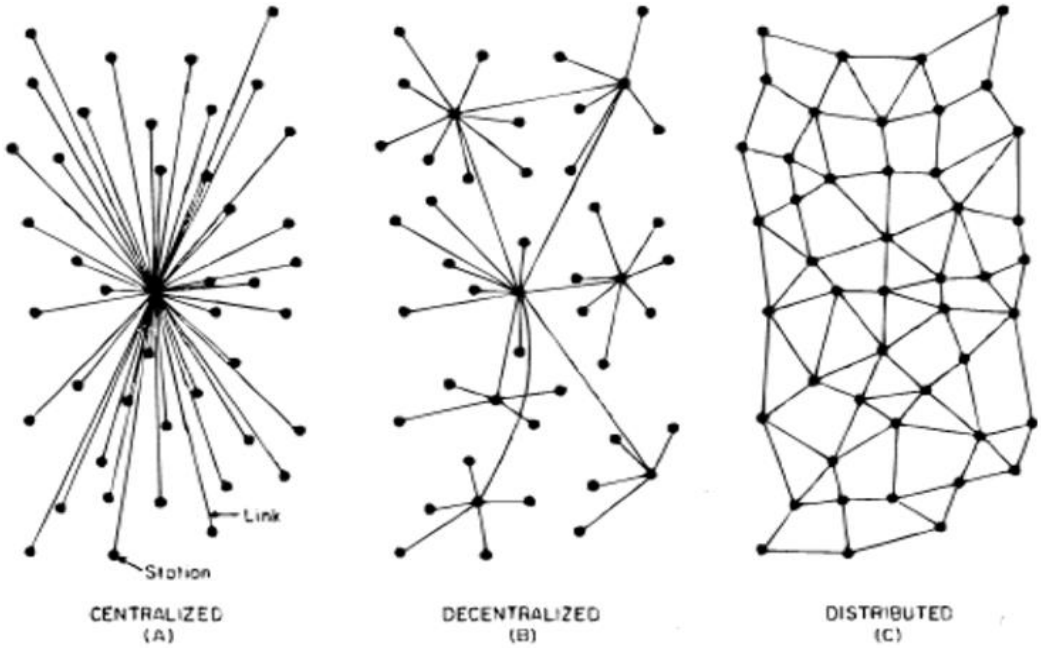


Figure 9. The configuration of traditional community units (A); representative bodies of the Ryukyuan people (B);[20] and stakeholders of language revitalisation efforts (C) (adapted from Liacas 2019).

6. Conclusion

In this article, I first introduced my PhD research project MAI-Ryukyus which is designed based on Hinton's Master/Mentor-Apprentice Language Learning Program, and discussed issues identified in the current language revitalisation efforts through ethnographic observation. The most prominent issue was a controversial attempt to promote language revitalisation while waiting for the completion of cultural assimilation. Paradigmatic issues were also identified on a personal level, which prohibited effective collaboration among stakeholders based on relational accountability and the Indigenous transformative paradigm. The author plans to establish a non-governmental organisation based on the MAI-Ryukyus project to provide long-term coherent language planning support.

References

1st MAW Marines. 2021. *Video of HMLA-469 conducting rocket and minigun live fire off the coast of Okinawa, Japan late last year!* Twitter. https://twitter.com/1stMAW_Marines/status/1377361306328518657 (27 April, 2021).

Abe, Tsukasa. 2016. *Seiji Jinruigaku no Riron to sono Gendai teki Igi [The theories of political anthropology and its significance in the contemporary world]*. Meiji University PhD Thesis.

Aikhenvald, Alexandra Y. 2014. *The art of grammar: A practical guide*. Oxford: Oxford University Press.

Anderson, Mark. 2014. Revitalisation attempts and language attitudes in the Ryukyus. In Mark Anderson & Patrick Heinrich (eds.), *Language crisis in the Ryukyus*, 1–30. Newcastle upon Tyne: Cambridge Scholars Publishing.

Austin, Peter K. 2013. Language documentation and meta-documentation. In Mari C Jones & Sarah Ogilvie (eds.), *Keeping languages alive: Documentation, pedagogy and revitalization*, 3–15. Cambridge: Cambridge University Press.

Austin, Peter K. 2020. Language documentation and language revitalization. In Justyna Olko & Julia Sallabank (eds.), *Revitalizing endangered languages: A practical guide*, 199–219. Cambridge: Cambridge University Press.

Austin, Peter K. & Julia Sallabank. 2018. Language documentation and language revitalization: Some methodological considerations. In Leanne Hinton, Leena Huss & Gerald Roche (eds.), *The Routledge Handbook of Language Revitalization*, 207–215. Routledge.

Bishop, Russell. 1998. Freeing ourselves from neo-colonial domination in research: A Maori approach to creating knowledge. *International Journal of Qualitative Studies in Education* 11(2). 119–219.

Bowern, Claire. 2011. Planning a language documentation project. In Peter K Austin & Sallabank Julia (eds.), *The Cambridge Handbook of Endangered Languages*, 459–482. Cambridge: Cambridge University Press.

Carlino, Salvatore & Michinori Shimoji. 2021. Nichi-Ryu Shogo Hogen no Denshi Jisho no Kochiku [Construction of an electronic dictionary of dialects of Japonic languages]. Presented at the 162nd Meeting of the Linguistic Society of Japan, Online.

Chang, Heewon, Faith Ngunjiri & Kathy-Ann C. Hernandez. 2016. *Collaborative autoethnography*. Routledge.

Chilisa, Bagele, Thenjiwe Emily Major, Michael Gaotlhobogwe & Hildah Mokgolodi. 2016. Decolonizing and indigenizing evaluation practice in Africa: Toward African relational evaluation approaches. *Canadian Journal of Program Evaluation* 30(3). 313–328.

Choi, Jung-Bong. 2003. Mapping Japanese imperialism onto postcolonial criticism. *Social Identities* 9(3). 325–339.

Clarke, Hugh. 2015. Language and identity in Okinawa and Amami: Past, present and future. In Patrick Heinrich, Shinshō Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan Languages: History, structure, and use*, 631–647. Berlin: De Gruyter Mouton.

Cruz, Emiliana & Anthony C Woodbury. 2014. Collaboration in the context of teaching, scholarship, and language revitalization: Experience from the Chatino Language Documentation Project. *Language Documentation & Conservation* 8. 262–286.

Dixon, R. M. W. 2010. *Basic Linguistic Theory: Methodology*. Vol. 1. Oxford: Oxford University Press.

Fishman, Joshua, A. 1991. *Reversing language shift: Theoretical and empirical foundation of assistance to threatened languages* Clevedon: Multilingual Matters.

Fitzgerald, Colleen M. 2018. Creating sustainable models of language documentation and revitalization. In Shannon T. Bischoff, and Carmen Jany. (eds.), *Insights from practices in community-based research: From theory to practice around the globe,* 94–111. Berlin: De Gruyter Mouton.

Froese, Fabian Jintae, Tomoki Sekiguchi & Mohan Pyari Maharjan. 2018. Human resource management in Japan and South Korea. In Fang Lee Cooke & Sunghoon Kim (eds.), *Routledge Handbook of Human Resource Management in Asia*. Taylor & Francis.

Galla, Candace Kaleimamoowahinekapu & Alanaise Goodwill. 2017. Talking story with vital voices: Making knowledge with indigenous language. *Journal of Indigenous Wellbeing* 2(3). 67–75.

Genee, Inge & Marie-Odile Junker. 2018. The Blackfoot Language Resources and Digital Dictionary project: Creating integrated web resources for language documentation and revitalization. *Language Documentation & Conservation* 12. 274–314.

Ginoza, Ayano. 2015. Intersections: Dis/articulation of ethnic minority and indigeneity in the decolonial feminist and independence movements in Okinawa. *Intersections: Gender and Sexuality in Asia and the Pacific* (37).

Gohard-Radenkovic, Anine. 2012. Kokusai teki na Ido no naka ni aru Akuta tachi no aratana Soten to Senryaku [New contentions and strategies for actors in international migration]. In Hideo Hosokawa (ed.), *Gengo Kyoiku to Aidentiti: Kotoba no Kyoiku Jissen to sono Kanosei [Language education and identity - language education practice and its potential]*. Yokohama, Japan: Shumpusha Publishing.

Guba, Egon G., Yvonna S. Lincoln & Susan A. Lynham. 2017. Paradigmatic controversies, contradictions, and emerging confluences, revisited. In Norman K. Denzin & Yvonna S. Lincoln (eds.), *The SAGE Handbook of Qualitative Research*, 108–150. 5th edn. SAGE Publications.

Hammine, Madoka. 2020a. Educated not to speak our language: Language attitudes and newspeakerness in the Yaeyaman language. *Journal of Language, Identity & Education* 1–15.

Hammine, Madoka. 2020b. Framing indigenous language acquisition from within: An experience in learning and teaching the Yaeyaman language. *The Language Learning Journal* 48(3). 300–315.

Hạnh, Thích Nhất. 2013. *The art of communicating*. London: Rider.

Heinrich, Patrick. 2011. Difficulties of establishing heritage language education in Uchinaa. In Heinrich, Patrick & Galan, Christian (eds.), *Language life in Japan: transformations and prospects*, 34–49. London; New York: Routledge.

Heinrich, Patrick. 2012. *The making of monolingual Japan. Language ideology and Japanese modernity*. Bristol: Multilingual Matters.

Heinrich, Patrick. 2018. Revitalization of the Ryukyuan languages. In Leanne Hinton, Leena Huss & Gerald Roche (eds.), *The Routledge Handbook of Language Revitalization*, 455–463. Routledge.

Heinrich, Patrick & Ishihara, Masahide. 2017. Ryukyuan languages in Japan. In Corinne A Seals & Sheena Shah (eds.), *Heritage language policies around the world*, 165-184. London: Routledge.

Heinrich, Patrick, Shinsho Miyara & Michinori Shimoji. 2015. Introduction: Ryukyuan languages and Ryukyuan linguistics. In Patrick Heinrich, Shinshō Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan Languages: History, structure, and use*, 1–10. Berlin; Boston: De Gruyter Mouton.

Henry, Ella & Hone Pene. 2001. Kaupapa Maori: Locating indigenous ontology, epistemology and methodology in the academy. *Organization* 8(2). 234–242.

Hinton, Leanne, Margaret Florey, Suzanne Gessner & Jacob Manatowa-Bailey. 2018. The Master-Apprentice Language Learning Program. In Leanne Hinton, Leena Huss & Gerald Roche (eds.), *The Routledge Handbook of Language Revitalization*, 127–136. New York: Routledge.

Hinton, Leanne & Barbra A. Meek. 2018. Language revitalization in Indigenous North America. In Leanne Hinton, Leena Huss & Gerald Roche (eds.), *The Routledge Handbook of Language Revitalization*, 375–383. Routledge.

Hokuto City Library. 1998. Kindaichi Haruhiko Kinen Toshokan Aakaibu [Kindaichi Haruhiko Memorial Library Archive]. https://trc-adeac.trc.co.jp/WJ11C0/WJJS02U/1920905100 (26 April 2021).

Howden, Kristin. 2001. Indigenous traditional knowledge and native title. *University of New South Wales Law Journal* 24(1). 60–84.

Ikeda, Mitsuho. 2020. Senjumin no Ikotsu to Fukusohin no Henkan ni tsuite: Dare ga Bunka Isan to Songen wo Shoyu (Hoyu) suru noka [Repatriation of human remains and burial materials of Indigenous peoples in Japan: Who owns their cultural heritages and dignity?]. *Co*Design* 7. 1–16.

ILCAA. 2021. Researchers training workshop. *Linguistic Dynamics Science 3 (LingDy3)*. https://lingdy.aa-ken.jp/activities/training-ws/210819-flws-ninjal/ (27 July 2021).

Ina, Hein. 2010. Constructing difference in Japan: Literary counter-images of the Okinawa boom. *Contemporary Japan*. Routledge 22(1–2). 179–204.

Inoue, Kota. 2012. A little story of settler colonialism: Imperialist consciousness and children's literature in the 1920s. In Michele Mason & Helen Lee (eds.), *Reading colonial Japan: Text, context, and critique*, 188–208. Stanford University Press.

Inoue, Masamichi. 2004. The community of subjects, the power, and the public-sphere of citizens: a new stage of the theory and discourse of appropriation and the U.S. base problems in Okinawa. *Japanese Journal of Ethnology* 68(4). 534–554.

Ishihara, Masahide, Katsuyuki Miyahira, Gijs van der Lubbe & Patrick Heinrich. 2019. Ryukyuan sociolinguistics. In *Routledge Handbook of Japanese Sociolinguistics*. Routledge.

