Prosodic Patterns in Ramari Hatohobei

Vasiliki Vita

675802

MA Language Documentation and Description

This dissertation is submitted in partial fulfilment of the requirements for the degree of MA Language Documentation and Description of SOAS, University of London

30 September 2020

Word Count: 9909

I have read and understood the School Regulations concerning plagiarism and I undertake:

- That all material presented for examination is my own work and has not been written for me, in whole or in part by any other person(s).
- That any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the dissertation
- That I have not incorporated in this dissertation without acknowledgement any work previously submitted by me for any other module forming part of my degree.

Table of contents

Abstract

With only one grammar describing the languages of Sonsorol-Tobi and only its phonetics, this dissertation focuses on describing prosodic patterns in Ramari Hatohobei, or Tobian, a severely endangered Micronesian language. The primary aim is to contribute to the description of Ramari Hatohobei based on data from the ELAR collection, "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language" (Black and Black, 2014). Another aim is to identify the extent to which such data could be useful for linguistic description and in particular to the field of phonology and phonetics. Spectrograms have been extracted using Praat from conversations, descriptions and stories and the ToBI conventions have been used for the analysis of prosodic patterns. Furthermore, the curators and speakers have been consulted in order to investigate particular hypotheses. Due to my personal interest in documenting Sonsorolese, a closely related language, this dissertation could potentially become an axis in distinguishing the different prosodic patterns between the two languages.

Acknowledgments

This research project would not have been possible without the feedback and support from my supervisor Lutz Marten. I would also like to thank George Starling for his teachings on ELAN and Praat and his help with the methodology and results chapters of the project.

Thank you to Aoife for helping me with R and Louis for the emotional support; the Lady Linguists of SOAS, Sydney, Yewa, Leo, Maja and Sravya, for their commitment from the first moment of following this path of diversity and excitement.

Peter and Barbara Black, I thank you for the work you have done all these years with the people of Tobi and for your cultural and linguistic insights, as well as your support and heartfelt advice throughout the research process.

Tintin and Justin Andrew, thank you, for being colleagues and, above all, teachers of the Tobian language and culture. Hapari ma hatawahi sewa.

I would also like to thank my family, Anastasia, Harris and Giannis, and friends, Christina, Ioanna, Maria, Ioanna, Federica, Hannah, Ella and Jihyun, for their support throughout the stressful process of writing this, and particularly for always welcoming me when I try to run away. Σας αγαπώ.

Finally, I would like to extend my heartfelt gratitude to Mitsuru Inoue, who has introduced me to Palau and its peoples. Sulang badaband.

List of figures

List of tables

Abbreviations

1	1st person
3	3rd person
C	Consonant
CONJ	Conjunction
DAMSL	Dialogue Act Markup in Several Layers
dB	Decibel
DEM	Demonstrative
ELAN	EUDICO Linguistic Annotator
ELAR	Endangered Languages Archive
ELDP	Endangered Languages Documentation Programme
EMP	Emphatic
EXCL	Exclamative
f0	Fundamental frequency
GIL	Gilbertese
Hz	Hertz
HAT	Hatohobeian (Tobian)
INCL	Inclusive
IPA	International Phonetic Alphabet
ISO	International Organisation for Standardization
KSR	Kosraean
MOD	Modality
MRS	Marshallese
NEG	Negation
OBJ	Object marker
PCC	Palau Community College
pdf	Portable Document Format
PL	Plural
PON	Pohnpeian
POSS	Possessive marker
PTK	Proto-Micronesian
PUA	Pulo Annian
PUL	Polowatese
RDP	Reduplication
SG	Singular
SOAS	School of Oriental and African Studies
SON	Sonsorolese
SpnCRL	Saipan Carolinian
TAM	Tense, Aspect and Mood
ToBI	Tones and Break Indices
V	a vowel
VV	a long vowel or diphthong
WOL	Woleaian

1. Introduction

This dissertation uses raw data of Ramari Hatohobei or Tobian (ISO 639-3 tox), an endangered Micronesian language, archived at the Endangered Languages Archive (ELAR) for the description of its prosodic patterns. The primary aim is to contribute to the description of Tobian and fill the gap in Regh's (1993) paper on Proto-Micronesian prosody. A secondary aim is to identify the extent to which archived data could be useful for linguistic description and in particular for the analysis of prosody. In the first part, a discussion on the role of language archives is presented with some background information on Tobian and Micronesian prosody. The second part describes the methodology used for the description of Tobian prosodic patterns, while in the third part, the analysis and results of the prosodic patterns in Tobian are presented, followed by a discussion on the usefulness and quality of the particular collection. In the concluding remarks, some ideas for future research are introduced.

According to Himmelmann (2006: v), language documentation is "concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties". Henke and Berez-Kroeker (2016: 411) underline the value of archiving data collected from a language documentation project. With the growing digitization of our world, linguists working with endangered language communities recognize the necessity of prioritizing digital preservation, especially in contexts where no speakers are left. Nevertheless, Holton (2012: 106) underlines that language archives nowadays have reached a point where they are considered digital reservoirs of not only linguistic structures but also sources of cultural documentation.

Data collected from a documentation project have often been used for the grammar writing of various endangered languages. However, the place of prosody in descriptive grammars has been a constant debate. Tradition exemplifies descriptive grammars with a limit of a phonemic inventory on the phonology section (Mosel, 2006: 52). That is because of various

reasons, primarily the lack of tools or resources, as well as techniques and training (Mosel, 2006: 51-52). Nevertheless, the description of the phonology and phonetics of the world's languages could contribute in distinguishing dialects, historical changes, variation as well as language acquisition (Ladefoged, 2003: 203). Additionally, studying sound distinction or as Ladefoged (2003: 203) calls it, phonetic fieldwork, could lead to the understanding of sounds that are difficult to be perceived by the non-native speaker and may be typologically unusual.

2. Language Archives

Austin (2010: 12) argues that by its nature, language documentation, is multidisciplinary since it combines ideas and methods from a variety of fields, such as linguistics, ethnography, computer science, and more. According to Austin (2010: 23), archiving is one of the five activities included in language documentation and, according to Himmelmann (2006: 15), one of its most significant features; preserving language data, whether that is in their raw, primary or theoritised form, for posterity, in order to ensure that it could be available to a variety of users and for various purposes.

Vinogradove (2016: 128) underlines that the techniques, principles and challenges of building a corpus of a small understudied or undocumented language are different in comparison to that of a major extensively-studied language. It is argued that such records or corpus are usually small, in comparison to the ones of major languages, not balanced and traditionally not available to researchers who did not participate in the collection (Vinogradove, 2016: 129-130). Additionally, certain data might originate from elicitations rather than naturally occurring speech in everyday contexts. Nevertheless, both corpus' material is machine readable and its size is finite (Vinogradove, 2016: 130). Then how should these records be used?

In the case of endangered-language documentation, such archived records include not only linguistic information but many other types of community knowledge (Holton, 2012:

106). They become the material and sources that document the language and any derivative or analytical material and provide data that could stimulate further research and analysis, which could be used for revitalisation efforts (Conathan, 2011: 236). Austin (2010: 25-26) also mentions that this should be the criterion for evaluating archival files, since they should be adequate for description and analysis and consequently potential maintenance or revitalisation.

These would essentially be the products of documentation and can develop over time as research progresses, in the sense that these descriptions and analyses are supported by evidence reducing the risk of making it "sterile, opaque and untestable" (Austin, 2010: 23). Berez-Kroeker et al. (2018: 5-6) underline the importance of possibilities of reproducible research in linguistics, particularly in the context of documentary linguistics, whose methods have the potential of providing "substantiation of scientific claims by promoting attention to the structuring and sharing of language data".

The importance of the metadata for such collections has been underlined extensively (Woodbury, 2003: 39-40, Himmelmann, 2006: 11, Austin, 2010: 19, Good, 2011: 231, Holton, 2012: 106, Berez-Kroeker et al., 2018: 8). Woodbury (2003: 41) mentions the "text curation" aspect of documentation, in the sense that depositors and linguists are encouraged to provide information on the location and date of the speech event, its participants and their roles, the genre, the communicative purpose and other features, all part of the ethnography of speaking. Furthermore, it is noted that this should not exclude linguistic elicitation (Woodbury, 2003: 42), rather it is the combination of both that creates a "representative and lasting multipurpose record of a natural language or one of its varieties" (Himmelmann, 2006: v).

Huvila (2008: 25) notes three main characteristics of the participatory archive. First of all, it assumes a decentralised curation, with archivists and participants curating the material together. Second of all, radical user orientation is encouraged where locating and using the appropriate material becomes more significant than the archival process. Finally, radical user

orientation also allows for community members and others to include their contributions.

3. The Republic of Palau: Linguistic landscape

The Republic of Palau is an independent nation state of the west Pacific (Matsumoto & Britain, 2012: 137). It exemplifies a case of diglossia (Matsumoto & Britain, 2015: 22) with the national language being Palauan, with both English and Palauan having official status (Matsumoto & Britain, 2000: 22). In other words, administrative issues and official documents are written in English, while Palauan is used in everyday, informal and local contexts (Matsumoto & Britain, 2015: 22).

Ramari Hatohobei, or Tobian (ISO 639-3: tox), is the language of Tobi, one of the Southwest islands of the Republic of Palau, and the main island of Hatohobei state (Matsumoto and Britain, 2006: 131). Belonging to the Chuukic, Micronesian group of the Austronesian family, it is part of a dialect continuum spoken in the southwestern islands of the Republic of Palau; Sonsorol, Pulo Ana, Merir and Tobi (Matsumoto and Britain, 2006: 131).

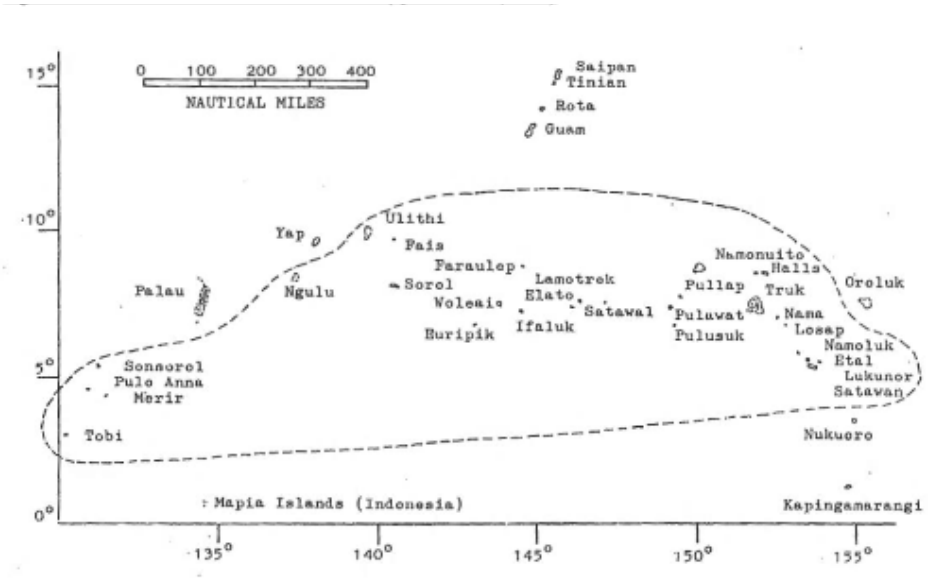


Figure 1. Boundaries of the Chuukic continuum (Quackenbush, 1968)

Although in the past the Southwest islands exhibited high numbers of population (van den Berg, 2014: 3), nowadays, the majority of islanders have migrated to the biggest island of

Palau for various reasons. Economic, health, educational and environmental (typhoons) reasons have led these peoples to the Echang village, where they are bilingual in Palauan and English (Taborosi et al, 2018). Young speakers tend to speak a mixture of Tobian, Sonsorolese (ISO 639-3 sov), a relative language, and English, different to the elder generation's linguistic choices (1).

(1) That's the thing, you know, ah, like now when I speak I always mix up my language. Like, I would use a Tobian language, like in a sentence, I could be mixing my Tobian, Sonsorolese and ah a few English words in it, haha, you know what I mean, haha (tb01-vva-fab-inter-2020-06-20-CH3-01).

Tobian and Sonsorolese are considered two of the most archaic languages spoken in the West Micronesian Sprachbund, at least at the phonological level. It is argued that the two languages are more similar to nuclear Micronesian languages rather than other Chuukic, since they preserve the word-final voiceless vowels on stems like Woleaian (ISO 639-3 woe) (Grant, 2017: 853). According to Capell (1969: 1), they exhibit close resemblance to Ulithian.

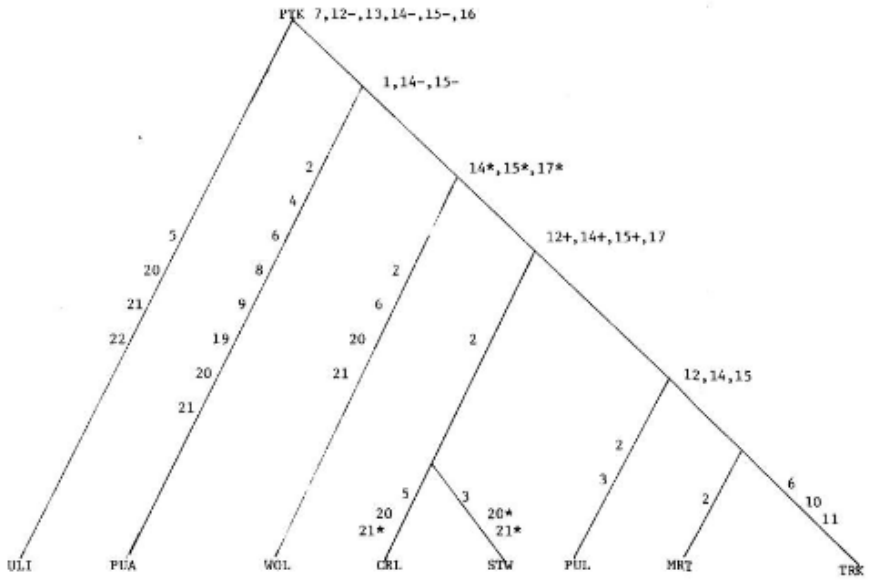


Figure 2. Genetic tree of Chuukic, based on diachronic phonological information (Jackson,

1983)

There is limited description of the language but there is an abundance of raw data on Kaipuleohone (Berg, 2013) and on the Endangered Languages Archive collection "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language" (Black and Black, 2014). Description of the language is limited to Capell's (1969) grammar and van den Berg's (2014) dissertation, both commenting on both languages. Both contain no information regarding prosody in the languages which are dealt together. Rather, most literature of Tobian and Sonsorolese is on how these two languages reflect word-final vowels (van den Berg, 2014: 13).