Johnson, Jay T., Garth Cant, Richard Howitt & Evelyn Peters. 2007. Creating anti-colonial geographies: Embracing Indigenous peoples' knowledges and rights. *Geographical Research* 45(2). 117–120.

Kan, Hironobu. 2011. Ryukyu Islands. In David Hopley (ed.), *Encyclopedia of Modern Coral Reefs: Structure, form and process*. 940–945. Dordrecht: Springer.

Karimata Shigehisa. 2003. Kiki Gengo to shite no Ryukyu Hogen no Kenkyu Jokyo: Nihon Fukki Go kara Kyo made no Katsudo ni tsuite no Oboegaki [The status of research on Ryukyuan dialects as endangered languages: A memorandum on

activities since the return to Japan until today]. *Ryukyuan Dialects as Endangered Languages in Japan* 39. 257–267.
Kiyama, Soichi. 2008. Okinawa ni yoru Amami sabetu: sono ryokai no kozo [Okinawa's discrimination against Amami: the structure of understanding], *Yoron Island Qualia*. http://manyu.cocolog-nifty.com/yunnu/2008/10/post-7ad3.html (26 August 2021).
Kondo, Ken'ichiro. 2014. Japanese language education in modern Okinawa until 1945. In Mark Anderson & Patrick Heinrich (eds.), trans. Yuka Ando, *Language Crisis in the Ryukyus*, 54–81. Newcastle upon Tyne: Cambridge Scholars Publishing.
Kroskrity, Paul V. 2009. Language renewal as sites of language ideological struggle: The need for "ideological clarification". In Jon Allan Reyhner & Louise Lockard (eds.), *Indigenous language revitalization: Encouragement, guidance & lessons learned*, 71–83. Flagstaff, Ariz: Northern Arizona University
Leonard, Wesley Y. 2018. Reflections on (de)colonialism in language documentation. (Eds.) Bradley McDonnell, Andrea L. Berez-Kroeker & Gary Holton. *Language Documentation & Conservation* SP15: Reflections on Language Documentation 20 Years after Himmelmann 1998. 55–65.
Li Wei. 2018. Translanguaging as a Practical Theory of Language. *Applied Linguistics* 39(1). 9–30.
Liacas, Tom. 2019. What's the difference between "distributed" and "decentralized" organizing? *NetChange Consulting Inc.* https://netchange.co/whats-the-difference-between-distributed-and-decentralized-organizing (25 July 2021).
Lim, John Chuan-Tiong. 2016. Lìnglèi tángrénjiē: "mǐn rén sānshíliù xìng" 'yǔ liúqiú jiǔmǐ cūn [The Alternative Chinatown: Lewchew (Ryukyu) Kume Village and "36 Min Families"]. *Translocal Chinese: East Asian Perspectives* 10(2). 316–339.
Lin, Angel M. Y. 2015. Researcher positionality. In Francis M. Hult & David Cassels Johnson (eds.), *Research methods in language policy and planning: A practical guide*. West Sussex, UK: John Wiley & Sons.
Macedo, Donaldo P, Bessie Dendrinos & Panayota Gounari. 2016. *The hegemony of English*. Routledge.
Madsen, Thomas. 2021. Between frustration and education: Transitioning students' stress and coping through the lens of semiotic cultural psychology. *Theory & Psychology* 31(1). 61–83.
Maeda, Tatsuro. 2014. Amamian language life experiences of migration and "dialect correction". In Mark Anderson & Patrick Heinrich (eds.), trans. Philippa Heming, *Language crisis in the Ryukyus*, 82–102. Newcastle upon Tyne: Cambridge Scholars Publishing.
Maher, John. 2019. Diglossia in multilingual communities. In Simona Montanari & Suzanne Quay (eds.), *Multidisciplinary perspectives on multilingualism: The fundamentals*, 103–122. De Gruyter Mouton.
Manohar, Narendar, Sameer Bhole & Amit Arora. 2017. Researcher positionality in cross-cultural and sensitive research. In Pranee Liamputtong (ed.), *Handbook of Research Methods in Health Social Sciences*, 1–15. Springer Nature Singapore.
Marcum, James A. 2013. From paradigm to disciplinary matrix and exemplar. In Vasso Kindi & Theodore Arabatzis (eds.), *Kuhn's The structure of scientific revolutions revisited*, 41–63. New York: Routledge.
Masiko, Hidenori. 2014. The politics of the movement to enforce standard Japanese under the U.S. occupation. In Mark Anderson & Patrick Heinrich (eds.), trans.

Philippa Heming, *Language crisis in the Ryukyus*, 82–102. Newcastle upon Tyne: Cambridge Scholars Publishing.
Matsuda, Misuzu & Gijs van der Lubbe. 2020. Japan as a multilingual society: Simakutwuba: The language in the Okinawa Main Island and its surrounding islets (21) language education that acknowledges Okinawan (dialect) variation. *Language and Society* (22). 202-216. https://ci.nii.ac.jp/naid/40022419940.
Merlan, Francesca. 2009. Indigeneity: Global and local. *Current Anthropology* 50(3). 303–333.
Meyer, Stanislaw. 2007. The rhetoric of the assimilation ideology in the remote islands of Okinawa: Becoming Japanese or Okinawan? *Eras* 9.
MOFA. 1960. Agreement regarding the status of United States Armed Forces in Japan. https://www.mofa.go.jp/region/n-america/us/q&a/ref/2.html (24 July 2021).
Murray, Andrea E. 2017. Okinawa's tourism imperative. In *Footprints in paradise: Ethnography of ecotourism, local knowledge and nature therapies in Okinawa*, 15–28. Berghahn Books.
Nakashima, Koji. 2010. Nature conservation and anti-base movement in Okinawa. *Geographical Sciences* 65(3). 231–241.
NHK. 2001. Renzoku Terebi Shosetsu: Churasan [Serial TV novel: Churasan]. https://www2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das_id=D0009010478_00000 (24 July 2021).
Ning, Wang. 1997. Orientalism versus Occidentalism? *New Literary History* 1, 57-67.
NINJAL. 2016. Documentation of endangered languages and dialects in Japan. *National Institute for Japanese Language and Linguistics*. https://www.ninjal.ac.jp/research/project-3/institute/endangered-languages/ (27 July 2021).
Okinawa Hands-On NPO. n.d. Okinawa Hands-On NPO. https://www.h-on.org/ (23 July 2021).
Okinawa Prefecture. 2020. "Shimakutuba Fukyu Suishin Keikaku" oyobi "Shimakutuba Fukyu Suishin Kodo Keikaku" ni tsuite [About "Ryukyuan languages spread promotion plan" and "Ryukyuan languages spread promotion action plan"]. https://www.pref.okinawa.jp/site/bunka-sports/bunka/shinko/simakutuba/keikaku.html (23 January 2021).
Ono, Mitsuaki. 2015. "Shin" Shokuminchi Shugi Ron toiu Hikari no motode "Okinawa Mondai" wo kangaeru [Considering "the Okinawa Problem" with the Theory of "Neo"-Colonialism]. *Ritsumeikan Studies in Language and Culture* 27(1). 193–207.
O'Rourke, Bernadette, Joan Pujolar & Fernando Ramallo. 2015. New speakers of minority languages: The challenging opportunity – foreword. *International Journal of the Sociology of Language* 2015(231). 1–20.
Pellard, Thomas. 2015. The linguistic archeology of the Ryukyu Islands. In Patrick Heinrich, Shinshō Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan languages: History, structure, and use*, 13–37. Berlin: De Gruyter Mouton.
Port Language Revitalization Project. 2020. Port Language Revitalization Project. https://plrminato.wixsite.com/webminato (24 August 2020).
Rangahau. n.d. Rangahau. http://www.rangahau.co.nz/rangahau/ (17 July 2021).
Roche, Gerald, Hiroshi Maruyama & Åsa Virdi Kroik (eds.). 2018. *Indigenous efflorescence: Beyond revitalisation in Sapmi and Ainu Mosir*. Australian National University Press.

Rots, Aike P. 2019. Strangers in the sacred grove: The changing meanings of Okinawan utaki. *Religions* 10(5). 298.

Rubin, Herbert J & Irene S Rubin. 2011. Research philosophy and qualitative interviews. In *Qualitative interviewing: The art of hearing data* 3rd edn, 13–24. Thousand Oaks, California: SAGE Publications.

Ryukyu Shimpo. 2003. Paarankuu. https://ryukyushimpo.jp/okinawa-dic/prentry-42566.html (16 April, 2021).

Sakihara, Masashi & Shinako Oyakawa. 2021. Shimakutuba fukyu suishin katudo dantai "Kutuba Surija: Ninufa bushi" katsudo hokoku matome [Report and summary of the activities of a group for promoting Ryukyuan languages "Language School: Ninufubushi"]. *Bulletin of National Institute of Technology, Okinawa College* 15. 9–12.

Samuels, Richard J. 1994. *"Rich nation, strong Army": National security and the technological transformation of Japan*. Cornell University Press.

Sanders, Robert. 2020. Is English the lingua franca of science? Not for everyone. *Berkeley News*. https://news.berkeley.edu/2020/10/14/is-english-the-lingua-franca-of-science-not-for-everyone/ (27 July 2021).

Sato Yu. 2020. Ryukyu Go on Seisho Ho: Chiteki Isan Fukko ni FUkaketsu [Orthography of the Ryukyuan language: Essential for the recovery of our intellectual heritage]. *Ryukyu Shimpo*. https://ryukyushimpo.jp/news/entry-1070998.html (25 April 2021).

Shimakutuba Fukyu Center. 2018. *#Tsukatte miyo Shimakutuba Navi [Let's use community languages Navi]*. https://shimakutuba.jp/ (28 August 2021).

Shimoji, Kayoko. 2020. Current status and issues of language succession activities in Ryukyuan languages: An analysis of preceding case studies. *Journal of Japanese Language and Japanese Literature* (24)2. 98-126.

Soja, Edward W. 1996. *Thirdspace: Journeys to Los Angeles and other real-and-imagined places*. Cambridge, Mass: Blackwell.

Sudo, Naoko. 2016. Okinawa he Iju suru Wakamono tachi: Kirino Natsuo "Metabora" ni miru Ijusha Zo [Young people migrants to Okinawa: The image of migrants in "Metabora" by Kirino Natsuo]. *Sociological papers: Official journal of the Waseda Graduate Students' Association for Sociological Research, Japan* 25. 17–34.

Sugimoto, Yoshio. 1999. Making Sense of Nihonjinron. *Thesis Eleven* 57(1). 81–96.

Swain, Merrill. 2013. The inseparability of cognition and emotion in second language learning. *Language Teaching* 46(2). 195–207.

Tada, Osamu. 2015. Constructing Okinawa as Japan's Hawai`i: From honeymoon boom to resort paradise. *Japanese Studies* 35(3). 287–302.

Tai Chugoku "Nansei shifuto" kasoku, "Taiwan yuji" ni Boeisho sonae [The Japan Ministry of Defence accelerate the "southwest shift" against China to prepare for 'Taiwan contingency']. 2021. *Jiji.com*. https://www.jiji.com/jc/article?k=2021043000851&g=pol (24 July 2021).

Ting-Toomey, Stella. 2015. Identity negotiation theory. In Charles R Berger, Michael E Roloff, Steve R Wilson, James Price Dillard, John Caughlin & Denise Solomon (eds.), *The International Encyclopedia of Interpersonal Communication*, 1–10. 1st edn. Wiley.