4. Micronesian Prosody

Jackson (1983: 232) notes that phonological, lexical and lexicostatistical evidence identify the close relationship of languages of the continuum to Woleaian. Chuukic languages tend to share grammatical characteristics but seem to vary in prosody and vocabulary (Ellis, 2012, in van den Berg, 2014: 9). According to Kennedy (2003: 35), "Micronesian stress does not clearly have a consistent manifestation of prominence like pitch or loudness". That is, stress is assigned and alternates based on the mora. The family presents variation in regards to primary stress, which correlates with the activity of the final vowels.

"At a certain time in history, Chuukic words must have all ended in vowels" (van den Berg, 2014: 13). Over time, these final vowels in many if not all Micronesian languages have been through changes, either as shortened unstressed vowels at the end of words, voiceless or disappeared (van den Berg, 2014: 13). In languages with full deletion patterns, like Pohnpeian (ISO 639-3 pon), Mokilese (ISO 639-3 mkj), and Marshallese (ISO 639-3 mah), primary stress occurs on the final surface vowel (Kennedy, 2003: 36).

For example, based on Regh (1993: 30), at the phrase level in Pohnpeian, primary stress occurs on the final mora, with secondary stress occurring on alternate preceding moras (Table

1). In Pohnpeian, long vowels are treated as sequences of double moras. “The unmarked pitch contour of a non-final prosodic phrase is 234 (mid, high, extra high); that of a final prosodic phrase is 231 (mid, high, low). High pitch occurs on the penultimate mora of the phrase” (Regh, 1993: 30).

Ponapean	Mokilese	Gloss
mwenge	mwinge	‘eat’
mesé	mijoa	‘face’ (3ps.)
àramás	armaj	‘person’
àperé	aproa	‘shoulder’ (3ps.)
menìpiníp	menipnip	‘thin’
isìpwukí	ijipwki	‘seven-hundred’
dìpwekèlekél	dipkelkel	‘stumble’
ìmwisèkalá	imwjekla	‘finished’

Table 1: A comparison between evidence from Ponapean and Mokilese (Regh, 1993: 29)

Regarding Ulithian (ISO 639-3 uli), Sohn and Bender (1973: 74) note that stress is non-phonemic and not clearly recognisable, while it accompanies a long vowel or a short vowel before a long consonant. But, if the short vowel before a long consonant is followed by a long vowel then it is not stressed. Sohn and Bender (1973: 37) have also identified four contrasting phonemic pitch levels: 1,2,3 and 4. The most common pattern in statements and interrogative-word questions is 231.

	PUA	WOL	PUL	PON	MRS	GIL	KSR
1)	S_0V́CV̄CV̲]	S_0V̄CV̲CV̲]	?	S_0V̄CV́C]	S_0V̄CV́C]	S_0V̄CV́C(V̲)]	S_0V̄CV̲C]
2)	S_0V́CV̄V̲]	S_0V̄CV̲]	S_0V̄CV́]	S_0V̄CV́]	S_0V̄CV́]	S_0V̄CV́(V̲)]	S_0V̄CV̲]

Figure 3: Surface representations of stress and pitch in Micronesian (Regh, 1993: 35)

In those languages with an incomplete vowel-weakening process (either as devoicing, or deletion of a subset of vowels), primary stress occurs on the penultimate mora (Kennedy, 2003: 36, Figure 3). Oda (1977, in Regh, 1993: 32)[1] notes that the intonation contour of a simple statement in Pulo Annian, a question-word question and a simple command is 231, with no pitch levels presented, and Regh (1993: 32) inferring that pitch level 3 is assigned on the penultimate mora. These conclusions are noted as hypothetical and if forms like the ones above occur in Pulo Annian, then, stress assignment is mora-sensitive while pitch assignment is syllable-structure sensitive (Regh, 1993: 32). That way, stress is assigned on the penultimate voiced mora, with voiceless moras being treated as extrametrical, and high pitch assigned to the penultimate syllable. Thus, stress and pitch might not co occur because of final vowel devoicing.

5. Methodology

5.1. Data

Covid-19 has affected many aspects of our life, including how we conduct research. Regarding research in linguistics, language archives can become useful tools for accessing a variety of data without conducting fieldwork. Originally funded by Arcadia in 2002, ELAR is a digital repository which became part of the School of African and Oriental Studies (SOAS) library in 2014 (*Endangered Languages Archive*). The main aim of the archive is to preserve and publish endangered language documentation materials from around the world (*Take-down policy*, 2018). SOAS staff and students, as well as Endangered Language Documentation Programme (ELDP) grantees are encouraged to deposit their material (*Endangered Languages Archive*). ELAR also allows external parties to deposit their material, while with the

[1] Oda's (1977) dissertation on Pulo Annian (Glottocode pulo1240) was not possible to retrieve and the data presented here is from Regh's (1993) paper on Proto-Micronesian prosody (Figure 4).

DELAMAN project, it participates in an international body of archives and initiatives, aiming at preserving the intangible culture of our world (*Endangered Languages Archive*).

The ELAR deposit "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language" is the outcome of an ELDP funded documentation project, curated by Peter Black and Barbara Black, who are part of the Friends of Tobi Island team, in 2013-14. The deposit's page (Figure 4) has a column on the left with five different search options based on the access protocol (**O**pen resources and ordinary **U**ser), Language (Tobian, English), Type (Document, Video, ELAN, Settings, Audio, Image), Genre (Story, Conversation, Meeting, Discussion, etc.) and finally Participants (Peter W Black, Barbara W Black, Felicisma Andrew (Ngiralbong), Regina Andrew, etc.).

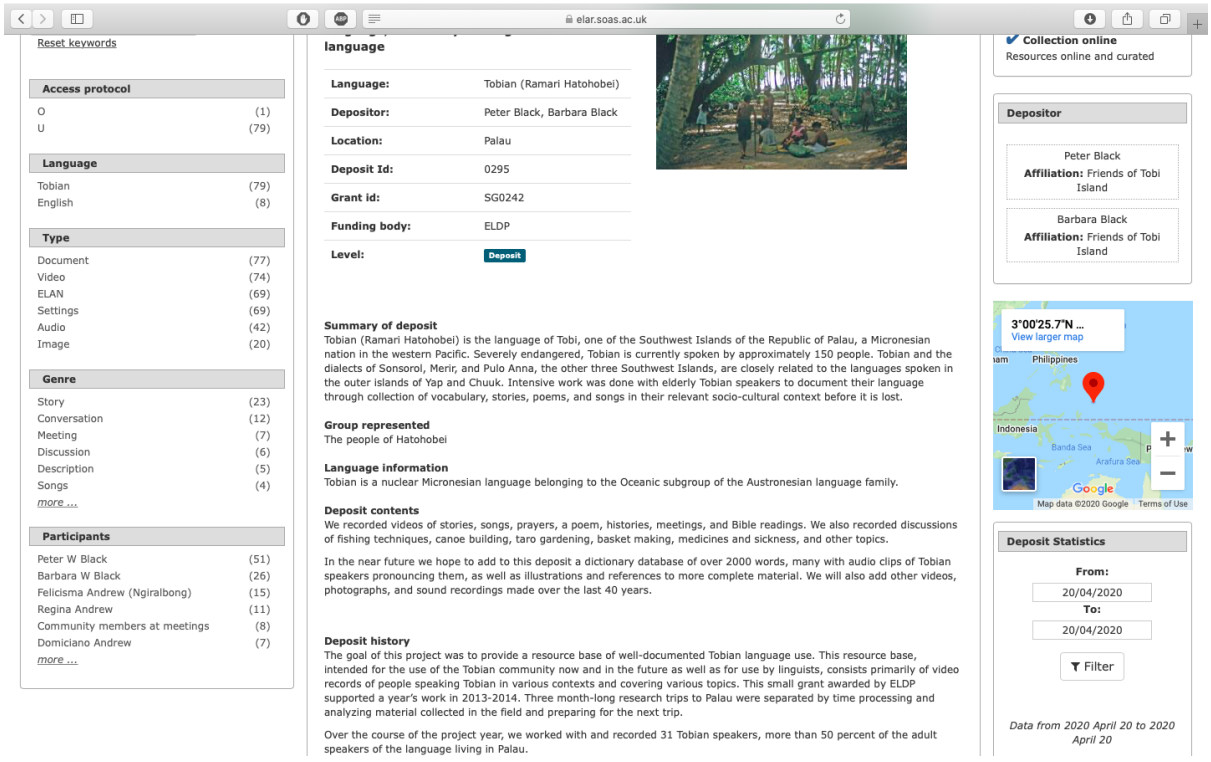


Figure 4. The deposit page

The deposit is divided into the main deposit page and the bundles and resources one. The main deposit page (Figure 4) contains general information, such as the name of the language, the depositors, the location, the deposit's id, the ELDP grant's id. There is also a summary of the deposit, the language group represented and information about the language,

the deposit's contents and history and finally acknowledgements and citation. A map is also included.

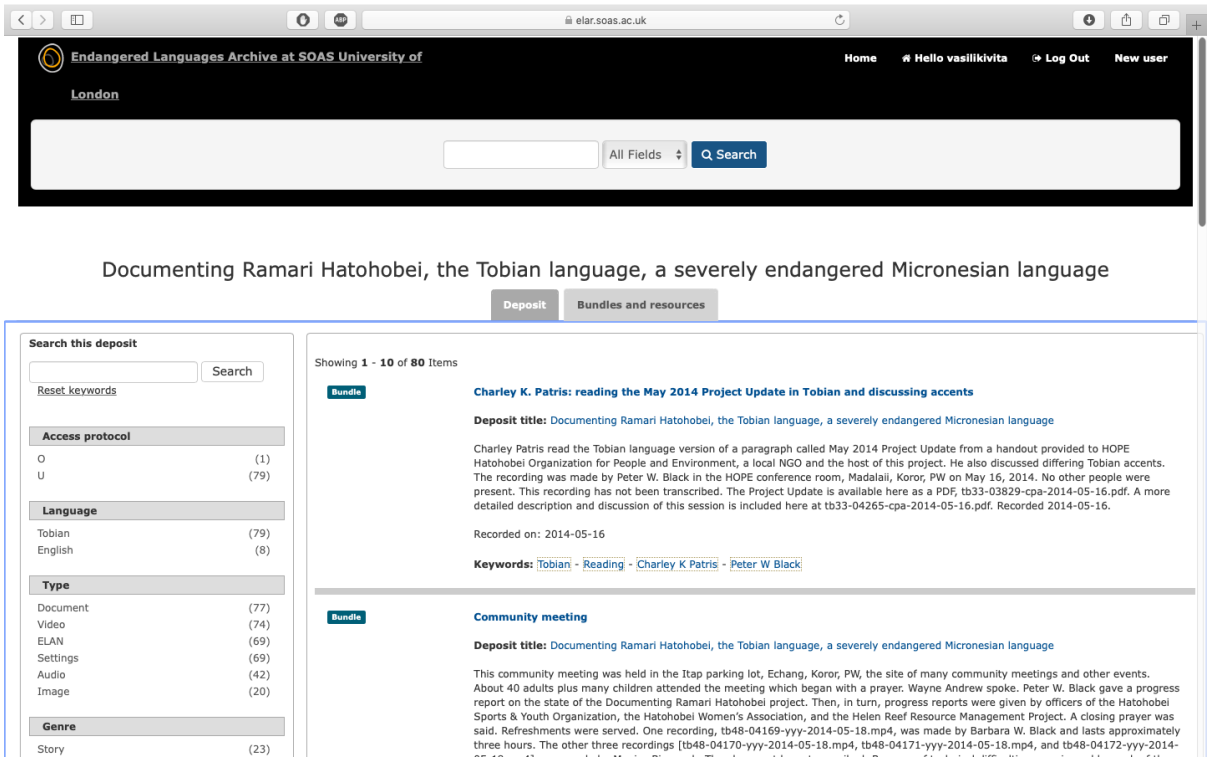


Figure 5. The bundles and resources page

As far as the bundles and resources page (Figure 5) is concerned, the title of each bundle is provided, as well as, a summary of what is included in the recording, the date of the recording and some keywords that could be of help when searching. When clicking on a bundle, more detailed information about the bundle is provided (Figure 6), such as, the genre, the location, the participants, and the actual files, the access protocol, the name, the type (ELAN, settings, document, video etc.) and the resource.

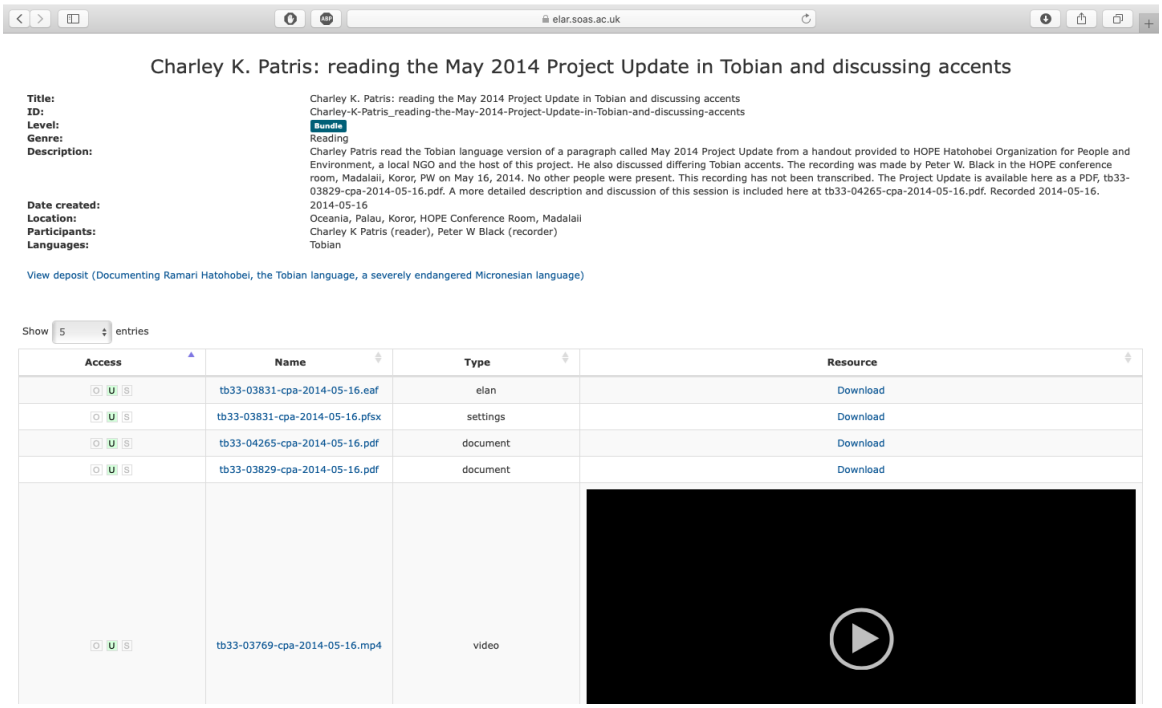


Figure 6. Bundle tb33-03831-cpa-2014-05-16

For the purposes of this project, 5 bundles, in total 1 hour of recordings, from the collection using ELAN (ELAN, 2020) and the IPA (*IPA Chart,* 2015) have been transcribed and glossed. The particular bundles contain ethnographically informed recordings. Two of the bundles (Figure 7 and 8) are narrations of the same story during different times, 2004 and 2014, while the other three consist of a monologue on how Nixon Andrew learnt how to fish (Figure 9) and two conversations, one on the topic of jobs (Figure 10) and another one on organising a community event (Figure 11). There are 10 different speakers, varying in terms of age, from 20 to 70 years old, gender and sociolinguistic background (education, languages, place of residence).