Tohyama, Nana. 2019. Ryukyu shogo ni okeru onsei kyozai ni tsuite: Gengo shutoku to hogen kyoiku to kikigengo no kiroku hozon no kaku kanten kara [Development of

audio materials for Ryukyuans as language documentation and dialectal education, language acquisition]. *Bulletin of the Humanities and Social Sciences University of the Ryukyus* 5. 161-183.
Tomigusuku City. 2020. Tomigusuku Shi: Keikan Keikaku [Tomigusuku City: Landscape Plan]. https://www.city.tomigusuku.lg.jp/municipal_government/1667/7881 (12 April 2021).
Tonaki jima nikki. 2013. Tonaki jima nikki [Diary of Tonaki Island]. Official blog of Tonaki Island, Okinawa. http://tonakijima.seesaa.net/archives/20130928.html (24 July 2021).
UN. 2017. United Nations Declaration on the Rights of Indigenous Peoples. https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html/ (8 April 2021).
UNDESA. 2008. *Resource kit on indigenous people's issues*. New York: United Nations Department of Economic and Social Affairs.
Viswanathan, Madhubalan, Sara Baskentli, Samanthika Gallage, Diane M. Martin, Maria Ramirez-Grigortsuk & Saroja Subrahmanyan. 2021. A demonstration of symbiotic academic-social enterprise in subsistence marketplaces: Researching and designing customized sustainability literacy education in Tanzania. *Journal of Public Policy & Marketing* 40(2). 245–261.
Widianingsih, Ida & Donna M. Mertens. 2019. Transformative research and the sustainable development goals: Challenges and a vision from Bandung, West Java. *International Journal for Transformative Research* 6(1). 27–35.
Wilson, Shawn. 2008. *Research is ceremony: Indigenous research methods*. Black Point, N.S.: Fernwood Pub.
Yamada, Masahiro, Yukinori Takubo, Shoichi Iwasaki, Celik Kenan, Soichiro Harada, Nobuko Kibe, Tyler Lau, Natsuko Nakagawa, Yuto Niinaga, Tomoyo Otsuki, Manami Sato, Rihito Shirata, Gijs van der Lubbe & Akiko Yokoyama. 2018. Experimental Study of Inter-Language and Inter-Generational Intelligibility: Methodology and Case Studies of Ryukyuan Languages. *Japanese/Korean Linguistics* 26.
Yokoyama, Akiko & Takayuki Kagomiya. 2019. Gengo Jikken ni motozuku Gengo Suitai no Jittai no Kaimei: Ryukyu Okinoerabu Jima wo Jirei ni [Elucidation of the actual situation of language decline based on language experiments: A case study in Okinoerabu Island in the Ryukyus]. *Studies in Dialects* 5. 353–375.
Yokoyama, Akiko. 2021. *Shimamuni Takarabako [A treasure box of community language]*. https://www.erabumuni.com/ (19 August 2021).
Yomitan Village History Editing Room. 2021. Shashin de miru Yomitan son no imin & dekasegi: sekai no Yuntanzanchu [Emigrant workers seen in the photos: Yomitan people in the world]. http://imin-visual.yomitan-sengoshi.jp/.
Zlazli, Miho. 2021. Atarashii Washa no Shiza kara mita Ryukyu Shogo no Kaika no Torikumi [Efflorescence of Ryukyuan Languages from Perspectives of New Speakers]. *Journal of Regional Science for Small Islands* 2. 163-181.
Zlazli, Miho. forthcoming. *Constant fear of ostracism*. https://www.researchgate.net/publication/354142196_Constant_fear_of_ostracism (26 August 2021).

SOAS Working Papers in Linguistics. Volume 20 (2021): 165–180

Interdisciplinary research in language documentation: The benefits and present limits of a more sustainable documentation methodology

Sarah Round

686422@soas.ac.uk

Abstract
Interdisciplinary Research (IDR) in language documentation furnishes linguists with the knowledge they require to access and explain linguistic practices inextricable from their extra-disciplinary context. By heightening the validity of documentation in this way, as well as generating linguistic data of novel importance to other disciplines, the use of interdisciplinary methods moreover creates a more sustainable documentary model. In spite of this, IDR does not secure any sizeable disciplinary and financial interest relative to mono-disciplinary models. The following paper examines the extent to which this lesser uptake of IDR within documentation projects may be caused by the particular disadvantages of its related methods, such as investigating language-specific Traditional Ecological Knowledge (TEK) or cross-disciplinary domain-based concerns, and to what extent this may be caused by its lower levels of exposure and funding in the academic community. Finding the answers to resolve IDR's lesser prominence within Linguistic academia should, in turn, facilitate its increased teaching and more proficient practice within language documentation. This paper does not present primary research, but instead aims to summarise relevant arguments and studies in a manner accessible to the general reader.

Keywords
Interdisciplinary Research (IDR), Traditional Ecological Knowledge (TEK), interdisciplinary domains, indigenous taxonomy, extra-disciplinary data, ethnobotany, ethnography, language documentation

1. Introduction

Interdisciplinary Research (IDR) involves linguists and specialists from other external disciplines mutually contributing to the investigative processes involved in the documentation of a language. This includes conceptualization of research concerns, methodology and the final analysis (Penfield 2020: 2). Conventional multi-disciplinary approaches often isolate non-linguistic data as evidence to complement linguistic documentary findings. In contrast, IDR integrates relevant "information… techniques… [and] perspectives" (Amith 2020: 72) from disciplines such as anthropology or ecology. Furthermore, the research concerns addressed by IDR are purposely designed to be of relevance and interest to all parties involved.

As this paper will demonstrate, the increased practice of Interdisciplinary Research within language documentation could offer considerable advantages. As a result of its integrated approach, IDR enables precise or nuanced "explanations of [certain] complex phenomena" which are best explored from a combined linguistic and extra-linguistic perspective. Its cross-disciplinary research concerns, moreover, generate revolutionary analyses which "no single discipline [could] create on its own" (Derrick et al. 2011: 3).

Holton draws attention to this principle in terms of linguists requiring collaborators to extricate particular language practices from their biological context. Without the collaboration of the right specialists, linguists could not make complex botanical distinctions such as identifying subtly different plant families; thus, their corresponding separate lexical terms might never be searched for (Holton 2018: 7).
In order to enumerate the further benefits of this approach, it is necessary to examine the methods that interdisciplinary language documentation can entail. Methods to be discussed include employing cross-disciplinary domain-based research concerns and recording extra-disciplinary but language-specific Traditional Ecological Knowledge (TEK). The ways in which shared domain-based concerns can be used to expedite and expand linguistic data collection, and collaborative TEK documentation can elucidate languages' cultural histories will be illustrated. There will, additionally, be an exploration of how collaboration with ethnographers, social anthropologists or biologists can give linguists the knowledge they require to document certain linguistic practices that are unexplainable without their extra-disciplinary context. It is important to note that this analysis looks at the benefits of IDR principally from the perspective of the linguist.

Apparent drawbacks of this approach will then be outlined. The practical implications of non-linguists joining fieldwork teams can both positively and negatively affect language consultant contributions and chances for funding. Domain-based language research can also distort linguistic documentation by overestimating the equivalency of linguistic representational strategies across languages. The author will argue, however, that IDR's lesser academic exposure, alongside linguists' inattention to the principle that IDR-based projects should provide equal benefits for collaborators, represent the commensurately large obstacles preventing the wider use of the methodology in linguistics. No primary research is presented; rather the paper aims to summarise relevant arguments and studies in a manner accessible to the general reader.

The publications of several linguists, in particular, are greatly drawn upon to prove these points. Niclas Burenhult (2020) is cited in order to assess the proficiency of using domain-based research concerns when documenting language. The aforementioned Gary Holton's (2018) work is used extensively; first, to indicate the multi-faceted nature of researcher-consultant rapport in IDR. He is also referenced to detail how biologist collaborators ensure documentation on language specific TEK or species classifications contains unambiguous lexical terms for phenomena. The first of these publications to be discussed is by Jeff Good who, in conjunction with Bulmer (1967, cited in Si 2011), introduces the complex ways in which different disciplines can intersect and elucidate linguistic phenomena.

2. Language practices inextricable from cross-disciplinary contexts

Interdisciplinary consultation is essential when the true meanings or motives underlying linguistic phenomena prove inextricable from their cultural or ecological environments. In such instances, linguists must consult extra-linguistic resources, and specialists who can provide the relevant ethnographical or ethnoecological expertise.

It is ethnographic knowledge that reveals how a language interacts with its community's "subjective cultural attitudes and belief systems" to encode meaning in "objective" semantic categorizations or lexical forms (Si 2011: 170-1). As a result, through the study

of the mythology of the Kalam speech community for instance, Bulmer was able to arrive at an accurate explanation of their categorisation of the cassowary bird. Distinguished from their broader category of *yakt* 'flying bird or bat' it was not, Bulmer found, due to its relatively large size or inability to fly, but rather because of the "special (kin) relations [that it shared] with humans" in their folklore (1967, cited in Si 2011: 171). Similarly, collaboration with ethnobotanists is critical to linguistic documentation of plant classifications. This is due to their ability to reliably identify organisms using standardised biological taxonomies, and to then delineate which extra-linguistic factors from botany or social anthropology (flora morphological aspects, traditions, agriculture) may have coincided to create their language-specific nomenclature (Holton 2018). Employing purely linguistic processes to analyse phenomena such as these could lead to the loss of fine distinctions of lexical meaning. Using methods uninformed by extra-disciplinary insight can, moreover, create unrepresentative research parameters.

One problematic parameter of linguistic research can be found in the documenting of multi-lingual populations. Traditional approaches commonly prioritise recording a single ancestral code. This derives from the assumption that the multi-lingual environment is forced on the indigenous language community by a "dominant colonial" linguistic presence. It also derives from the assumption that this second language serves only to push the ancestral language towards shift or endangerment (Good 2020: 67). Documentary linguists additionally presume speech communities view any consequent lesser use of their first language as a lessening of their identity. This is because, according to "Herderian equation" language ideology theory, they interpret language as "a marker of immutable ethnicity" (Good 2020: 68). This interpretation comes at the expense of multi-lingual practices not being sufficiently documented. It also heightens the risk that linguists will not fully consider the social significance of multi-lingual practices to the speech community in question.

By contrast, ethnographers tend to place emphasis on discovering what the languages used in a community culturally represent to their speakers. Collaboration with them can thus reveal to linguists how ancestral and contact language varieties can positively and meaningfully coexist. An example of this is in Good's (2020: 68) documentation of the highly multi-lingual West Cameroon Lower Fungom community, where his investigations into the social contexts of their varying language choices were directly instigated by ethnographic interdisciplinary research.

These investigations found that the individual linguistic varieties used by the Lower Fungom community represented different "socio-political identities". Said identities could be taken on at will by any variety user to gain powerful affiliation with the relevant linguistic group. Language use, therefore, did not indicate permanent but rather *ephemeral* linguistic identities that users moved between according to circumstantial needs or interests, such as ensuring social "co-operation" or to "access…the resources of different groups" (Good 2020: 68). Ethnographic collaboration consequently can be seen to encourage documentary methodologies inclusive of multi-lingual or other sociolinguistic contexts and augments documentary explanations surrounding language perception and use.

Furthermore, historical, and geographical documents can vitally supplement linguistic theory with knowledge of the social causes behind atypical patterns of linguistic divergence. If this knowledge is not harnessed in the documentation of a language, unusual patterns may be written off as insignificant, or the explanations behind them could remain incomplete.

In his same study of Lower Fungom, Good (2020: 48) identified a Mungbam language variety from the village of Missong that showed significant lexical and grammatical divergence from other varieties in neighbouring villages. Linguistics might typically attribute such linguistic divergence to language drift. This concept identifies language change as dependent on how much time has elapsed between the migration of speakers from a formerly united language community to new dispersed settlements. As time elapses, linguistic divergence in the new settlement areas increases. In this process, individual alterations emerge, and over time distinguish the new variety from its linguistic starting point (Sapir 1921: 74). Generally, such alterations are seen to occur unconsciously and cumulatively. However, in consulting historical documents from the Buea National Archives, Good (2020: 55) evidenced that the Missong village community had only "recently" re-settled. By citing colonial documents from the early twentieth century that described the community as a "historically shallow…break-off [settlement] from the village of Munken" (Good 2020: 64), Good uncovered that language change following resettlement must have been relatively rapid.

Good (2020: 62) then created his own geo-linguistic material which corroborated these historical findings. By mapping the locations in Lower Fungom that Mungbam speakers identified as "sites of memory" (places referenced in their ethnohistorical lore), he unearthed that in 1860, Missong had far fewer memory sites than other nearby villages with similar population sizes at that time. Having established the divergence of the Missong variety through linguistic analysis, and the recentness of the village settlement through geographical and historical documents, Good (2020: 55–56) was then able to explain the cause of rapid language change by documenting material of an ethnographic nature: Missong origin stories. The speakers he recorded suggested that the village had been created "by immigrant groups" from outside the local area. These groups, they claimed, had actively adopted, or "stolen" Mungbam from other villages.

From all this multi-disciplinary material, Good (2020: 65) thus concluded that the reason behind the language change was not due solely to one community of Mungbam speakers having resettled and quickly innovated. Instead, he clarified the highly divergent Missong variety was the result of considerable grammatical and lexical borrowing from the multiple and diverse language groups who settled in the village. The rapidness with which this new variety emerged, Good proposed, could be attributed to the need to quickly create a lingua franca. What's more, drawing upon the ethnographic observation that Lower Fungom speech varieties were commonly used to represent different identities, (Good 2020: 68) he posited that some aspects of the variety may have been purposefully designed to distinguish its speakers as a unique "kin group federation" (Good 2020: 57). This demonstrates how extra-linguistic information can address documentary linguistic queries in situations where linguistics alone might not suffice.

Linguists could, as Good (2020) has done, compile cross-disciplinary detail for documentation either on their own or with strategic extra-disciplinary consultation. However, switching to IDR, which is here defined as collaboration with other specialists throughout the research process, might more efficiently signpost them to linguistic phenomena that intersect closely with other disciplines. While it is not always easy to predict which disciplines these may be, and so still hard for linguists to know what their interdisciplinary team should look like, examining previous documentation can most certainly indicate common connections. Extrapolating from the studies shown in this section, for example, diachronic linguistic change could be elucidated through directly collaborating with ethno-historians instead of using historical material; while accurate representation of any ecologically influenced taxonomies or biological nomenclature could be assured by working alongside social anthropologists and ethnobotanists.

3. IDR: a more sustainable model of language documentation

IDR also promotes a more sustainable model of language documentation in several ways. Firstly, by drawing attention to previously unexplored cross-disciplinary research concerns, it generates a larger scope for future documentary inquiry. Secondly, IDR enables extra-linguistic information to be incorporated accurately, and in a way that makes documentary findings directly applicable to the outside discipline(s) involved in the collaboration.

By having biologists or botanists onsite to assist with linguistic fieldwork, for example, the linguist avoids being over-reliant on field guides that might lead to erroneous classifications of organisms. Botanical field guides, Holton (2018: 5-6) argues, often do not represent all relevant plant varieties, "emphasize [certain] morphological aspects of the species [that are not the] most salient to the linguist," or contain pictures that misconstrue "relative and absolute sizes". Linguists may also lift from field guides the generalised glossed terms used for several subtypes of phenomena and so misidentify them, naming something *ginger*, for instance, when it is another type of *rhizome* (Holton 2018: 5). Others might neglect technical classification entirely and only record language-specific terms for organisms out of focus on linguistic interpretation over content (Holton 2018: 4). Extra-disciplinary collaborators on the other hand can employ their own knowledge of specialised classification systems: such as using Linnaean binomial equivalents for plants and animals (Holton 2018: 5). As a result, they guarantee that all phenomena are identified, and that said identifications are unambiguous and valid. Gaining more reliable documentation in this way benefits the linguist, as well as their collaborators. In particular, language documentation that appropriately records the biodiversity of lesser documented rural environments can be invaluable to taxonomists and conservation biologists (Ramstad et al. 2007; Si 2011; Klubnikin et al. 2000).

By creating such documentation that is varied, accurate and widely re-applicable, yet which most importantly can only be developed through continued collaboration, IDR secures its own sustainability. One instance of this can be observed in the documentation of the Eastern Indonesian Tobelo language. Cooperation of biologists and linguists in this study ensured a reliable examination of previously undocumented covert taxonomic categories of certain sexual biotic forms. In contrast to other asexual corals and algae, it emerged that black coral was classified linguistically as having separate male and female forms. This finding has enabled and encourages further investigation by future

interdisciplinary groups into "what distinguishing biological feature of black coral might underlie [this] Tobelo folk classification." As a new research concern, this could help both linguists, by informing understanding of Tobelo noun class systems, and biologists, by improving the accuracy of Western binomial categories (Holton 2018: 7).

4. Revealing language contact and honouring ethics through TEK

Traditional Ecological Knowledge (TEK) is the knowledge a community has of its environmental resources, such as crops, plants, and other living organisms, as well as a set of beliefs and practices that illustrate how they as humans relate to these resources. Recording TEK, especially that of non-urban communities, is another research concern that requires collaboration, this time with ecologists, biologists, or ethnobiologists. According to Si, this is because rural language speakers have been found to provide knowledge that is significantly specialised, "rival[ling] that of…professional naturalist[s]" (Si 2011: 175). Nevertheless, linguists will document a community's language specific TEK as they find it beneficial to tracing the groups' cultural history.

Eliciting this TEK affords practical advantages in terms of linguistic research, along with access to a domain that is prone to reflect language contact. As Si further states, TEK is a substantial source of easily elicitable data in rural speech communities because speakers from these areas must "routinely engage with their natural environment…to be a fully-functioning member of a community." They might need to identify edible foods, avoid "dangerous organisms", or know the "seasonal cycles for agricultural purposes" (Si 2011: 172). Moreover, a specific part of this ecological knowledge, the semantic domain of taxonomy, is highly perceptible to change where languages interact (Amith 2020: 101). Indigenous taxonomies based on TEK thus show layers of lexicosemantic change, which can manifest as the borrowing of loan words, or loan translations from various languages. Linguists can use these layers to diachronically track when one speech community interacted with another, or where its speakers might have migrated to new settlements (Gadgil, Berkes & Folke 1993: 155).

In the Sierra Nororiental de Puebla region of Mexico for example, Amith recorded that many basic taxonomic loan words from the Totonac language were being used within Nahuatl speech. From this documentation, he was able to postulate that there had been historical migration of Nahuatl speakers into these previously Totonac community areas (Amith 2020: 102).

However, to establish whether an organism's taxonomical term or a loan translation of the term has been shared across speech communities, linguists must be able to reliably signpost specimens of that organism to their language consultants. Linguists attempting this on their own would again encounter the issues raised in Section 3. Lack of biological knowledge, for instance, might cause the linguist to misidentify one species or even subspecies of plant as being the same as another. This might lead them to request the taxonomy of one plant from members of the first speech community but accidentally ask for the taxonomy of a completely different plant from consultants in the second. Alternatively, an unaccompanied linguist may correctly elicit an organism's taxonomy at species level but be unaware of the subspecies of that organism which could also be identified through a loan word or translation, or vice versa. Therefore, as aforementioned

by Si, comprehending and fully exploiting the specialised taxonomic data within TEK calls for biological or ethnobiological collaboration.

Past documentations on the borrowing of taxonomic lexical terms across languages have been instrumental in corroborating the delineation of certain cultural and historical regions, such as Mesoamerica. Smith-Stark recorded fifty-two cases of compound words that were shared across Mesoamerican languages in the form of loan translations (Campbell, Kaufman & Smith-Stark 1986). Loan translation here signifies the literal, referent-for-referent translation of a term from one language to another. Smith-Stark noted that the use of these loan translations was widely distributed within Mesoamerica, and tightly restricted outside its perceived borders (Campbell, Kaufman & Smith-Stark 1986: 554). Of these terms, eighteen percent referred to flora or fauna (Amith 2020: 102). In other words, they related to and required knowledge of an extra-linguistic, biological domain.

Documenting TEK could also be considered part of the linguist's greater moral imperative to combat areas of language shift or loss caused by "problematic social inequalities" (Dobrin & Schwartz 2016: 259). According to Dobrin and Schwartz (2016: 259), one of the ethical principles embedded within language documentation is that the linguist should actively address any "problematic" loss they observe. They put forward the idea that this tendency arose shortly after 1990, when the disciplinary movement towards documenting endangered languages revealed that language shift in such speech communities was more commonly driven by the far greater, and "fundamentally political, economic, and cultural" pressures they faced, rather than any one linguistic force. Documentary linguists thus believed that to prevent further language loss within a community, their work must also address the social and environmental roots of the problem. In the same vein, the part of a speech community's lexicon that contains local ecological knowledge has been observed to be more inclined to undergo loss. This is, furthermore, often as a result of similarly extra-linguistic, environmental factors including urbanization, speaker migration, or even environmental degradation (Si 2011; Burenhult 2020). Additionally, speech communities typically define this loss of their TEK as unwelcome, as they view it to contain vital information on agricultural or non-urban ways of life (Si 2011: 174–175). Si, for example, cites that Solega speakers in southern India expressed strong remorse at their community's decreased retention of their former local ecological knowledge (2011: 176).

Whilst biologists have been known to independently record and thus preserve the botanical and taxonomic knowledge held by rural speech communities, the resulting resources are usually documented and published in a dominant language, for the easier access of Western academic audiences (Si 2011: 18). This does not address the speech community's own need to access records of such knowledge in their native language, and so conserve its use in the local environment. In contrast, linguistic documentation of TEK can ensure that said knowledge is also preserved in its original lexical form. Providing practical ecological resources to a community desirous to conserve its endangered TEK, especially if that TEK is at risk due to social factors, also upholds documentary ethical standards. The UNESCO People and Plants project demonstrates how the return of ecological information to speech communities in the appropriate language can greatly assist its conservation (Martin 2004).

In summary, documentary linguists represent a group who have the disciplinary capacity to record a community's TEK in the native language, as well as having their own academic and ethical motives to do so. These motives derive from the potential TEK documentation has in terms of tracking languages' cultural histories, and addressing problematic language loss or speech community concerns, respectively. Having understood that this form of interdisciplinary documentation is both feasible within, and beneficial to the discipline of Linguistics, one must also consider that its technical accuracy can be efficiently ensured when linguists collaborate with other specialists who have the right ecological expertise. Language-specific TEK records, therefore, can be collected, they should be collected, and above all are best collected as part of an IDR-based team.

5. Documenting interdisciplinary domains of shared human experience

Contrary to common belief, the research concerns of a documentation do not need to be based on linguistic queries to draw targeted linguistic conclusions. Burenhult (2020) argues that basing investigations instead on more accessible, thematic *domains* of human experience can generate far larger corpora of data that are equally applicable to linguistic theory. These domains also tend to have broader parameters that commonly stipulate crossing disciplinary boundaries.

Burenhult claims that basing research concerns on recording the linguistic representational strategies used by a speech community to express a certain set of fundamental, "universally relevant" human experiences (Burenhult 2020: 9) is highly conducive to the in-depth exploration of that language (Burenhult 2020: 10–11). This is because due to the essentialness and relatability of these experiences, speakers from all environments are capable of representing them "in language and thought," and, above all, at complex levels. In contrast, if linguists choose to base documentation on more esoteric research concerns, such as "serial verb constructions or phrase-level prosody" (Burenhult 2020: 9), any language consultants who are non-linguists might not be able to comprehend, nor offer in any great measure the information they are looking for.

Take the domain of landscape: it is structurally fundamental to language because it is needed to reference locations where activities take place, such as "named places, [and] landforms" (Burenhult 2020: 15). It also has great "spatial and temporal constancy," making it a deeply relatable domain for all speech communities (Burenhult 2020: 13). Burenhult witnessed that the Austroasiatic Jahai speech community consultants responding to domain-based landscape prompts showed rapid comprehension of concepts and were subsequently more attentive. Large amounts of linguistic data were thus collected within relatively short timeframes. Moreover, he remarked that speakers took more creative control with their responses by independently volunteering specialist words and phrases (Burenhult 2020: 15).

As aforementioned to be common amongst domain-based documentation, this study also prompted the sharing of diverse linguistic knowledge that intersected with other disciplines. There were "hunting stories" of relevance to ecological natural resource management, place names that were useful to anthropological study since each was attached to its own creation myth, and narratives conducive to learning local geography such as "travel stories" and "life histories [which were] always tied to different locations

and movement between them" (Burenhult 2020: 15). Typically, domain-based linguistic projects involve collaboration with many such "scientific branches and fields" (Burenhult 2020: 12). Participating linguists tend to draw upon the extra-disciplinary knowledge found in initial documentation in order to further cater to interdisciplinary research concerns in their subsequent work (Burenhult 2020: 16).

Such was the case for the LACOLA project, which carried out domain-based documentation on "the relationship between language, thought and landscape in several diverse and endangered language settings" (Burenhult 2020: 17). Motivated by its findings, Hägerhäll and Sang (in progress, as cited in Burenhult 2020) then spearheaded a subproject which combined the disciplinary resources of landscape architecture and environmental psychology and held the further specified aim of investigating "human landscape preference from a cross-cultural [and cross-linguistic] perspective" (Burenhult 2020: 17).

Domains are therefore an interdisciplinary medium of expediting data collection, as well as extending data ranges and sizes to enable efficient analysis of linguistic and cross-disciplinary community practices. Nonetheless, the extent to which documenting these spheres of shared human experience will reveal comparable representational strategies across languages can be overestimated. Certain categories within the domain of landscape, like place names, may indeed be stable enough to provide the means for cross-linguistic analysis on spatial representations (Burenhult 2020: 10–11). However, there is still significant "linguistic variation in geographical ontology and conceptualisation." What might be perceived as fundamental geographical classification methods by documentary researchers, e.g., Geographic Information Systems (GIS), still operate through Western specific parameters. "Linguistically and culturally attuned" GIS must be developed before accurate landscape-based data collection in other speech communities can occur. Without such methods, there may be "inter-cultural misunderstandings" that distort the documentation of indigenous landform categories (Burenhult 2020: 16–17).
The success or failure of a domain-based documentation can thus potentially depend on whether the linguist is aware of any relevant non-corresponding linguistic representational strategies used by the respective speech community. For some linguists following this interdisciplinary method might seem an all too precarious path to tread, as could potentially come at the expense of their linguistic data's validity.