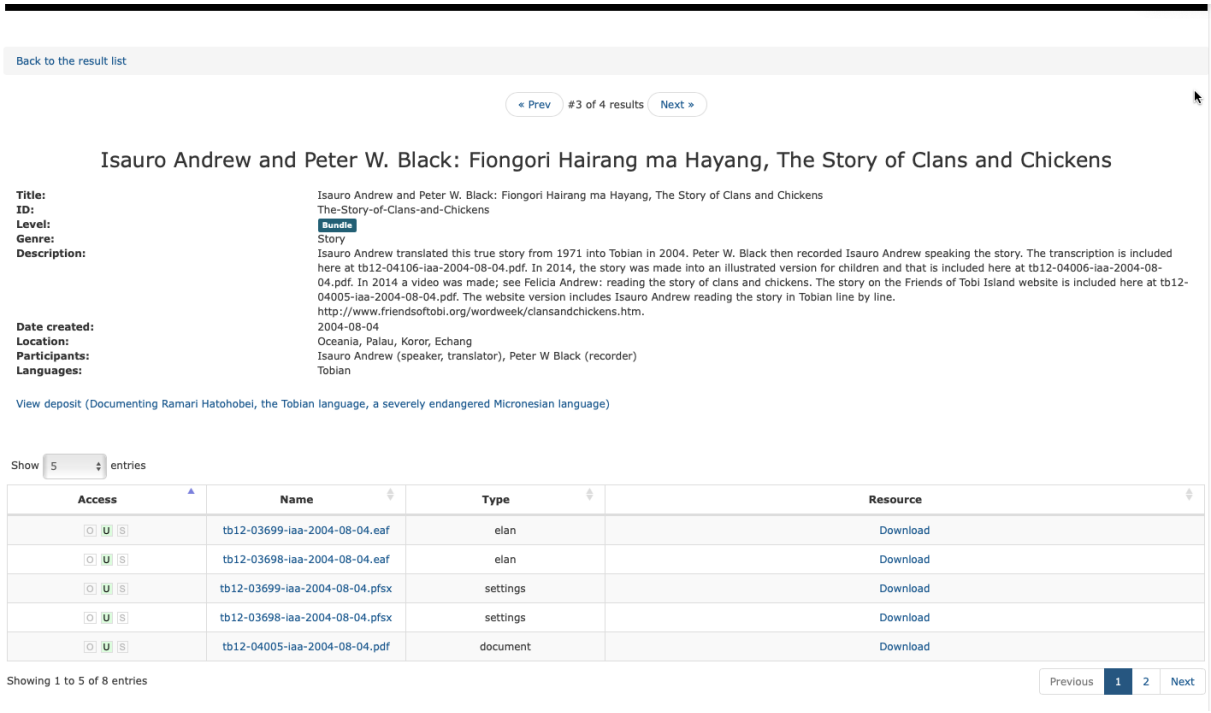


Figure 7. Bundle tb12-03698-iaa-2004-08-04

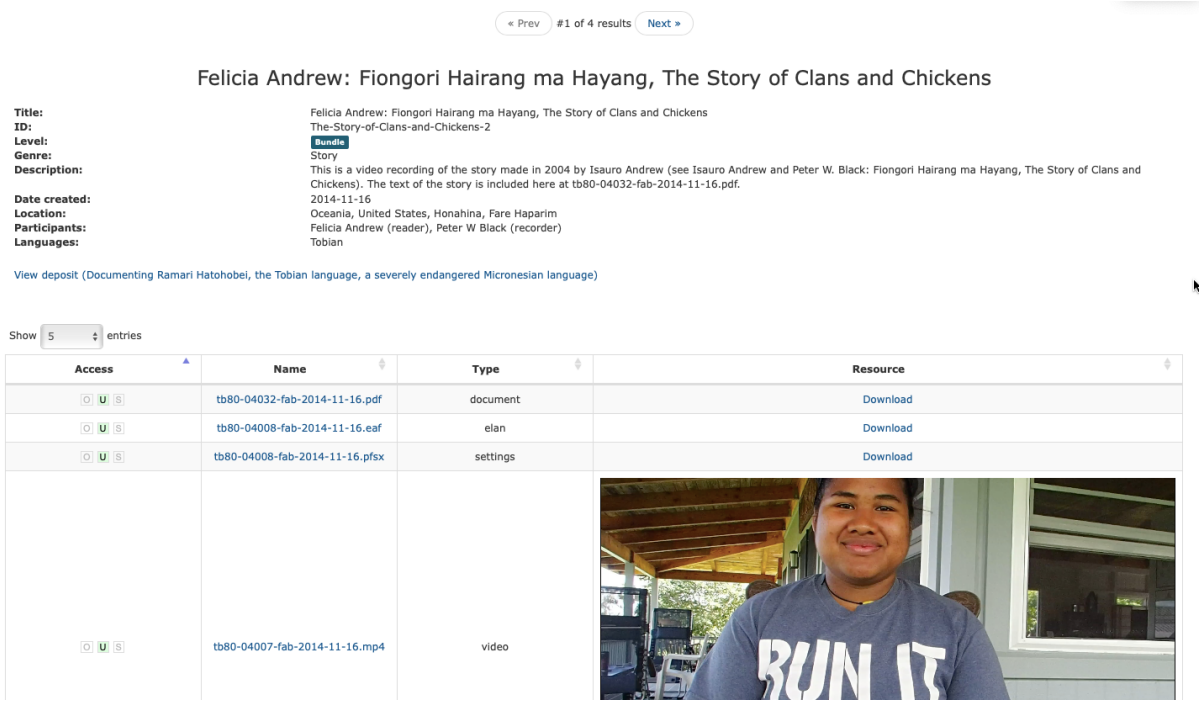


Figure 8. Bundle tb80-04007-fab-2014-11-16

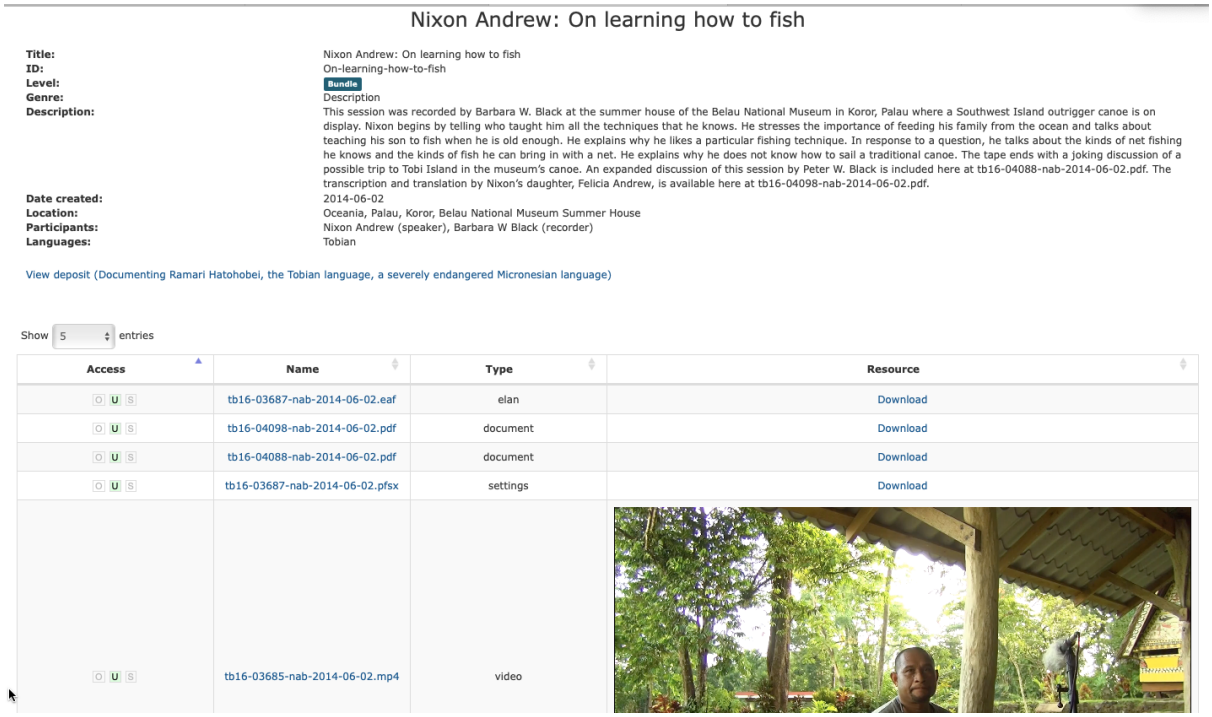


Figure 9. Bundle tb16-04088-nab-2014-06-02

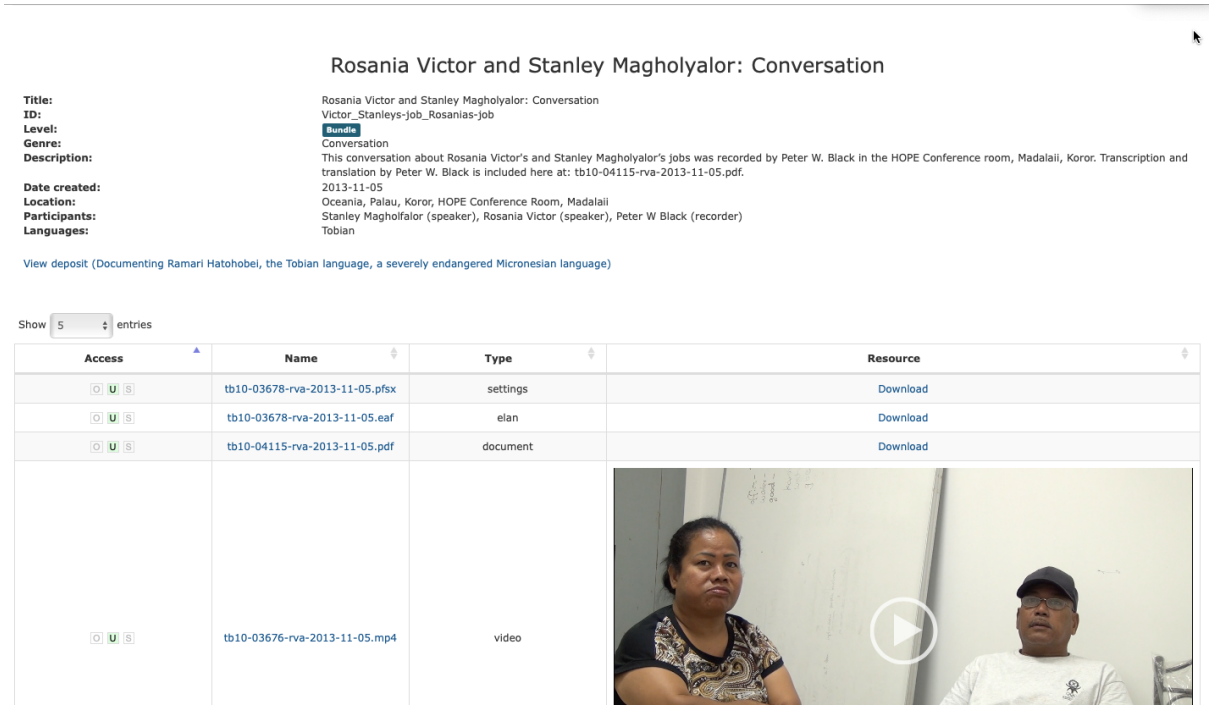


Figure 10. Bundle tb10-04115-rva-2013-11-05

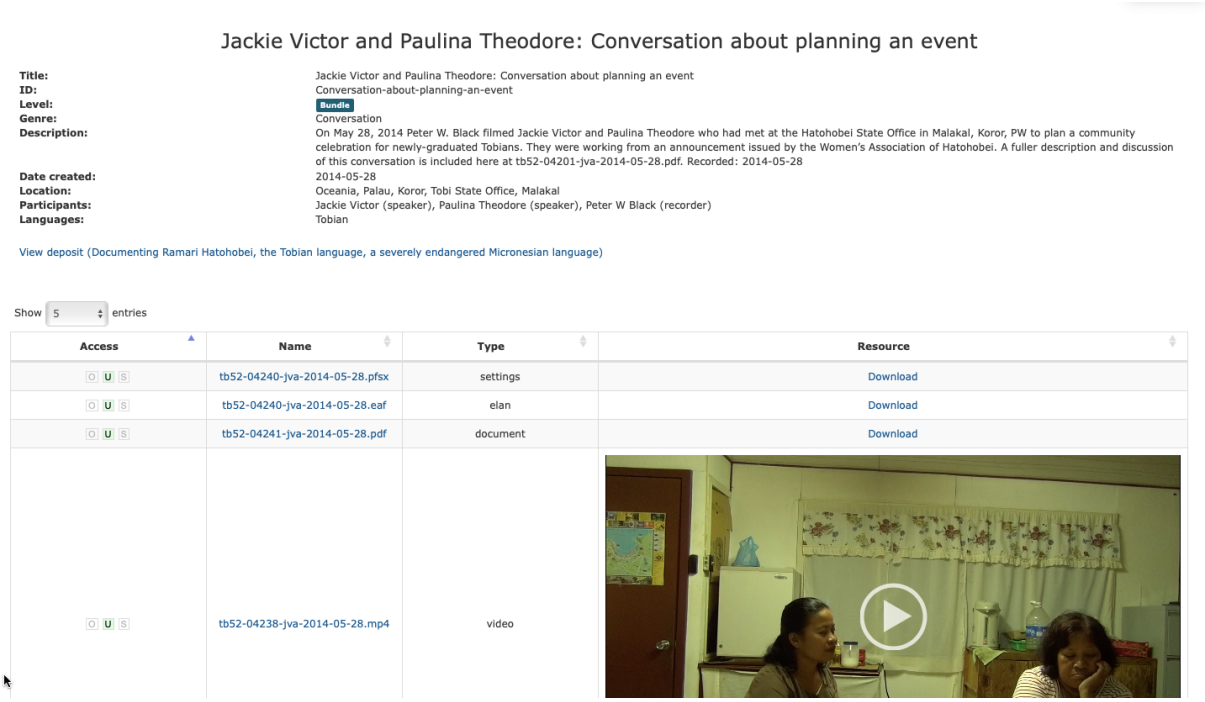


Figure 11. Bundle tb52-04241-jva-2014-05-28

The curators have assisted me during my research project in navigating the collection faster and have always been helpful regarding questions related to cultural information. Through email and face-time correspondence due to not only physical distance but also because of Covid-19, they have assisted with cultural insights and their experiences and relationship with the speakers and the community in general. They have also introduced their online dictionary which became a valuable resource of ethnographic information.

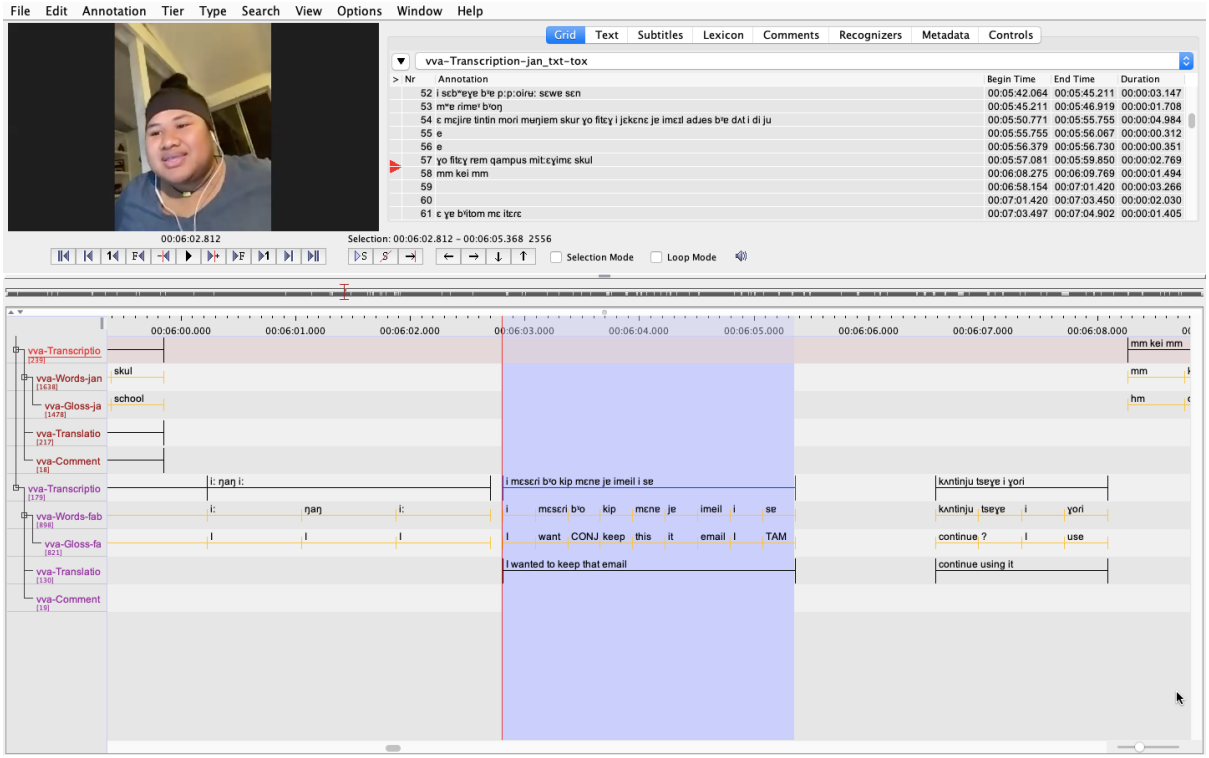


Figure 12. Bundle tb03-vva-jan-fab-2020-07-03

Sessions[2] with two speakers, Tintin and Justin Andrew, have been conducted, thanks to the curators who introduced me to them. Before the recorded sessions, speakers were given a consent form based on the SOAS Research Ethics Policy (*Research Ethics at SOAS*) (Appendix 10.1). The speakers assisted with translating certain recordings, clarified syntactic aspects of the language and had a conversation on Zoom (Zoom Video Communications Inc, 2016), talking about Covid-19 and their news since it had been a long time since they saw each other (Figure 12). The conversation was organised with the purpose of acquiring data on questions in naturally occurring language settings.

Interviews with the speakers have also been conducted with the purpose of discovering their linguistic background, their attitudes, and thoughts about the languages they speak. Zoom (Zoom Video Communications Inc, 2016) was used for its possibility of sharing and recording

[2]The particular bundles are to be uploaded to the collection with different names, with a note in their descriptions for the names used in this dissertation. For now, the recordings can be found on the following link: https://drive.google.com/drive/folders/1uTfX8V41jwWdxfSYrgDvDaPWTFVgqEu1?usp=sharing.

one's screen. Some limitations include the fact that the recording was not able to capture all participants, rather only the individual speaking every time. Furthermore, the quality depended on the participant's device and the connection was sometimes lost. The interviews (Figure 13 and 14) were conducted as in a Master-Apprentice session of learning basic questions in Tobian, such as "what is your name", "how old are you", "where are you from" etc, for the purpose of making the consultants feel comfortable. Tintin pointed out that she would like to conduct our sessions in such a manner and that it was more fun compared to just translating content.

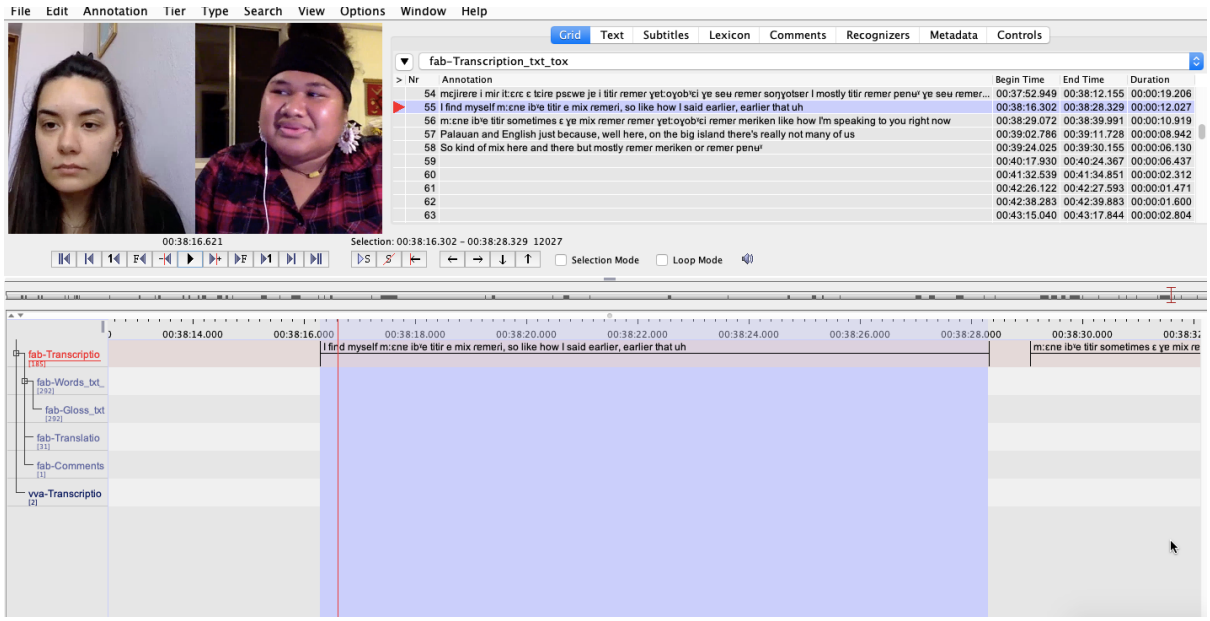


Figure 13. Bundle tb01-vva-fab-inter-2020-06-20

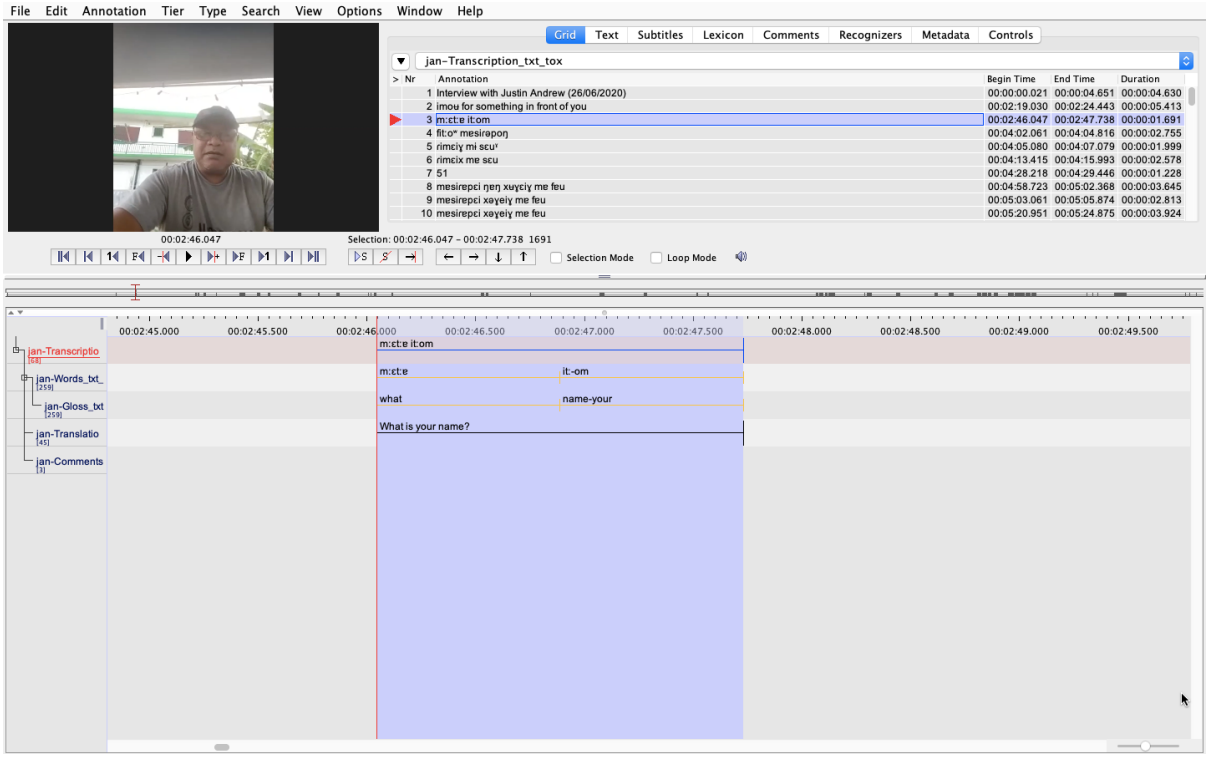


Figure 14. Bundle tb02-vva-jan-inter-2020-06-26

5.2. Analysis

As far as the analysis of the recordings is concerned, first of all, ELAN (ELAN, 2020) was used for the transcription and annotation of the recordings since the curators are familiar with the particular software. The tiers used are an IPA (*IPA Chart*, 2015) tier, a Word tier where the Leipzig Glossing Rules (Comrie et al., 2008) were used for the annotation, while Capell's (1969) and van den Berg's (2014) phonetic analyses where taken into account. Translation and comments tiers were also included. Afterwards data had been labelled based on the Switchboard DAMSL (Dialogue Act Markup in Several Layers) tags[3] (Jurafsky et al., 1997) (Table 2) and on criteria such as sound quality and length and extracted using Praat (Boersma and Weenink, 2020).

Tag	Example
Statement	Me, Im in the legal department.

[3] Analysis of only Statements, Opinion, Yes-No-Question and Wh-Questions are only presented due to the word-limit.

Backchannel	Uh-huh.
Opinion	I think its great
Agree/Accept	Thats exactly it.
Abandoned/Turn-Exit	So, -/
Appreciation	I can imagine.
Yes-No-Question	Do you have to have any special training?
Non-verbal	<Laughter>, <Throat Clearing>
Yes answers	Yes.
Conventional-closing	Well, its been nice talking to you.
Uninterpretable	But, uh, yeah
Wh-Question	Well, how old are you?
No answers	No.
Response Ack	Oh, okay.
Hedge	I dont know if I'm making any sense or not.
Declarative Question	So you can afford to get a house?
Other	Well give me a break, you know.
Backchannel-Question	Is that right?