6. Non-linguist fieldworkers and their effect on consultant contributions

It is now important to turn our attention to the complex effects of being an interdisciplinary team in the field. The physical presence of extra-disciplinary specialists has the potential to either improve or impede the building of researcher-consultant rapport. Building such rapport is crucial since without it, language consultants may be less inclined to share the relevant linguistic information or may offer the researcher fewer opportunities to record it.

Whether the external researcher will improve or impede rapport building with language consultants depends on how informed they are about the language of that speech community, their cultural conventions, or even the area of extra-disciplinary knowledge under investigation.

Researchers who do not have the appropriate language knowledge, for example, may limit their team's opportunities to uncover linguistic practices unfamiliar to them. Extra-disciplinary collaborators are less likely to be familiar with a speech community's indigenous language, especially if it is endangered or minoritised, which could complicate their communication with mono-lingual community members. This difference in turn could act as a social barrier and restrict their interactions with the community at large. Linguists on the same team might then also, by association, become more "socially removed" from the speech community (Holton 2018: 10). As a consequence, chances to observe speaker behaviours at "off-duty times" could be missed. Holton identifies such experiences as being extremely valuable because it is within this setting that linguists often come across subsidiary speech community practices previously unknown to them (Holton 2018: 10).

The same extra-disciplinary experts, however, may be capable of building strong rapport with language consultants through drawing upon the type of knowledge they share with these speakers. Botanists or ecologists, for instance, will display a greater understanding of Traditional Ecological Knowledge than the linguist. Subsequently, they may be better at encouraging speakers to provide more detailed accounts of this information, which may also be linguistically relevant. Holton similarly remarked that Indonesian Abui language consultants were eager to discuss plant knowledge and its lexical terms in detail with the team botanist because they visibly comprehended the subject matter. Whereas uninformed enquiries made by the linguist, Holton noted, were more likely to receive "a one-word answer" (Holton 2018: 11).

As the above instances show, having good social relations with language speakers offers multiple documentary benefits, and these relations frequently come from greater familiarity with language community conventions. On the other hand, over-familiarity with a language community and their concerns can lead the linguist to be over-cautious, in that they may avoid broaching uncomfortable topics of conversation or questioning local knowledge that is seemingly "common sense." In contrast, extra-disciplinary team members, who are entering the local environment for the first time, will be less aware of what questions might be interpreted as controversial or mundane. This could work to the team's advantage.

In Holton's case, a botanist colleague posed questions to consultants regarding women's "eating habits and use of plants during pregnancy" (Holton 2018: 11). This was community-based knowledge that Holton had not attempted to document for several reasons. Firstly, these were habits that he, as a long-term researcher, had become accustomed to and thus not queried. Secondly, based on his earlier social interactions with the community, he feared this topic might be labelled as taboo. It instead transpired that the consultants were eager to share this traditional knowledge and so the team recorded a substantial source of formerly undocumented, culturally-specific and language-specific information (Holton 2018: 11).

From assessing Holton's experiences, one might postulate that the social dynamics of interdisciplinary documentation have greater potential to increase the amount of linguistic data collected than to decrease it. Furthermore, the supplementary linguistic knowledge obtained will be highly specialised and directly pertinent to the interdisciplinary research

concern in question. Conversely, there can be no guarantee that extended informal interaction with speakers in mono-disciplinary documentation would provide the linguist with knowledge that was suitably relevant or detailed enough to be useful to their research aims.

Apprising extra-disciplinary colleagues of language community conventions ahead of fieldwork would likely dissuade posing indelicate questions unnecessarily to speakers. Otherwise, a team could altogether deliberate on whether a certain query might yield results promising enough to risk weakening researcher-consultant rapport.

One way that IDR can adversely affect rapport between researcher and consultant, however, proves harder to resolve. Language consultants may become disillusioned with the documentation process if there are discrepancies between the linguist's and other collaborator's remuneration procedures. Whilst linguists can pay consultants for concrete instances of linguistic production like transcriptions or recordings, collaborating anthropologists, for example, will require more extended contact with consultants to draw meaningful observations, and so are likely to deem non-monetary payment more appropriate (Widlok 2005: 16). Employing these disparate procedures in the same interdisciplinary team could cause language consultants to not have a fixed notion of what their remuneration will be, or to feel that the form of compensation they receive from one researcher is less valuable than that given by the other (Akumbu 2020; Widlok 2005). A speech community could therefore become less willing to participate in future documentary sessions.

Interdisciplinary teams should, in this instance, avoid jeopardising rapport with speakers through lack of clarity. They may do so by offering more comprehensive explanations of the different remuneration methods and their motivations. Nevertheless, it is hard to see what interdisciplinary research teams could do to change the necessary varied ways in which these different disciplinary approaches must collect their data; and, when their remuneration methods must be implemented side by side, what researchers can do to discourage speech communities from comparing them.

7. The trials of interdisciplinary documentation in a mono-disciplinary world

Despite increasing interest in IDR within language documentation (Widlok 2005), this approach still does not secure any sizeable disciplinary and financial interest relative to mono-disciplinary models (Penfield 2020). This may stem from the fact interdisciplinary documentation projects have been habitually afforded lesser exposure within the academic linguistic community.

Many research journals, from both linguistic and other academic institutional bodies, only feature papers based within a single discipline. This greatly discourages linguists from engaging in IDR since academic publication is crucial to improving career prospects (Penfield 2018: 12). Linguistics journals especially are also prone to raise concerns with papers co-authored by non-linguists, as they traditionally prefer longer, contextually informed prose and extra-disciplinary writing styles might be more "concise [or]…formulaic" (Holton 2018: 17). On the other hand, interdisciplinary linguistic documentation that is published in extra-linguistic journals can equally be subject to

scrutiny, as linguists may find themselves needing to justify the work's linguistic value to internal "graduate…or tenure review committee[s]" (Holton 2018: 19).

Most academic funding bodies, moreover, are either apprehensive, or ill-adapted to support IDR-based documentations. This may firstly be due to their greater monetary demands. Interdisciplinary teams who are investigating a speech community's taxonomic lexicon, for example, will entail larger equipment expenses such as "specimen collecting vessels, plant dryers [and] photo stands" (Holton 2018: 18). They will also need to finance larger research assistant groups to help collect and process biological samples. Such botanical collaboration can also raise logistical issues or involve acquiring time-consuming field permits. The increased amount of equipment IDR requires, on top of standard documentation digital recording equipment, is likely to cause transport problems, particularly if equipment needs to be transferred to a remote fieldwork setting (Holton 2018: 18). Furthermore, before collecting any specimens, botanists must request and await approval from both local herbaria, and the CITES international ethical trade body to establish their collection process will not threaten species survival (Holton 2018: 19). For such reasons many funding review panels, including the US National Science Foundation, have been seen to view interdisciplinary proposals as "high-risk ventures" which thus disadvantages them in funding application processes (Amith 2020: 73–74).

As a language documentation methodology, IDR is also "rarely taught" by universities or other professional organizations (Penfield 2018; Holton 2018). Penfield notes, additionally, that present limited teaching of its methods fail in particular to reinforce one key principle of IDR: that the needs of all involved disciplines be integrated from the very "conception" of linguistic research (Penfield 2020: 3). This means the research concern of an interdisciplinary documentation should be designed so that its answers equally benefit all collaborating disciplines. Nonetheless, owing to lesser disciplinary mindfulness of this principle, linguists typically require their extra-disciplinary collaborators to provide a level of "technical service" within documentation that is greater than the extent to which its information benefits "their own scientific research agendas" (Amith 2020: 74).

Take Amith's example of taxonomists who are commonly recruited, as part of ethnobiological language projects, to classify specimens of a species already known to them. Even if a new species is discovered through the documentation, the classification of one species alone does not constitute sufficient data for a the taxonomist to write their own publication. Taxonomic disciplinary traditions dictate that the description of a certain species can only be published if accompanied by more extensive detail on the genus to which they belong (Amith 2020: 74). The lesser relative advantages such external collaborators receive when working with linguists, in turn, make them less likely to engage in extended or future interdisciplinary documentation (Amith 2020: 76). Collaborative documentation that shows no proof of "integration" through useful extra-disciplinary materials may also ultimately fail to qualify for interdisciplinary funding (Amith 2020: 74).

All these factors could, in fact, be seen as causally interlinked. IDR's difficulty in garnering academic recognition has led to a lack of opportunity for the method's funding,

correct teaching, and subsequent practice. Poor adherence to undertaught collaborative principles in the field then restarts the same cycle, wherein interdisciplinary documentation is viewed as "a random, unsystematic occurrence" (Penfield 2020: 3) that remains largely unable to prove its full worth.

There is, nevertheless, evidence that attitudes towards funding IDR within linguistics, as well as in wider academia, have noticeably changed. Language documentation initiatives including the Volkswagen Foundation DoBeS programme (est. 2000), and the Hans Rausing Endangered Languages Project (est. 2003) offer grants whose guidelines stipulate interdisciplinary methods as requisite to project approval (Widlok 2005). Whilst previously noted for their disinclination to invest in IDR-based projects, large funding bodies like the US National Science Foundation (NSF) have set up initiatives to finance multi-disciplinary projects, such as the Creative Research Awards for Transformative Interdisciplinary Ventures (CREATIV) and the Integrated NSF Support Promoting Interdisciplinary Research and Education (INSPIRE) (Amith 2020). This NSF INSPIRE award has been won, for example, by interdisciplinary teams with research concerns based in Computational Linguistics (University of Michigan 2013).

Penfield also argues that for certain funders interdisciplinary documentations represent a more affordable investment opportunity. This is because projects with cross-disciplinary research concerns can be subsidised by both linguistic and non-linguistic funding bodies. Internal funders may now share the financial responsibility for collaborative projects with sponsors from external disciplines. These external sponsors similarly benefit as they can invest more economically in extra-linguistic research concerns; they need only provide support equivalent to the documentary aspects relevant to them (Holton 2018: 18). Other interdisciplinary documentations may gain financing as a direct result of their broader research agendas because funding agencies perceive them to yield "more research outcomes for their money" (Penfield 2020, 2018).

Additionally, of course, interdisciplinary language documentation may gain greater external funding than other linguistic work due to the specific cross-disciplinary research concerns it covers. Recording TEK in Native American languages, for example, is a particularly high interest area for funding bodies (Penfield 2018).

Owing to its wider scope, its research parameters that improve cost-benefit ratio for investors, and most importantly the increased number of initiatives endorsing this model, IDR is now far less subject to "haphazard…funding" than it was in the past (Penfield 2018: 12). Instead, it is comparatively better promoted as a 'profitable' documentation methodology for career linguists and investors, a factor which might ultimately encourage its needed wider tutelage as a linguistic research method (Penfield 2018; 2020).

That said, in order for IDR to be able to equitably compete with mono-disciplinary models for academic approbation and financing, opportunities need to be accorded to this linguistic research method on a much larger scale. Whilst collaborative linguists could seek publication in long-standing linguistics journals specifically designed to cover IDR (e.g. Text & Talk, est. 1981), this does nothing to amend the method's comparatively restricted market. Traditionally mono-disciplinary linguistics journals must make a conscious effort, en masse, to address any biased publication regulations they might still

enforce. Linguistics funding bodies should renegotiate, if feasible, any budget ceilings typically set for mono-disciplinary documentations so as to similarly better accommodate IDR. Alternatively they could attempt to establish long-standing, shared financing links with other sponsors for IDR-based documentation, to take the unnecessary burden of "selling" their research away from the interdisciplinary linguist.

8. Conclusion

In conclusion, IDR provides linguists with the tools to build more precise and reliable documentation. Unambiguous lexicon for biological phenomena can be ensured by collaborating with botanists or biologists, linguistic practices inextricable from their cultural context can be elucidated by collaborating with ethnographers or social anthropologists, and language cultural histories by collaborating with ecologists or ethnobiologists. By assuring greater validity of research for linguists, culminating in cross-disciplinary materials re-applicable to other disciplines, and necessitating continued collaboration to further develop its interdisciplinary concerns, IDR moreover renders our model of documentation more sustainable.