Table 2. Most frequent SWBD-DAMSL labels, from Jurafsky et al. (1997)

Regarding the description of the data, ProsodyPro (Xu, 2013) was used for the extraction of accurate f0s and the ToBI (Silverman et al., 1992) conventions for the description of the patterns observed. R (R Core Team, 2013, Appendix 10.2) was used for the statistical representation of moraic feet in Tobian (Figure 18). Prosody Pro's (Xu, 2013) values were used for the analysis. That is, the F0 range is from 75-600 Hz with the maximum formant for male

speakers being 5000 Hz and for female speakers 5500 Hz (Figure 15). Intensity[4] is measured in dB and pitch in Hz (Figure 15). The horizontal axis shows the time and the vertical axis shows intensity and pitch (Figure 16). A representation of both in the same Figure (16) has been opted for all examples for the purposes of exhibiting that Tobian stress (intensity, dB) and highest pitch (Hz) do not coincide. The bold, or sometimes dotted, line in the example (Figure 16) represents pitch (Hz), while the regular mountainous line represents intensity (dB).

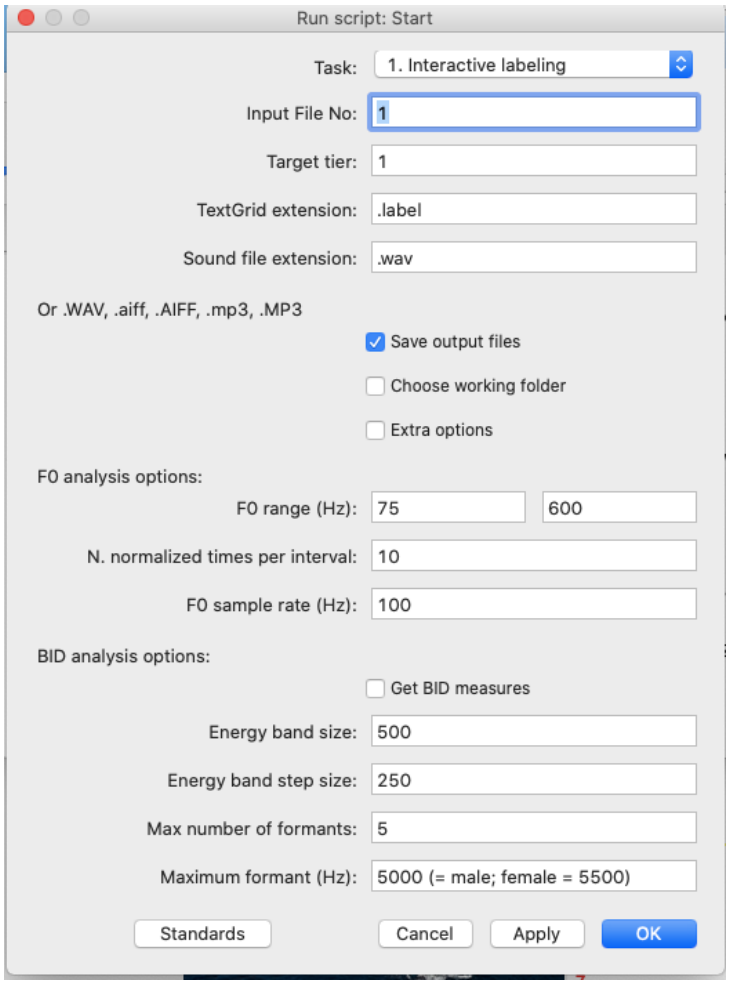


Figure 15. Prosody Pro (Xu, 2013) script

[4] Values for Intensity (dB) and Pitch (Hz) of all data used for the purposes of this dissertation can be found here: https://docs.google.com/spreadsheets/d/1XYiYJJFsEh353-U9-h8FlG_5Ch0-cwyuIHwTKgXlOoU/edit?usp=sharing.

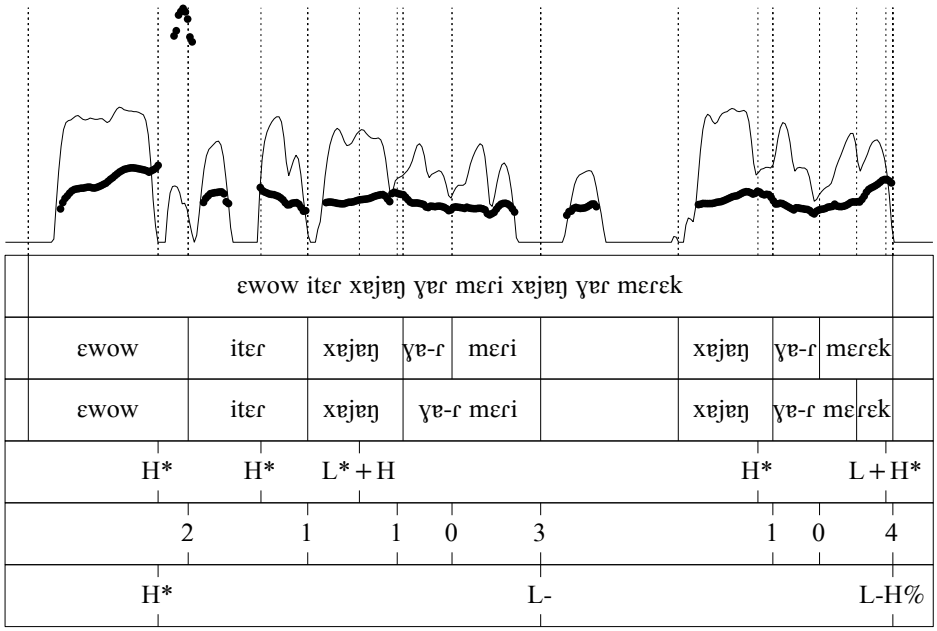


Figure 16. Example of representation

The tiers for each TextGrid are the IPA (IPA Chart, 2015) phrase, word and feet transcription and division tiers, the tone tier, describing "distinctive pitch events, transcribed as a sequence of high (H) and low (L) tones" (Silverman et al., 1992), the break index tier "which marks the prosodic grouping of the words in an utterance by labelling the end of each word for the subjective strength of its association with the next word, on a scale from 0 (for the strongest perceived conjoining) to 4 (for the most disjoint)" (Price et al, 1991), and the phrase tones tier, following the pitch contour of phrases (Figure 17).

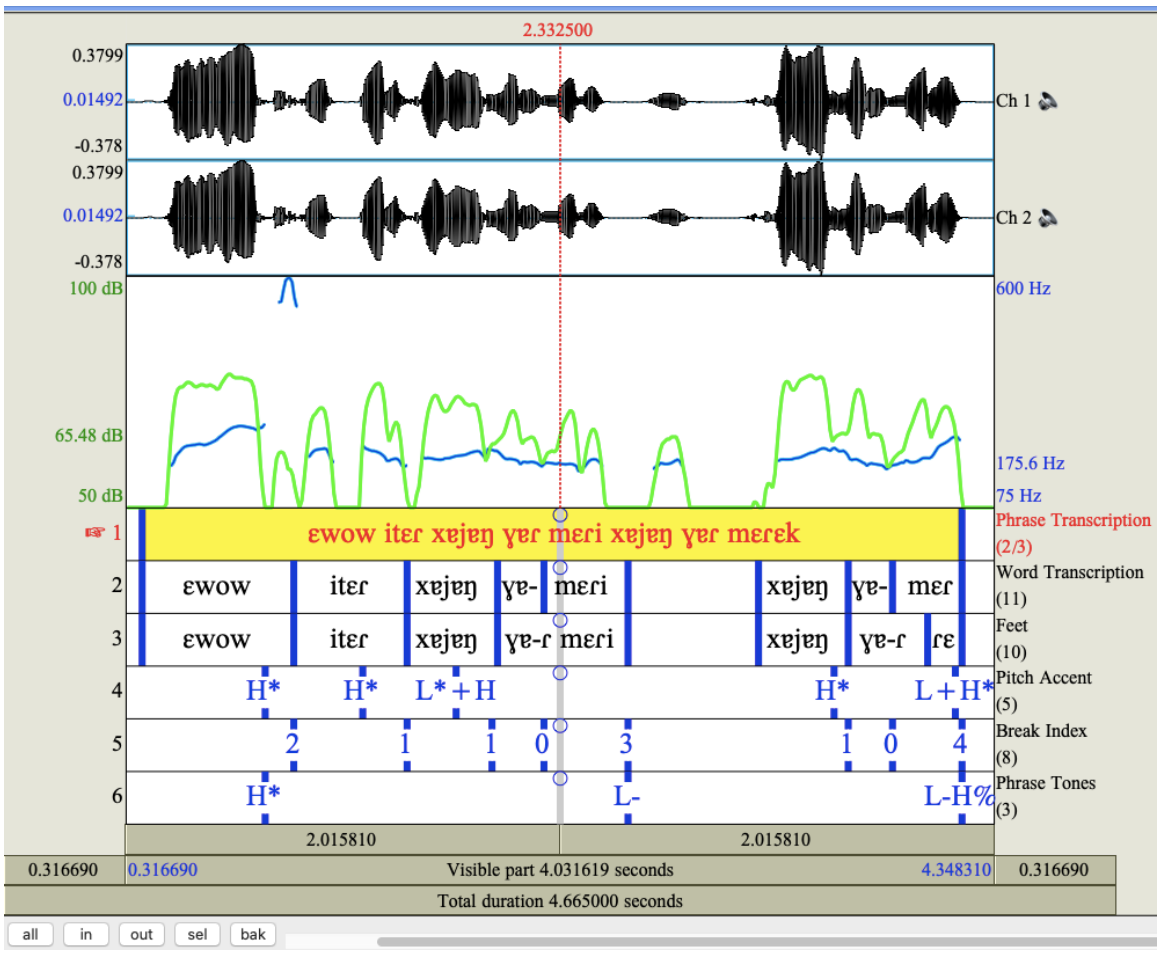


Figure 17. Example of analysis

6. Results

6.1. Phonotactics

According to van den Berg (2014: 14), Tobian seems to allow a variety of syllables, primarily due to the fact that it has dropped its final vowels. All double consonants in van den Berg's (2014: 14) table (Table 3) are geminate consonants and clusters of two different consonants occur only in borrowed words like *skuur*, school, *ɹeɪndʒəɹ*, ranger and *steit*, state. Codas in isolated words are often being 'reunited' with the final vowel.

	SpnCRL	HAT	
V:	i	i	sbj:1.sg
VV:	ii		
CV:	ma	ma	coord.conn

CVV:	maa	ngii	‘tooth’
CCV:	mmasa	cca	‘blood’
CVC:	mat	yehamat	‘person’
CVVC:	maat	yang	‘wind’
CCVC:	mmat	ppōr	‘dirt, soil’
CVCC:	makk	rapp	‘big’
VC:	aw		
VVC:	aaw	iih	‘fish’
VCC:	aff		

Table 3. Syllables for Tobian based on a comparison from Ellis (2012: 133) in van den Berg (2014: 13)

For example, in reduplication, Tobian seems to be suffixing (Table 4). According to Kennedy (2018: 300), in reduplication, “some or all segments of a root are repeated, and the appearance of the repeated sequence carries some morphological function”. In Tobian, reduplication is used to express owning or occupying the object referred to (Capell, 1969: 54) or to express a continued past or present (Capell, 1969: 39).

VC		
Syllabify	sɛi.ɾoɣ (fish)	
Reduplicate	sɛi.ɾoɣ	sɛiɾoɣ
Associate	sɛi.ɾoɣ	sɛi.**ɾoɣ**
Stray erasure	sɛiɾɐɣ	ɐɣ
VCVCV		
Syllabify	ɛ.soɣ(ʊ) (he-say)	
Reduplicate	ɛ.soɣ(ʊ)	ɛ.soɣʊ
Associate	ɛ.soɣ(ʊ)	ɛ.soɣʊ

Assimilation	ɛ.soɣ	ʊ	**tsʊ.ɣʊ**

Table 4. Tobian reduplication

The final vowel lenition processes of Micronesian languages are observed (Regh, 1993: 25). Because of lenition, the final vowels are dropped, but return in inflectional processes such as reduplication (van den Berg, 2014: 14). In other words, Tobian allows codas in word-final position, but then in reduplication the hidden vowels of Proto-Micronesian return (Table 5).

VC: *ʉx-ʊx*, net-RDP, *sɛiɾɐɣ-ɐɣ*, fish-RDP

CV: *bʸʊ-bʸʊ*, go-RDP

VCVCV: *ɛ-soɣ-ʊtsʊɣʊ*, he-say-RDP

VCVC: *ɛ-ɣots-œɣots*, it-black-RDP

CVC: *bʸoɣ-bʸoɣ*, go-in-RDP, *bʸoŋ-i-bʸoŋ*, night-RDP

Table 5. Types of Tobian reduplication

As van den Berg (2014: 14) mentions vowel initial syllables exist in Tobian (e.g. [u:x], net). However, there are two rules which constraint them to a VVC type:

A consonant coda in a CVC or VC syllable will be re-segmented as the onset of a new second syllable in running speech

(2) *sɐntos,* Santos -> *sɐntos**ə** bʸɐ,* Santos that (tb12-03698-iaa-2004-08-04-CH6-06)

A type of compensation rule that lengthens vowels in words that are monosyllabic when elicited on their own

(3) *ŋɐ i ɣʉrɐ foɣʉr ɣoɣor ux,* CONJ I know do use net (tb16-03685-nab-2014-06-02-CH6-07)

(4) *u:x,* net (Friends of Tobi, 2020)

Hughes (2020: v) has observed that Micronesian languages, Tobian as well, are weight-sensitive, which means that heavy syllables (diphthongs, double vowels) and in particular moras within those syllables are assigned stress. Moras can be assigned to sonorant consonants

and, of course, vowels. There is uncertainty as to whether geminate consonants are assigned moras in Tobian. It is hypothesized that only the sonorant geminate [m:] is assigned a mora. Diphthongs and double vowels are assigned two moras.

1 sg	i
2 sg	ho
3 sg	ye
1 pl incl	si
1 pl excl	hei
2 pl	hau
3 pl	ha/he

Table 6. Ovierview of subject markers in van den Berg (2014: 35-36)

Similar to Gilbertese all rime-internal elements are moraic, while onsets are nonmoraic (Blevins and Harrison, 1999: 209). There are restrictions as to which codas are assigned moras which will be explained in the next section (6.2.1.). The only monomoraic lexical items are minor categories like (proclitic) emphatic subject markers, subject markers (Table 6), modal-aspectual particles and prepositions. However, these items never occur in isolation and are followed by a major category item. Thus, a content word must at least be bimoraic.

(5) *i=bʸu* I go

Regh (1993: 34) argues that Proto-Micronesian most probably "distinguished between minimal 'phonological words' and minimal 'phonological phrases'. A minimal 'phonological word' consisted of two morae, forming a bimoraic foot". This dissertation argues that in Tobian as well, the minimal lexical and prosodic word is bimoraic.

(6) a. *fɐɣɯɾ,* do

b. *i miɾ,* I stay/live

c. *ɛ=moɣ,* it=good

d. *ɣʉɾɐ,* know

e. *ɛ=moɾ,* it=already

The majority of feet in Tobian, as in Proto-Micronesian, since it has been argued that Tobian has preserved phonological constituents of Proto-Micronesian (Grant, 2017: 853), are trimoraic (Regh, 1993: 25), with quadrimoraic feet emerging as well.

(7) a. *jɐɣɐmɐt,* person -> *jɐ/ɣɐ/mɐt*

b. fioŋo-ri, story-of -> *fi/o/ŋo/ri*

Figure 18 statistically presents the amount of words in the data that are monomoraic, bimoraic, trimoraic and quadrimoraic. The x axis shows the number of moras assigned to each prosodic word in the data and the y axis presents the amount of prosodic words containing each number of moras (Figure 18). It is evident that the amount of trimoraic prosodic words in the data outnumbers the bimoraic.

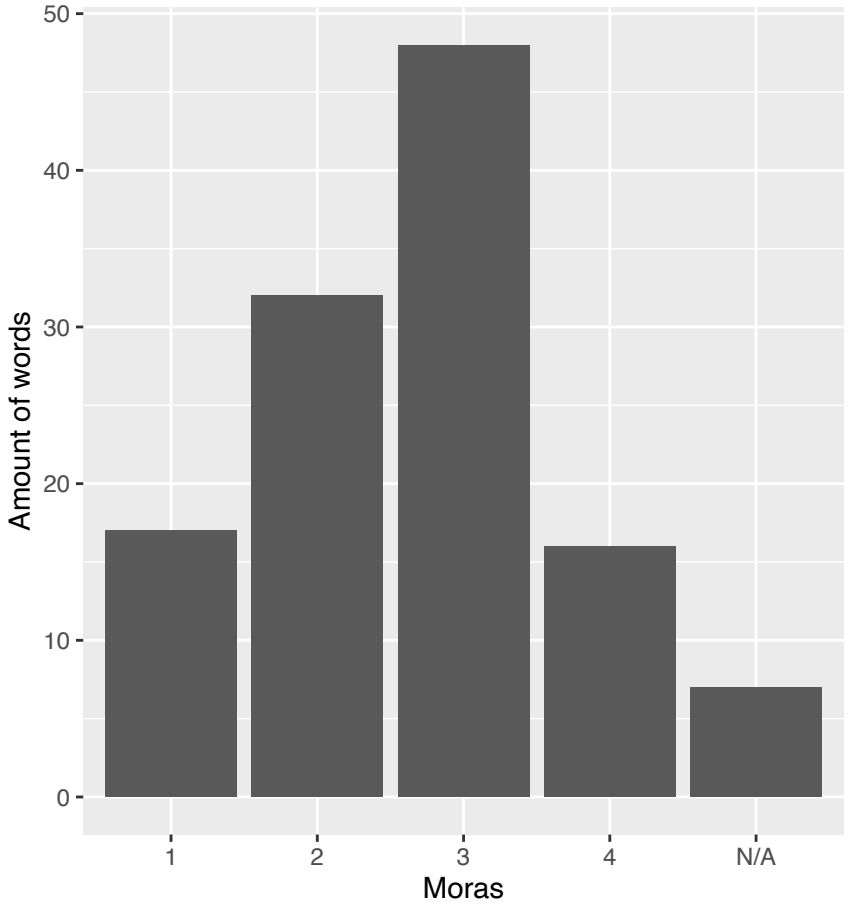


Figure 18. Feet in Tobian

6.2. Prosody

6.2.1. Stress

According to Kennedy (2003: 35), Micronesian stress is not manifested clearly in terms of pitch or intensity. It is argued that primary stress is assigned on the right edges of words, with secondary stress alternating towards the left (Table 1). In particular, Regh (1993: 29) argues that high pitch occurs on the second to last voiced vowel and low pitch on the last voiced vowel. Another characteristic of Micronesian stress is the fact that on the prosodic level, it is assigned and alternates depending on the mora (Kennedy, 2003: 35 and Regh, 1993: 29).

A bimoraic syllable must not be skipped, that is, "if the first stress occurs on the mora immediately after a heavy syllable, the first mora of that syllable must also bear stress" (Kennedy, 2003: 35). Thus, primary stress is usually assigned on the penultimate mora, which receives a salient pitch drop, which characterizes languages of the Micronesian family (Regh, 1993: 29). Nevertheless, there is variation in the family as to the locus of primary stress, which correlates with the activity of final vowel weakening (Kennedy, 2003: 35).

Tobian is weight-sensitive and seems to be following the hypothesis about the Pulo Annian pattern (Regh, 1993: 32), with stress assignment (intensity) being mora sensitive. Primary stress is assigned on the penultimate (from left to right) mora of the prosodic word. For example, in *fioŋo-ri*, story-of (Figure 19 and Table 7), a quadrimoraic prosodic word, *fi/o/ŋo/ri*, it is the penultimate mora ŋ[o]=73.32 dB that presents the highest intensity, while secondary stress is assigned on the preceding mora, fi[o]=72.82 dB.

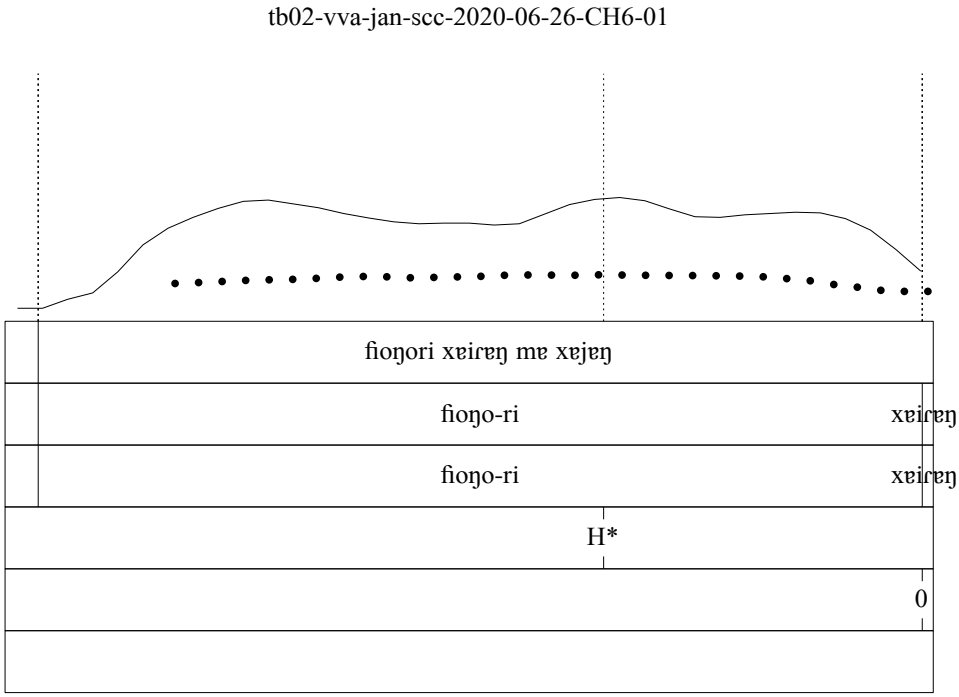


Figure 19. Spectrogram: [fioŋo-ri], story-of (tb02-vva-jan-scc-2020-06-26-CH6-01)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
fioŋo-ri, story-of	ŋ[o]=73.32 dB	fi[o]=72.82 dB	[ŋo]=173.5 Hz	H*

Table 7. [fioŋo-ri], story-of (tb02-vva-jan-scc-2020-06-26-CH6-01)

As mentioned before, all sonorant consonants can be assigned a mora. However, there are certain restrictions.

- In disyllabic words ending with sonorant codas if the penultimate syllable is heavy, (usually a diphthong) it is the second mora of that syllable that is stressed instead of the penultimate mora of the prosodic word (Figures 20-22 and Tables 8-10).

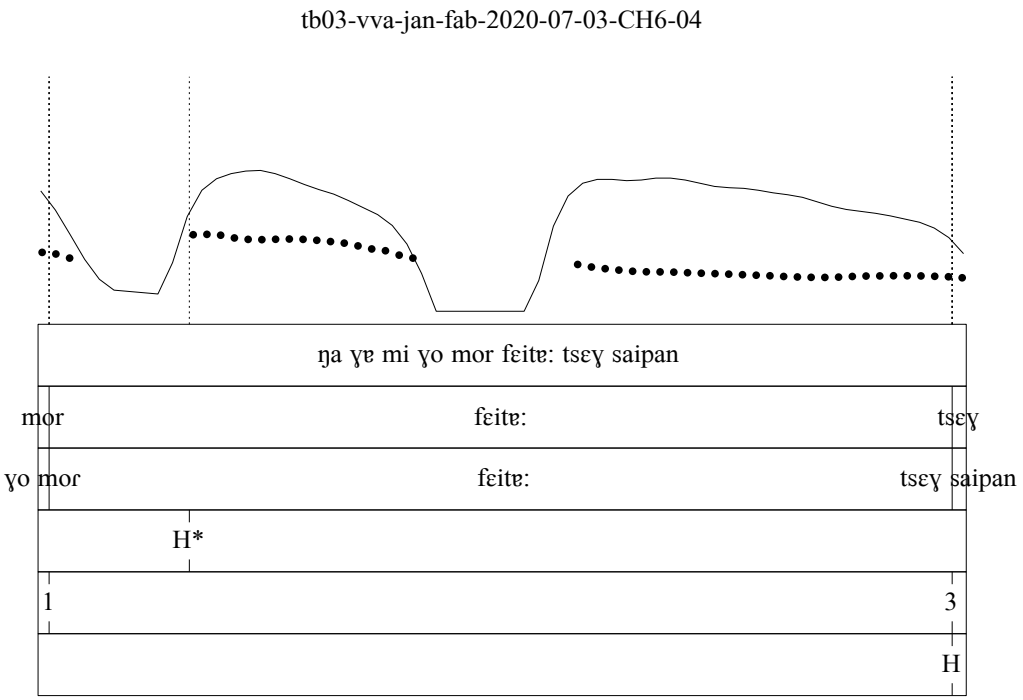


Figure 20. Spectrogram: [fɛitɐ], doing (tb03-vva-jan-fab-2020-07-03-CH6-04)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
fɛitɐ, doing	[i]=79.35 dB	t[ɐ]=78.03 dB	[fɛi]=265.5 Hz	H*

Table 8. [fɛitɐ], doing (tb03-vva-jan-fab-2020-07-03-CH6-04)

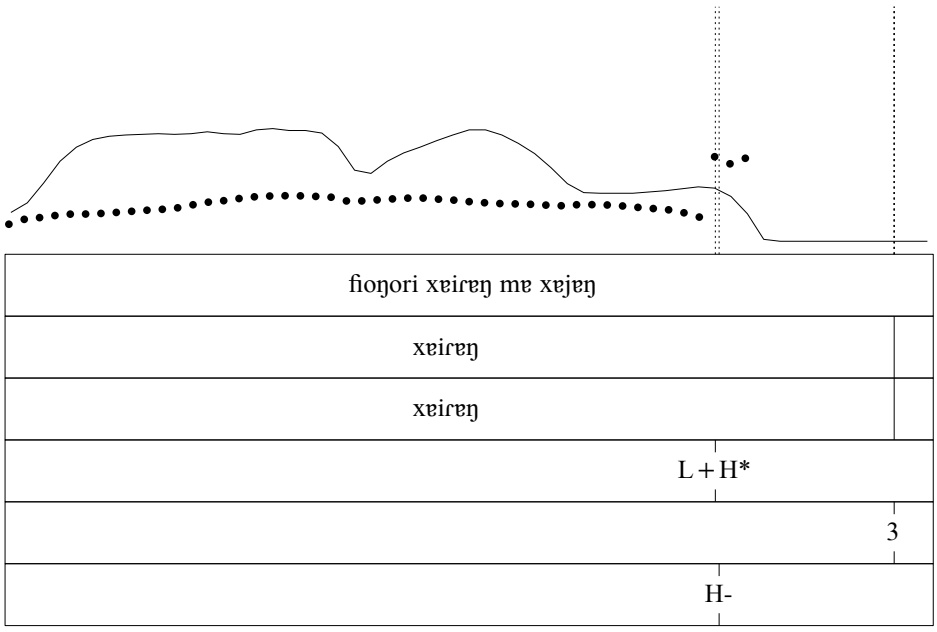


Figure 21. Spectrogram: [xɐirɐŋ], clan (tb02-vva-jan-scc-2020-06-26-CH6-01)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
xɐirɐŋ, clan	[i]=70.82 dB	r[ɐ]=73.66 dB (next mora)	[rɐŋ]=193.8 Hz	H*

Table 9. [xɐirɐŋ], clan (tb02-vva-jan-scc-2020-06-26-CH6-01)

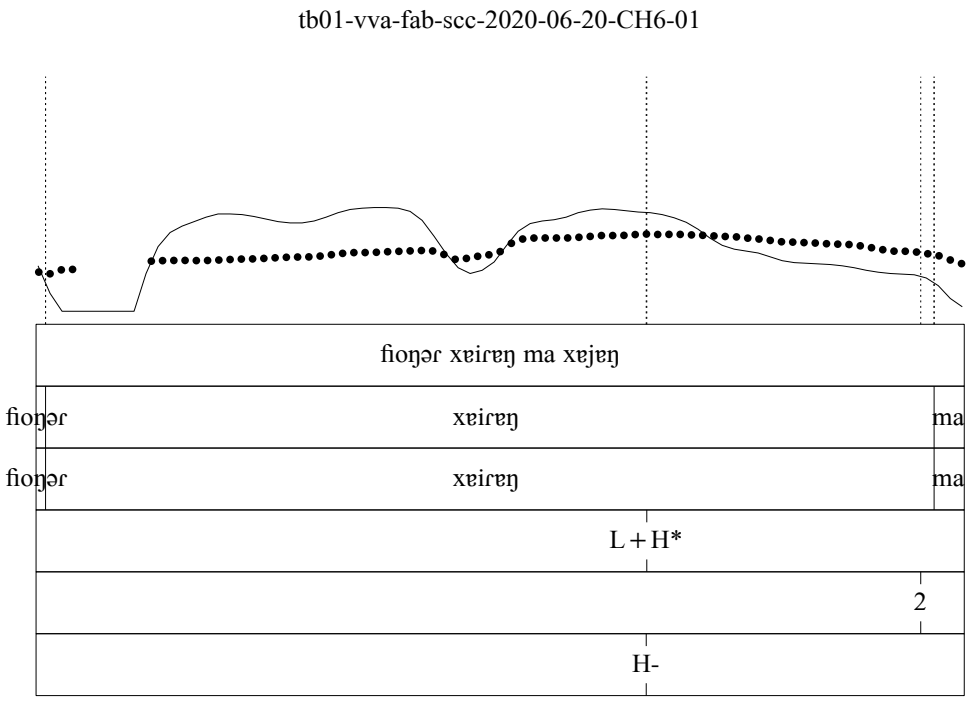


Figure 22. Spectrogram: [xɐiɾɐŋ], clan (tb01-vva-fab-scc-2020-06-20-CH6-01)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
xɐiɾɐŋ, clan	xɐ[i]= 72.02 dB	x[ɐ]i=70.72 dB	[ɾɐŋ]= 266.1 Hz	H*

Table 10. [xɐiɾɐŋ], clan (tb01-vva-fab-scc-2020-06-20-CH6-01)

- If the disyllabic word is at the end of the utterance even if the penultimate syllable is not a heavy syllable it is still that penultimate mora that is stressed (Figures 23 and 24 and Tables 11 and 12).