Documentary linguists must nonetheless be mindful of the limits of interdisciplinary methods. Domain-based language research can overestimate the equivalency of linguistic representational strategies across speech communities and so distort linguistic documentation through the Western lens. What is more concerning, however, is that the logistical challenges faced by interdisciplinary collaboration (its lesser academic exposure, fewer opportunities for funding, and resultant poor teaching) have arguably been equally instrumental in bringing about its "haphazard" and "high-risk" practice as a language documentation methodology.

Beyond greater funding and publishing allowances being made for interdisciplinary documentation, increased disciplinary attention on fostering appropriate IDR practices could represent a solution to these problems. Such practices refer both to linguists offering a fair deal in terms of their collaborators' research concerns, as well as being sufficiently well-versed in IDR approaches to know their benefits and limitations within different linguistic contexts. This could be accomplished in part by university and other independent linguistics research programs reviewing to what extent and efficiency they teach interdisciplinary methods. Awareness of proper practices could also be raised by linguists engaged in IDR regularly publishing papers separate from their documentary results, which reflect solely on whether/how they succeeded at conducting their study in accordance with correct interdisciplinary methodology. Writers would discuss in what ways they balanced linguistic and extra-disciplinary research aims, outline any contextual obstacles for IDR related methods they encounter, detail the practicalities and offer suggestions. Above all, these papers would reinforce the idea that linguists should be actively assessing such qualities within their work throughout an IDR-based documentation. Improvement in the tutelage of interdisciplinary linguistic research could secure a yet higher level of validity within future IDR documentary results, and by this merit also plausibly heighten the method's internal academic standing. At the very least, advising linguists on how to better support other disciplines during their collaborative work could secure greater extra-disciplinary approval and funding.

References

Akumbu, Pius W. 2020. Reflections on a community-based approach to writing grammars of endangered African languages. *Studies in African Languages and Cultures* 54. 71–96.

Amith, Jonathan. 2020. Endangered language documentation: The challenges of interdisciplinary research in ethnobiology. (Ed.) Susan D. Penfield. *Language Documentation & Conservation* SP21: Interdisciplinary Approaches to Language Documentation. 72–112.

Burenhult, Niclas. 2020. Domain-driven documentation: The case of landscape. (Ed.) Susan D. Penfield. *Language Documentation & Conservation* SP21: Interdisciplinary Approaches to Language Documentation. 9–23.

Campbell, Lyle, Terrence Kaufman & Thomas C. Smith-Stark. 1986. Meso-America as a linguistic area. *Language* 62(3). 530–570.

Derrick, Edward G., Holly J. Falk-Krzesinski, Melanie R. Roberts & S. Olson. 2011. Facilitating interdisciplinary research and education: A practical guide. report from the Science on FIRE: Facilitating Interdisciplinary Research and Education workshop of the American Association for the Advancement of Science.

Dobrin, Lise M. & Saul Schwartz. 2016. Collaboration or participant observation? Rethinking models of "linguistic social work." *Language Documentation & Conservation* 10. 253–277.

Gadgil, Madhav, Fikret Berkes & Carl Folke. 1993. Indigenous knowledge for biodiversity conservation. *Ambio*. 22(2/3). 151–156.

Good, Jeff. 2020. Interdisciplinarity in areal documentation: Experiences from Lower Fungom, Cameroon. (Ed.) Susan D. Penfield. *Language Documentation & Conservation* SP21: Interdisciplinary Approaches to Language Documentation. 43–71.

Holton, Gary. 2018. Interdisciplinary language documentation. In Kenneth L. Rehg & Lyle Campbell (eds.), *The Oxford Handbook of Endangered Languages*, 738–760. Oxford University Press.

Klubnikin, Kheryn, Cynthia Annett, Maria Cherkasova, Michail Shishin & Irina Fotieva. 2000. The sacred and the scientific: Traditional ecological knowledge in Siberian river conservation. *Ecological Applications* 10(5). 1296–1306.

Martin, Gary J. 2004. *Ethnobotany: A Methods Manual*. London: Routledge.

Penfield, Susan. 2018. Funding the documentation and revitalization of endangered languages. In Kenneth L. Rehg & Lyle Campbell (eds.), *The Oxford Handbook of Endangered Languages*, 784–802. Oxford University Press.

Penfield, Susan D. 2020. Introduction: Interdisciplinary approaches to language documentation. (Ed.) Susan D. Penfield. *Language Documentation & Conservation* SP21: Interdisciplinary Approaches to Language Documentation. 1–8.

Ramstad, Kristina M., N. J. Nelson, G. Paine, D. Beech, A. Paul, P. Paul, F. W. Allendorf & C. H. Daugherty. 2007. Species and cultural conservation in New Zealand: Maori traditional ecological knowledge of tuatara. *Conservation Biology* 21(2). 455–464.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace.

Si, Aung. 2011. Biology in language documentation. *Language Documentation & Conservation* 5. 169–186.

University of Michigan. 2013. Rada Mihalcea selected for NSF INSPIRE Award. *Computer Science and Engineering* https://cse.engin.umich.edu/stories/rada-mihalcea-selected-for-nsf-inspire-award/ (1April 2021).

Widlok, Thomas. 2005. Cross-disciplinarity in the documentation work of anthropologists and linguists. *Language Documentation and Description* 3. 12–21.

SOAS Working Papers in Linguistics. Volume 20 (2021): 181–189

Linguistic triage: Documentary linguistics and language revitalisation

Victoria Baldwin

Abstract
There has been considerable discussion of the question as to whether a linguist engaged in language documentation should also be involved in language maintenance and revitalisation projects. This begs the question of what qualifications are needed to meaningfully engage in revitalisation and maintenance, and whether a linguist should participate even if they lack them. The relationship of the linguist to the language community has altered significantly in recent years, and yet outdated perceptions of the linguist's role persist. Simultaneously, a new paradigm has emerged wherein there is an expectation on field-working linguists to engage with and contribute to language communities in ways that some fear may interfere with the goal of rigorous linguistic scholarship. This paper will explore the changing expectations of linguistic fieldwork, who is really qualified for revitalisation and maintenance, and how the relationship and responsibilities of the linguist to the community in which they work may impact the way in which language documentation is carried out.

Keywords: language documentation, language revitalisation, language acquisition, applied linguistics, linguistic fieldwork

As founder of the Piegan Institute, Darrell Kipp was at the forefront of researching, promoting and developing materials for Native American languages. Possessing more than one master's degree and years of experience in the field, Kipp was well qualified to talk about the state of endangered languages and the process of their revitalisation:

> "Don't hire linguists… They can speak the language, but the kids won't and in bilingual education they still won't. Nothing against linguists, they can talk the language, but they don't act like us. They are not us; they are recorders." (Kipp 2000: 192).

Kipp's sentiment emphasises the notion that the linguist is primarily concerned with recording data. At the same time, Kipp refers to linguists as speakers and teachers of language, an idea which Gerdts (2010) considers to be outdated but which persists, nonetheless. Kipp seems here to suggest that linguists are singularly unsuited to participate in maintenance and revitalisation and this attitude reflects a tension that has been present in discourses of language documentation and revitalisation. Kipp's attitude also echoes that of Dorian, who, regarding the revitalisation of an endangered language, said, "such rewards cannot be supplied from the outside, they are to be had from within the social web of the community itself or not at all" (Dorian 1998: 21). This statement reflects the current dominant paradigm in documentation and revitalisation: that it is not only a matter of enlisting the participation of the language community in question but that the language community must, in all respects, take ownership of the endeavour. And yet, as Grenoble and Whaley remind us: "as field linguists and anthropologists know, it is rare to be working on the documentation and description of a potentially endangered language without confronting the issue of revitalisation" (Grenoble & Whaley 2005: 192).

It is clear then, that it is not simply a case of *if* linguists should participate in language revitalisation but rather that linguists *do* participate. The question that might more helpfully be asked is, how can linguists meaningfully contribute to language revitalisation while they are working with language communities. To understand this situation, it is helpful to clarify firstly, what is the purpose of language documentation and what is its relationship to revitalisation. Secondly, it is worth reflecting on how linguists might best approach work on language revitalisation and the issues that arise from those approaches, in particular the role of applied linguists and the particular challenges they face working with endangered or minority languages.

1. Academic Rigour

A good starting point is to consider the question of what the primary focus of documentary linguistics really is. Newman states: "the primary justification for doing research on an endangered language has to be the scientific value of providing that documentation and in preserving aspects of that language and culture for posterity" (Newman 2003: 6). Newman's concern is that excessive focus on servicing the needs of the community, or as he describes it, participating in "social work", detracts from the ability of the linguist to carry out effective, rigorous research. This concern is not unreasonable since for Newman, as for any serious linguist, academic contributions are paramount.

However, it is excessive to argue that social considerations will necessarily diminish the quality of the research. Himmelman, for example, explicitly links language documentation with language maintenance: "Language documentations are not only [seen] as data repositories for scientific enquiries but also as important resources for supporting language maintenance" (Himmelmann 2006: 5). The issue here then is perhaps the fear that considerations of supporting revitalisation efforts might somehow negatively impact the quality of scholarship or perhaps the suggestion that the motivations of language documentation are driven by a kind of saviour complex rather than a desire to produce sound research. Without doubt, the majority of language documenters would surely vigorously reject such a suggestion. As Ameka argues: "Language documenters and theorists strive for the scientifically sound and intellectually sophisticated empiricism that Newman is calling for" (Ameka 2015: 20).

It is in fact disingenuous to express concern that consideration of revitalisation is somehow detrimental to the scientific rigour of language documentation projects when the consensus among documentary linguists themselves is that many documentation project proposals merely "pay lip service to revitalisation" (Hinton 2011). This is echoed by Austin & Sallabank who contend that: "language revitalisation has been seen by some documentary linguists as a simple technical 'add-on' to their research […], rather than as a field of research or activity that requires theoretical or applied knowledge" (Austin & Sallabank 2018: 10). This criticism has been shared by Nathan & Fang who argue that language documentation as a whole does not hold itself accountable for the health of languages and that there are, in fact, urgent human needs arising from language endangerment that must be addressed (Nathan & Fang 2013). The reality is that scientific rigour has remained the primary focus of documentary linguistics, to the detriment of revitalisation rather than the other way around. So, while it is of course the concern of every serious scholar to maintain rigorous standards, it simply isn't the case that factoring

the needs of revitalisation into planning and execution of documentation projects negatively impacts the work.

However, while there is increasing support for the view that language documentation "should involve the community so as to support the maintenance and revitalisation of the language as well as increase its documentary capacity" (Akumbu 2018: 267), it is often the case that the interests of the community and the linguist do not neatly align. Even where the linguist approaches documentation from a revitalisation perspective, the connection between the research and the community's language goals may not immediately be apparent. "Community members are often more interested in revitalisation than documentation, which to them has less obvious immediate benefits" (Austin & Sallabank 2018: 10). It is necessary therefore to consider the underlying assumptions that lead to this misalignment.

2. The embattled linguist

As Newman (2003) discussed, there has been a divergence in recent decades between linguistics and anthropology, disciplines that were previously closely entwined. Gerdts (2010) points out a similar phenomenon whereby linguistics and teaching were formerly much more closely aligned. This is pertinent in two respects: firstly, it demands a reflection on the attitudes of linguists towards fieldwork, and secondly, an examination of the perception that linguists know how to teach language or are interested in learning the language that they are documenting.

While Newman (2003) complains of a dearth of linguists willing to undertake fieldwork, it might be argued that the development of language documentation in the intervening years proves that this was never or is no longer true. It cannot, in all honesty, be claimed that there are no linguists willing to go out in the field, and there are numerous examples of linguists who are also community members for whom fieldwork in their own language is their primary focus. However, Newman's observations about the decline of fieldwork may go some way to explaining negative perceptions of linguists as lacking the right qualifications and being unsuitable for work towards the maintenance and revitalisation. Where the community has the expectation that the linguist should be committed to social fieldwork, with an interest not only in the description of hard linguistic data but all the socio-cultural aspects that accompany it, as might be an anthropologist, linguists may in fact have approached fieldwork simply as a means to complete research for academic purposes and not because of their understanding of or commitment to the language community. Furthermore, Gerdts (2010) suggests that recognition for fieldwork in documentation has been limited and claims that research produced in the field has generally received less acclaim from the wider linguistic field. This would support Newman's claims about the decline of linguistic fieldwork.

Additionally, as language teaching and acquisition have diverged into distinct subfields of linguistics, the field-working linguist lacks motivation or knowledge about how the process of learning and teaching language operates. There remains perhaps a hangover from the days of the missionary school, a favoured haunt of linguists and anthropologists in times gone by. Linguists today however will often be singularly unsuited to this type of work simply due to their academic training and interests, as well as a desire to rebuff any association with missionary work and all the connotations that term brings. Add to

this the common misconception that linguistics means learning languages, and we have a situation where the community may have had false perceptions of the academic outsider. This would give rise to expectations of the linguist which cannot be met and result in attitudes like that of Kipp. Essentially, a situation arises whereby community perceptions of linguists in the field rely on outdated notions of the linguist as anthropologist, teacher and linguist all in one, expected to carry out the function of all three and as a result doing none of them particularly well.