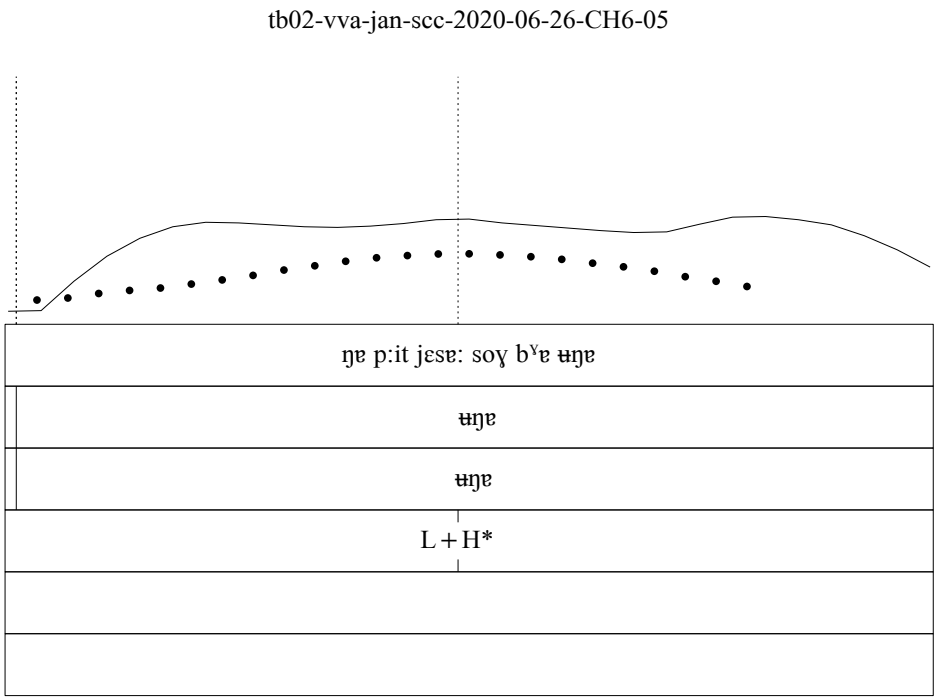


Figure 23. Spectrogram: [ʉŋɐ], yes (tb02-vva-jan-scc-2020-06-26-CH6-05)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ʉŋɐ, yes	[ʉ]=69.58 dB		[ʉ]=224.3 Hz	H*

Table 11. [ʉŋɐ], yes (tb02-vva-jan-scc-2020-06-26-CH6-05)

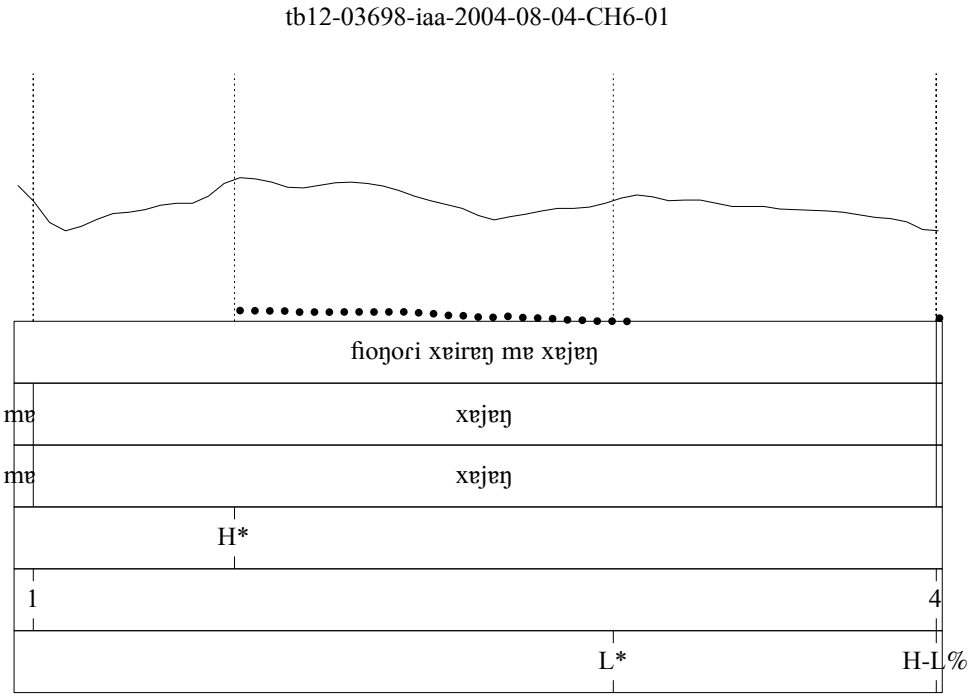


Figure 24. Spectrogram: [xɐjɐŋ], chicken (tb12-03698-iaa-2004-08-04-CH6-01)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
xɐjɐŋ, chicken	x[ɐ]=77 dB		[xɐ]=97.73 Hz	L*

Table 12. [xɐjɐŋ], chicken (tb12-03698-iaa-2004-08-04-CH6-01)

- If the disyllabic word is not at the end of the utterance and its penultimate syllable is a light syllable, then it is the penultimate mora, usually a mid-vowel, [ɐ], [ɛ], [o], [ə], that is assigned the stress and the sonorant coda becomes the final mora of the prosodic word (Figures 25 and 26 and Tables 13 and 14).

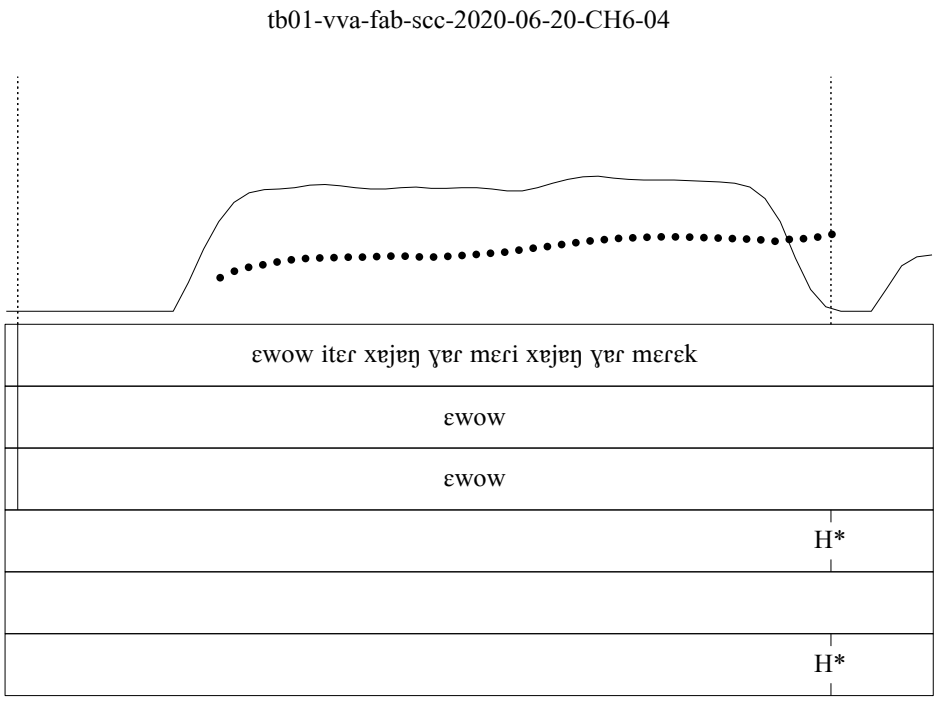


Figure 25. Spectrogram: [ɛ-wow], it-DEM (tb01-vva-fab-scc-2020-06-20-CH6-04)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ɛ-wow, it-DEM	w[o]w=78.72 dB	[ɛ]=76.93 dB	[wo]w=265.3 Hz	H*

Table 13. [ɛ-wow], it-DEM (tb01-vva-fab-scc-2020-06-20-CH6-04)

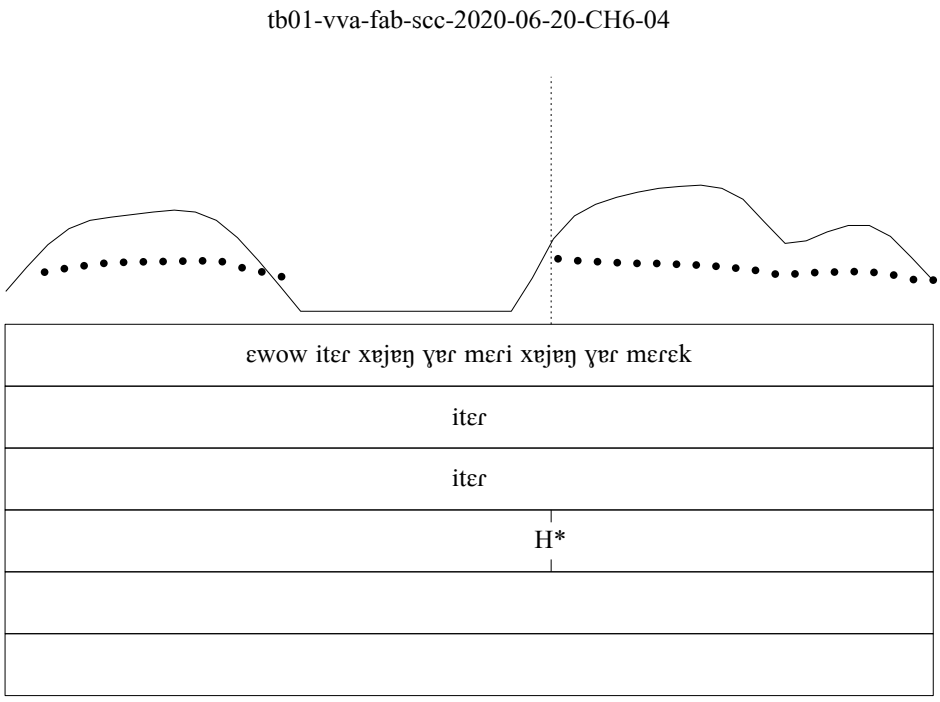


Figure 26. Spectrogram: [itɛr], name (tb01-vva-fab-scc-2020-06-20-CH6-04)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
itɛr, name	[ɛ]=76.67 dB	[i]=71.33 dB	[tɛ]ɾ=212.2. Hz	H

Table 14. [itɛr], name (tb01-vva-fab-scc-2020-06-20-CH6-04)

Pitch assignment, on the other hand is syllable sensitive, with the highest pitch being exhibited on the penultimate syllable in three syllable words[5] (Figure 27 and Table 15). However, in disyllabic words it is exhibited on the final syllable (Figures 21, 22, 25 and 26 and Tables 9, 10, 13 and 14). If these disyllabic prosodic words are at the end of the utterance, then highest pitch is observed on the penultimate syllable (Figure 24 and 25 and Tables 11 and 12). In Figure 20 and Table 8, for example, there is a slight hesitation that lengthens the vowel, thus

[5] Find all data, along with audio and TextGrid files here: https://docs.google.com/spreadsheets/d/14BJJSTqIekDH1F1L4YupZCoMoq_DtV-ZuXVT9krZzAc/edit?usp=sharing

it could be considered the last prosodic word of the utterance, thus the penultimate syllable exhibits high intensity and pitch.