There are of course broader social and political issues which must also be considered, notably the consequences of colonialism, and the extent to which documentary linguists should strive to "decolonise" their work and practice. There is the suggestion that the discipline of language documentation and its introduction of academics into endangered language situations perpetuates colonial inequalities since language documenters bring colonial norms with them through their linguistic training (Leonard 2018). This echoes Ladefoged, who cautions that: "We should always be sensitive to the concerns of the people whose language we are studying. But we should not assume that we know what's best for them" (Ladefoged 1992). Of course, it might equally be argued that minority and endangered language communities often languish under colonial socio-political structures that were imposed on them from the outside and that ignoring them is as equally problematic as engaging with them (Dorian 1998). Newman, however, shares a word of caution with the would-be linguist: "Language policy in fragile multi-ethnic states is not a simple sociolinguistic matter, but rather it is a serious, highly contentious matter in which a foreigner should not become embroiled" (Newman 2003: 6). This warning likewise applies to local linguists, be they part of the community or the wider state which it inhabits, or indeed be legally a foreigner while simultaneously a member of the language community. They too must navigate the prevailing political winds, and often with less protection and fewer means of escape from the ramifications of their work.

3. Documentation isn't revitalisation

It is clear that there are strong arguments in favour of the view that language documentation cannot ignore calls to engage with revitalisation. However, having considered whether or not documentation ought to be driven by revitalisation, and the position of the linguist to it, it must also be noted that methods of documentation and revitalisation often remain quite distinct. The classical outputs associated with documentation such as making audio-visual recordings, producing dictionaries, corpora etc, do not in themselves guarantee successful or meaningful revitalisation or maintenance of a language. Such materials are often not easily accessible to those developing curricula, be that as a result of poor file organisation, inadequate metadata or simply lack of communication between academic and community as to where documented materials will end up. Additionally, the academic and scientific jargon of linguistic analysis and a tendency to focus on classical linguistic features of phonology and morphology are not especially helpful for a teacher attempting to develop communicative competence (Austin & Sallabank 2018). This stems partly from a lack of consideration for pedagogical applications in the planning phase of a language documentation project. As such, the archived materials may not contain the kind of subject matter that is helpful for language acquisition, revitalisation, or maintenance. Topics may often be unsuitable for learners, particularly children, documenting sacred or taboo material, or relating to sex and death (Austin & Sallabank 2018). On top of this,

there is often disagreement within communities and across generations as to what constitutes "proper" language and of who speaks it "nicely" enough to serve as a model for future learners, as well as issues around contending with a lexicon that needs to work in the modern world (Grenoble & Whaley 2005). When these factors combine, the result may be that archived materials sit like specimens under glass in a rarely visited wing of a museum, hardly if ever seen and of no particular use even to those that might be interested in them. This is the complaint of Nathan & Fang (2013) who argue that the process of documentation is a "one way journey" of a language into an archive, ignoring its human characteristics and re-enacting the devalorisation which endangered it in the first place.

4. Applying Linguistics

Further to the discussion of how linguistic outputs can be relevant to revitalisation, Nathan & Fang (2013) complain that pedagogical materials have been weakly fostered by language documentation and generally not encouraged by practice. Hinton observed that "to a large extent, the models, methods and materials for second language teaching and learning are developed by bootstrap strategies within revitalization programs" (Hinton 2011).

The obvious solution seems to lie with applied linguists, however, as Penfield & Tucker make apparent, the applied linguist is not immediately suitable to work in the field since firstly, their training mostly lacks many of the classical tools of linguistics used in documentation (and description) for example comprehensive knowledge of typology or historical linguistics, and secondly, that the majority of scholarship in applied pedagogy is not suitable for an endangered or minority language (Penfield & Tucker 2011). Applied linguistics has an overwhelming focus on dominant and majority languages, such as English, Spanish or Chinese. As Grinevald (1998) cautions, revitalisation is more complicated than simply getting applied linguists out into the field: "Language revitalisation needs to be recognised as a special area of second language acquisition and second language teaching. It is not enough for 'straight' linguists to think that such projects are the domain of educators and applied linguists." (Grinevald 1998: 158).

While research into this area is growing, for example, McPake et al (2017) who carried out a comparative study into professional development for teachers in Maori, Basque, Catalan and Welsh-medium teaching; minority language research in applied linguistics is as yet comparatively underdeveloped. Additionally, there remain two further issues that may not so easily be resolved. Firstly, the research underway is occurring in wealthy states who possess already well-developed scholarship on both majority and minority languages, as is the case with McPake et al. And secondly, the issue that such training alone does not necessarily make an applied linguist suitable for revitalisation of an endangered language.

While there are clearly sound reasons for increased use of applied linguistics and pedagogical methods in language documentation and revitalisation, it must be noted that dominant discourses of pedagogy and second language acquisition rely heavily on both orthography and literacy, which in themselves entail a host of issues for endangered languages. Grenoble & Whaley argue that literacy is in fact not helpful for language diversity because literacy is an arena that requires and thrives on standardised language forms (Grenoble & Whaley 2005). Indeed, they are critical of situations where a well-

meaning applied linguist has attempted to use majority language materials to create educational programmes for minority and endangered languages. This may have the effect of reiterating the same colonial processes that operated to endanger the language in the first place. This is supported by Ameka, who claims that language standardisation does not preserve diversity but rather decreases it (Ameka 2015). Mufwene goes further, stating: "It appears that the development of writing systems for, and literacy in, some endangered languages guarantees not their revitalisation but their (lifeless) preservation in a jar" (Mufwene 2003: 5). With this in mind, the creation of literacy programmes, a standard part of language pedagogy, will not in fact support language revitalisation but instead perpetuate the "one-way journey" into archival obscurity that Nathan & Fang (2013) so strongly protest.

However, it must be noted that a meaningful contribution towards considerations of revitalisation does not solely require the linguist to create an orthography or become an expert in applied linguistics. Nathan & Fang (2013) state that at the planning stage of documentation, consideration of the domains of language should include culturally and socially relevant material that can be used in the development of materials for revitalisation. Nathan & Austin (2004) highlight that a focus on metadata and annotation can yield positive outcomes by increasing its accessibility to those who may not possess the technical linguistic knowledge to deal with it otherwise. Austin & Sallabank comment that: "a more sociolinguistic approach to documentation can identify learner groups, […] as well as potential teachers and consultants and their particular skills" (Austin & Sallabank 2018: 212). So, while calls for more attention to pedagogy and applied methods, in general, are valid, it is not necessarily a requirement for every linguist approaching documentation to acquire training in those areas. It is rather a call for more applied linguists to participate in language documentation projects as well as for a greater focus on minority and endangered language research within the fields of applied linguistics and language acquisition.

5. Conclusions

Reflecting on this discussion, it is clear that there has been extensive debate in the field of language documentation as to the nature of both the discipline and the role of the linguist. While it might be reasonably argued that linguistics as a discipline should be mostly concerned with gathering solid data in order to produce rigorous, scientifically driven analysis, it must be acknowledged that language, as a fundamentally human phenomenon, cannot be divorced from the social reality in which it exists. As such, it demands that fieldworkers and theorists alike operate in accordance with a set of ethics that respects this fact. Indeed, as Ameka made clear, language documentation is not, as Newman (2003) suggests, degraded by this social aspect, but instead has developed frameworks to ensure that it produces meaningful and rigorous data while attending to the ethical demands that accompany it (Ameka 2015).

Additionally, while the dominant paradigms in language documentation have made great efforts to ensure the highest possible standards in the field, there have been continuing criticisms that the practices of linguists have in fact often perpetuated the very causes of language endangerment, be that the reinforcement of colonial inequalities or the reification of languages into artefacts rather than considering them as living, albeit

intangible, entities. In response, there has been an examination of how language documentation might best be re-conceptualised to support and enable revitalisation.
Perhaps most strikingly, it has become clear that there has been a false perception of linguistics and as such linguistic training has lacked adequate focus on applied linguistics that would enable language documentation projects to effectively contribute in the manner that language communities expect. Likewise, it is clear that knowledge of applied linguistics alone is insufficient. Indeed, this simply reiterates the notion that language documentation must become increasingly interdisciplinary and collaborative. While linguistic training in language documentation is becoming increasingly more specialised so that it can produce scientifically rigorous work, as well as promoting a collaborative approach that accounts for and includes the rights and needs of language communities, it remains problematic to simply suggest that language documenters study some applied linguistics courses or vice versa. There remains the issue that applied linguistics brings with it its own set of potentially detrimental effects. These include an emphasis on pedagogy derived from models which are rooted in western concepts of literacy-based knowledge, which itself can operate against language diversity instead of encouraging it.

And yet, while it might seem that there is little support today for the notion that language documentation can operate in isolation from the concept of revitalisation, it must be noted that a lack of rebuttal from linguists is not necessarily a sign of agreement. It is too simplistic to accept linguists' silence on the matter as evidence of agreement or commitment to revitalisation. If a linguist, applied or otherwise, has little or no interest in minority languages or revitalisation, it is doubtful they would feel compelled to enter into a debate on how and why they should include it in their work, let alone make efforts to factor it into their practice. It is, in fact, the language documenters themselves who have made the effort to address the perceived disconnect between documentation and revitalisation. It is their support for revitalisation that presses them to address valid concerns about the scientific rigorousness of data, or about the contentious political and ideological circumstances that surround endangered languages and of the failings of documentation as a discipline to enable meaningful revitalisation and maintenance. And so it remains incumbent on documentary linguists to make noise on this matter, as loudly and as publicly as possible, because to have engaged in a meaningful way with this issue can only lead one to the conclusion: that the linguist can no longer be viewed as the detached researcher operating as a kind of phantom butterfly collector, intent on gathering specimens to put under glass but more as a physician tending to a living, breathing patient whose health is under threat. Documentary linguistics, while imperfect, has thus far operated much like a kind of battlefield medic performing linguistic triage, diagnosing the patient but perhaps not entirely qualified to cure them. However, until specialist pedagogy for endangered and minority languages is more fully developed and applied in the field, a battlefield medic is better than no medic at all.

References

Akumbu, Pius W. 2018. Babanki literacy classes and community-based language research. In Shannon T. Bischoff & Carmen Jany (eds.), *Insights from practices in community-based research: From theory to practice around the globe*, 266–279. De Gruyter Mouton.

Ameka, Felix K. 2015. Unintended consequences of methodological and practical responses to language endangerment in Africa. In James Essegbey, Brent

Henderson & Fiona Mc Laughlin (eds.), *Language documentation and endangerment in Africa*, 15–36. John Benjamins Publishing Company.

Austin, Peter K. & Julia Sallabank. 2018. Language documentation and language revitalization: Some practical considerations. In Leanne Hinton, Leena Huss & Gerald Roche (eds.), *The Routledge Handbook of Language Revitalization*, 207–215. New York: Routledge.

Dorian, Nancy C. 1998. Western language ideologies and small-language prospects. In Lenore A. Grenoble & Lindsay J. Whaley (eds.), *Endangered Languages*, 3–21. Cambridge: Cambridge University Press.

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

Grenoble, Lenore A. & Lindsay J. Whaley. 2005. *Saving languages: An introduction to language revitalization*. Cambridge: Cambridge University Press.

Grinevald, Colette. 1998. Language endangerment in South America: A programmatic approach. In Lenore A. Grenoble & Lindsay J. Whaley (eds.), *Endangered languages: Language loss and community response*, 124–160. Cambridge: Cambridge University Press.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*, 1–30. Berlin: Mouton de Gruyter.

Hinton, Leanne. 2011. Language revitalization and language pedagogy: New teaching and learning strategies. *Language and Education* 25(4). 307–318.

Kipp, Darrell R. 2000. *Encouragement, guidance, insights, and lessons learned for Native language activists developing their own tribal language programs*. Browning, MT: Piegan Institute's Cut-Bank Language Immersion School.

Ladefoged, Peter. 1992. Another view of endangered languages. *Language* 68(4). 809–811.

Leonard, Wesley Y. 2018. Reflections on (de)colonialism in language documentation. (Ed.) Bradley McDonnell, Andrea L. Berez-Kroeker & Gary Holton. *Language Documentation & Conservation* SP15: Reflections on Language Documentation 20 Years after Himmelmann 1998. 55–65.

McPake, Joanna, Ann Macdonald, Mona Wilson, Fiona O'Hanlon & Mary Andrew. 2017. *Transformative pedagogies for Gaelic revitalisation: Report to Soillse of a study of Gaelic-medium teachers' perspectives on the potential of translanguaging as a classroom pedagogy*. Soillse.

Mufwene, Salikoko S. 2003. Language endangerment: What have pride and prestige got to do with it. In Brian D Joseph, Johanna DeStefano, Neil G. Jacobs & Ilse Lethiste (eds.), *When languages collide: Perspectives on language conflict, language competition, and language coexistence*, 324–326. Columbus: The Ohio State University Press.

Nathan, David & Peter K. Austin. 2004. Reconceiving metadata: language documentation through thick and thin. *Language Documentation and Description* 2. 179–187.

Nathan, David & Meili Fang. 2013. Re-imagining documentary linguistics as a revitalization-driven practice. In Mari C. Jones & Sarah Ogilvie (eds.), *Keeping languages alive: Documentation, pedagogy and revitalization*, 42–55. Cambridge: Cambridge University Press.

Newman, Paul. 2003. The endangered languages issue is a hopeless cause. In Mark Janse & Sijmen Tol (eds.), *Language death and language maintenance: Theoretical, practical and descriptive approaches*, 1–14. Amsterdam: John Benjamins Publishing.

Penfield, Susan D. & Benjamin V. Tucker. 2011. From documenting to revitalizing an endangered language: Where do applied linguists fit? *Language and Education* 25(4). 291–305.

SOAS Working Papers in Linguistics. Volume 20 (2021): 190–194

A review of *Archiving for the future: Simple steps for archiving language documentation collections*

Victoria Baldwin

Archiving for the future: Simple steps for archiving language documentation collections (Kung et al. 2020) (henceforth AFTF) is an online course that sets out to comprehensively prepare those engaged in language documentation projects to meet a framework of archiving for language documentation which has been developed over the last two decades. To effectively evaluate it, it is necessary to briefly explore key aspects of that framework.

In response to changing technology, Bird and Simons (2003) set out a blueprint to ensure the coherence, quality, and longevity of digitised data in language documentation. As the volume of documented material increased, Nathan & Austin (2004) advocated creating detailed metadata to enhance accessibility to an increasing "quicksand" of information. Holton (2012) showed that language archives have scientific and pedagogical applications beyond linguistics and therefore should be designed with these broader uses in mind. Nathan (2014) emphasised the need for access by communities to materials about their language and culture. This idea has since been established in doctrines such as the FAIR Principles (Wilkinson et al. 2016), and the CARE Principles for Indigenous Data Governance (Carroll et al. 2020).

On this basis, substantial infrastructure has emerged with archives such as the Endangered Languages Archive (ELAR) and the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). Specialist software, like ELAN and Lameta, has also been developed. This documentary infrastructure aims to ensure that data for archiving is good quality, technologically future-proofed and well supported by metadata. It must be accessible to a variety of users and adhere to established ethical principles.

AFTF attempts to integrate these approaches and infrastructure into a single, accessible training resource. The training is divided into three phases, plus an introduction and conclusion. The introduction establishes its purpose and goals and presents the structure for subsequent sections. The introduction goes on to provide an overview of language documentation since the time of linguistic pioneers like Franz Boas, much in the same vein as the manner described by Henke and Berez-Kroeker (2016). A clear visual (see Figure 1) allows participants to develop a picture of how the course will proceed. Participants are then invited to carry out a kind of "pre-task" activity, aimed at eliciting prior knowledge of collecting and organising data by considering their own collections and asking them to create their own inventory. The aim of such a task is to contextualise the topics that will be covered in the course and prepare participants to learn. The introduction concludes with a summary of key vocabulary and provides suggestions for further reading. This basic structure of introducing concepts, reflection, vocabulary, and reading is repeated throughout subsequent sections.

Figure 1 Phases One to Three, from Introduction to Archiving for the Future: Simple Steps for Archiving Language Documentation Collections

Phase One covers the planning stage of language documentation projects, guiding language documenters through the contemporary framework of the discipline at every step of their project. Step One introduces some well-known archives, including many belonging to The Digital Endangered Languages and Musics Archives Network (DELAMAN) and others. Step Two discusses the organisation of files, while Step Three delves into the kind of technical and media quality issues raised by Jukes (2011), for example, opting for open formats rather than proprietary ones, or planning video for streaming rather than archiving large video files that are cumbersome to download. In this way, accessibility is built into the archiving process. Step Four introduces the concept of metadata, using visuals to simplify concepts. (See Figure 2.)

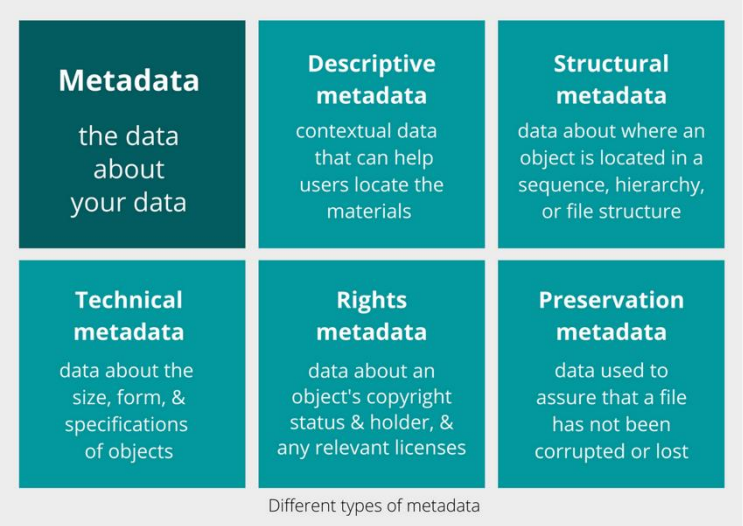


Figure 2 Visual aid for understanding metadata, from Section Two, Phase One, Step Four

Phase Two develops the discussion around metadata, with Step Five introducing various tools and approaches used to create and manage metadata in the field. For example, advocating the use of software such as Lameta or SayMore for collating and tracking

metadata, or inviting documenters to consider issues around access such as copyright or licensing. Step Six discusses appraisal of collected materials, including sensitivity to Traditional Knowledge, and other forms of intangible cultural heritage. For example, communities having strong views toward certain cultural artefacts which make their deposit in an archive problematic. AFTF makes useful and ethical suggestions on how to resolve such issues. Examples of issues discussed are shown in Figure 3.

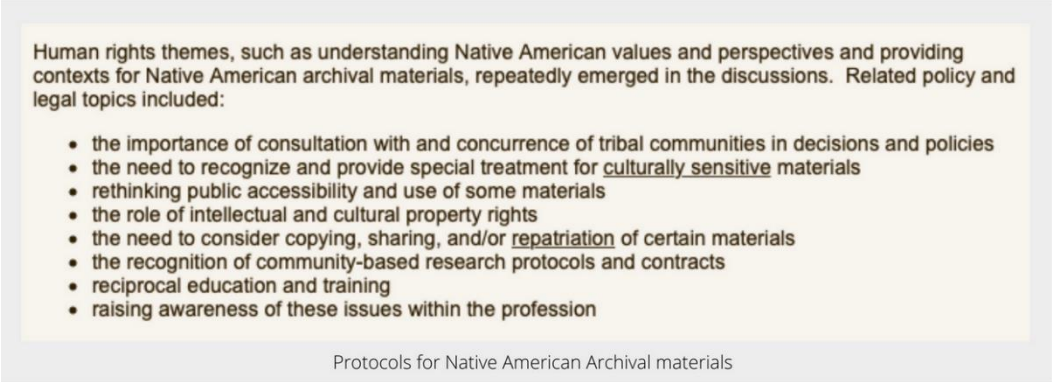


Figure 3 Considerations in Archiving Traditional Knowledge, from Step Six

Phase Three discusses deposit into an archive. Step Seven explores strategies for arranging materials into coherent bundles of data. Suggestions include opting for a flat file structure, rather than a nested one since it is easier for archive users to navigate. It might also encourage depositors to arrange files by association (e.g., all files relating to a particular recorded event) rather than by type (i.e., all video files in one folder, all transcripts in another and leaving archive users to muddle their way through). Not only does this keep materials coherent, but it also increases accessibility. Step Eight introduces the concept of progressive archiving, as discussed by Nathan (2013), where essentially materials are added to a collection one piece at a time, rather than as one unwieldy mass of data that must be organised before the entire collection can be archived. This is helpful both as a means to maintain quality but also as a tool for managing workflows around large volumes of data. Step Nine discusses issues of accessibility through the production of a collection guide, as expanded most recently by Sullivant (2020). This essentially is a document detailing the key features and content of the archived collection with the intent of making it easier for people to access and make use of that content.

The conclusion shares citations, acknowledgements, and attributions. Participants may also request a certificate of completion.

AFTF is clearly built around the core framework of contemporary language documentation. As an educational tool, it provides a thorough introduction to the processes and tools of language documentation. The format has been designed with pedagogical principles in mind. Visual aids and video are used in addition to text, as well as a 'pre-task' and reflection upon each topic. The activities at the end of each section enable learners to develop practical experience of each step of the actual archiving process, from planning to submission. There are also suggestions for further reading and resources to allow participants to engage with topics in more detail outside the scope of the course.

This resource is helpful for those undertaking their first language documentation project, particularly those coming from multidisciplinary teams who may not have much language documentation background to draw on, or for language community members who are carrying out documentation under their own steam. The discussion of progressive archiving is also helpful since it makes the whole enterprise of language documentation less intimidating by encouraging smaller, neater submissions to archives. This has the effect of encouraging language documentation since it reframes it as something more achievable and manageable than the commonly held perception of language documentation projects as huge, complex undertakings. In addition, considering Dwyer's urging to "do some good" – not only in the sense of compensating language communities for their contributions but also by adding value through education or empowerment – the certificate is a nice touch, particularly for community members collaborating in language documentation projects (Dwyer 2006). This, of course, depends on their ability to read English, which is perhaps the only drawback of this resource. However, since it is so new, having only been released in 2020, it cannot be too harshly criticised for not yet being available in multiple languages. Additionally, The Archive of the Indigenous Languages of Latin America (AILLA 2020) have already begun to make the video content for AFTF available in Spanish via their YouTube channel, and it is likely that this multilingual expansion is something that will continue to be developed in future. Overall, *Archiving for the future: Simple steps for archiving language documentation collections* is a well-designed and well-structured tool that easily achieves its stated goals of bringing together the main concepts and tools of contemporary language documentation into a single, accessible resource.

References

AILLA. 2020. *Tipos de arquivo*. AILLA Archive of the Indigenous Languages of Latin America. https://youtu.be/trl5EEVg950

Bird, Steven & Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3). 557–582.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19. 43.

Dwyer, Arienne. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 31–66. Berlin, New York: Mouton de Gruyter.

Henke, Ryan & Andrea L. Berez-Kroeker. 2016. A Brief History of Archiving in Language Documentation, with an Annotated Bibliography. *Language Documentation & Conservation* 10. 411–457.

Holton, Gary. 2012. Language archives: They're not just for linguists any more. (Ed.) Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts & Paul Trilsbeek. *Language Documentation & Conservation* SP3: Potentials of language documentation: Methods, analyses, and utilization. 105–110.

Jukes, Anthony. 2011. Culture documentation and linguistic elicitation. In Nicholas Thieberger, Linda Barwick, Rosey Billington & Jill Vaughan (eds.), *Sustainable data from digital research: Humanities perspectives on digital scholarship*, 47–62. Melbourne: Custom Book Centre, University of Melbourne.

Kung, Susan Smythe, Ryan Sullivant, Elena Pojman & Alicia Niwabaga. 2020. *Archiving for the Future: Simple steps for archiving language documentation collections*. https://archivingforthefuture.teachable.com/p/home (22 July 2021).

Nathan, David. 2013. Progressive archiving: Theoretical and practical implications for documentary linguistics. Presented at the 3rd International Conference on Language Documentation and Conservation (ICLDC).

Nathan, David. 2014. Access and accessibility at ELAR, an archive for endangered languages documentation. (Ed.) David Nathan & Peter K. Austin. *Language Documentation and Description* 12. 187–208.

Nathan, David & Peter K. Austin. 2004. Reconceiving metadata: language documentation through thick and thin. *Language Documentation and Description* 2. 179–187.

Sullivant, Ryan. 2020. Archival description for language documentation collections. *Language Documentation & Conservation* 14. 520–578.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1). 160018.