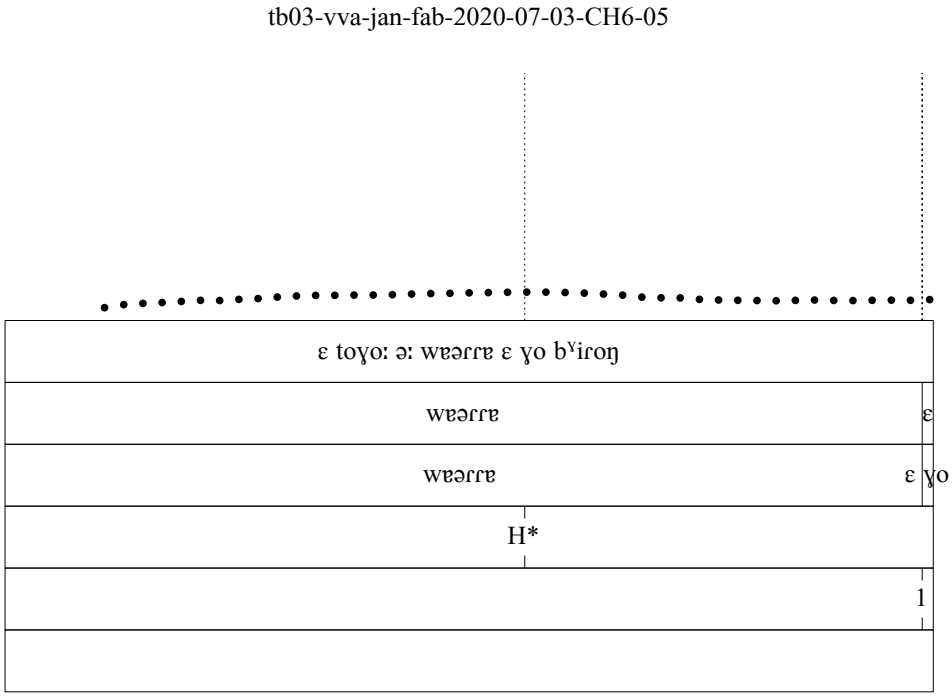


Figure 27. Spectrogram: [wɐəɾ-ɾɐ], airplane-DEM (tb03-vva-jan-fab-2020-07-03-CH6-05)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
wɐəɾ-ɾɐ, aeroplane-DEM	[ə]ɾ=74.49 dB		[əɾ]=135.2 Hz	H

Table 15. [wɐəɾ-ɾɐ], airplane-DEM (tb03-vva-jan-fab-2020-07-03-CH6-05)

When emphasis is needed to be expressed then the final syllable exhibits high pitch (Figures 28 and 29 and Table 45). For example, in Figure 28, the word *bʸi-roŋ*, although in final position, presents highest pitch on the final syllable because the speaker wished to add more information on the topic.

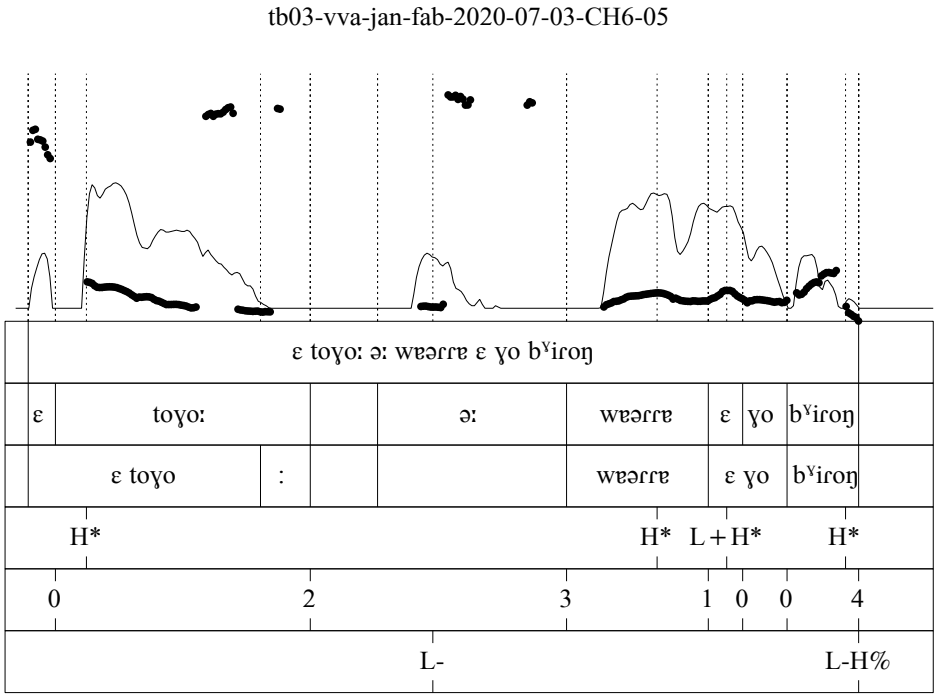


Figure 28. Spectrogram: There is no airplane that can come in (tb03-vva-jan-fab-2020-07-03-CH6-05)

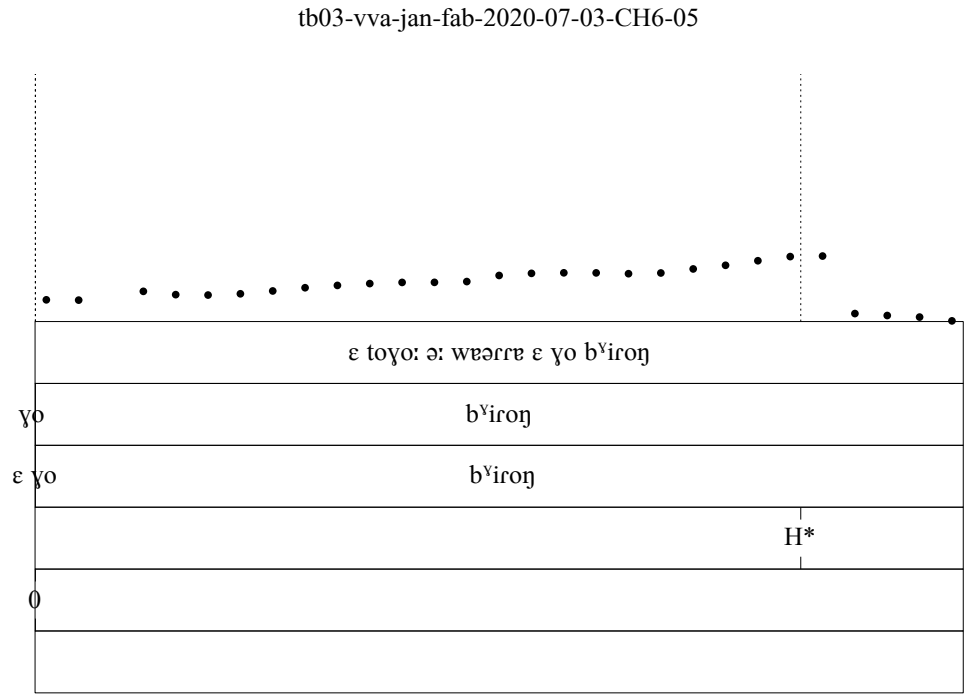


Figure 29. Spectrogram: [bʸi-ɾoŋ], go-in (tb03-vva-jan-fab-2020-07-03-CH6-05)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ɛ=toɣo, it=NEG	t[o]=76.71 dB	ɣ[o]=66.67 dB	[to]=158.5 Hz	H
əː, FILLER				
wɐər-ɾɐ, airplane-DEM	[ə]ɾ=74.49 dB		[əɾ]=135.2 Hz	H
ɛ=ɣo, it=DEM	[ɛ]=71.76 dB		[ɛ]=140.3	H
bʸi-ɾoŋ, go-in	[i]=61.28 dB		[ɾoŋ]=151.1 Hz	H

Table 16. Values: There is no airplane that can come in (tb03-vva-jan-fab-2020-07-03-CH6-05)

Additionally, in the indirect question "do chickens have names in America", the pitch rises in *ɛ=wow*, as in a surprise: "ARE THERE names for chickens in America" (Figure 25 and Table 17).

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ɛ=wow, it=there	w[o]w=78.72 dB	[ɛ]=76.93 dB	[wo]w=265.3 Hz	H*
itɛr, name	[ɛ]=76.67 dB	[i]=71.33 dB	[tɛ]ɾ=212.2. Hz	H
xɐjɐŋ, chicken	x[ɐ]=74.05 dB	j[ɐ]ŋ=73.92 dB	[jɐŋ]=192.4 Hz	H
ɣɐr mːɛri				
xɐjɐŋ, chicken	j[ɐ]=78.29 dB	x[ɐ]=77.21 dB	jɐ[ŋ]=210.9 Hz	H
ɣɐ-ɾ mɛrɛk, they-of America	ɣ[ɐ]=74.77 dB	m[ɛ]=73.13 Hz	[ɾɛk]=206 Hz	H

Table 17. Values: if chickens have names in America (tb01-vva-fab-scc-2020-06-20-CH6-04)

In unmarked questions, like, "well THEY, STAYING, how are you DOING, in Saipan" (Figure 30 and Table 18), the prosodic words with the highest pitch are disyllabic and within them the

highest pitch is on the penultimate syllable, rather than the final, contrasting with disyllabic words uttered in a statement, which demonstrate high pitch on the final syllable, or in the case of *ɣo=mor* (Table 18).

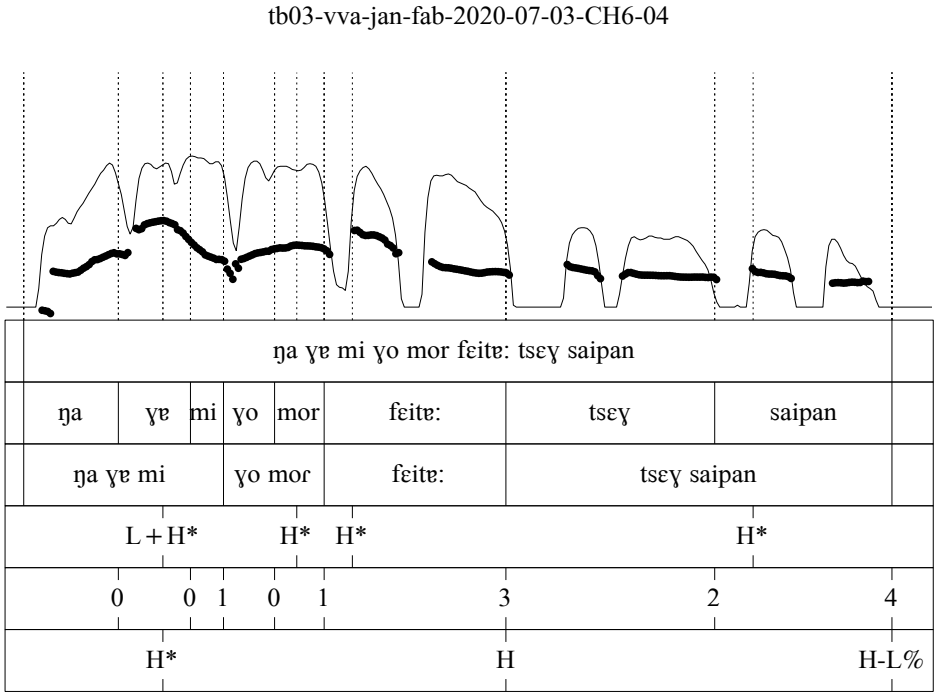


Figure 30. Spectrogram: How is it for you, who live in Saipan? (tb03-vva-jan-fab-2020-07-03-CH6-04)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ŋa, CONJ				
ɣɐ=mi, they=stay	m[i]= 82.12 dB		[ɣɐ]=286 Hz	H*
ɣo=mor, you=already	[ɣo]=89.4 dB		[mor]=234.2 Hz	H
fɛitɐ, do	[i]=79.35 dB	t[ɐ]=78.03 dB	[fɛi]=265.5 Hz	H*
tsɛɣ saipan, in Saipan	[i]=66.32 dB	p[a]=64.08 dB	[sai]=183.1 Hz	L

Table 18. Values: How is it for you, who live in Saipan? (tb03-vva-jan-fab-2020-07-03-CH6-04)

Finally, in 4 syllabic feet, the antepenultimate syllable exhibits highest pitch (Figures 31 and Table 19).

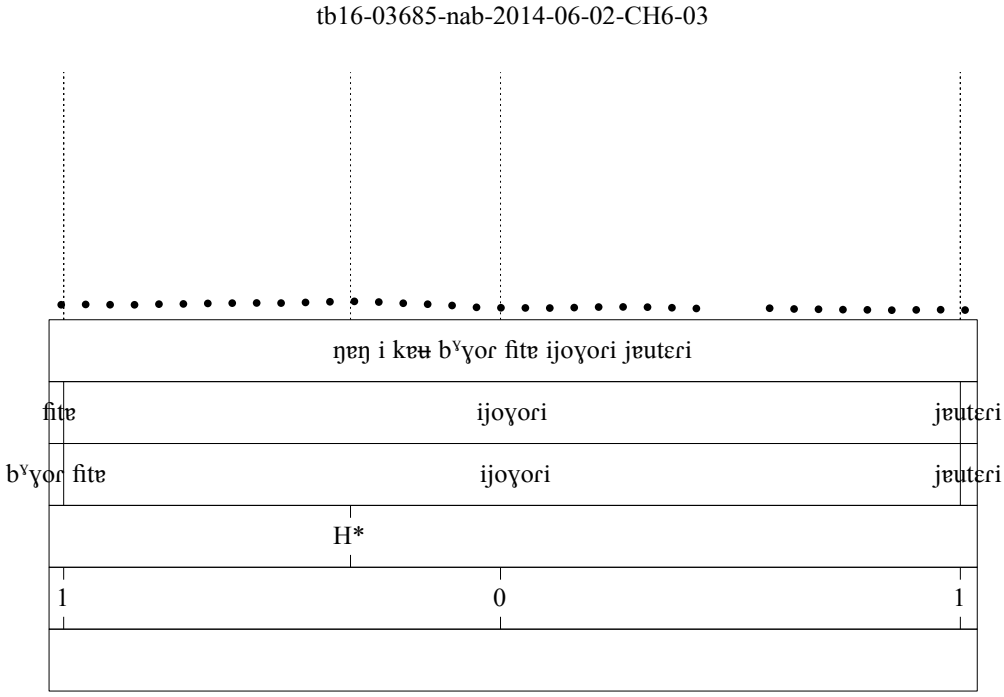


Figure 31. Spectrogram: [ijo-ɣo-ɾi], him-from-of (tb16-04088-nab-2014-06-02-CH6-03)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ŋɐŋ i=kɐʉ, I I=learn	[i]=86.36 dB		[ŋɐŋ]=144 Hz	H
bʸ-ɣoɾ fitɐ, go-out fish	f[i]=85.26 dB	bʸɣo[ɾ]=85.09 dB	[fi]=128.5 Hz	L
ijo-ɣo-ɾi, him-from-of	j[o]=84.44 dB		[jo]=113.7 Hz	L
jɐutɛɾi, elder	j[ɐ]=85.17 dB		[tɛ]=133.3 Hz	H

Table 19. Values: I learned how to fish from my father. (tb16-04088-nab-2014-06-02-CH6-03)

6.2.2. Intonation

Four phrase accents have been identified (Table 20). A high tone (H*), reaching 200 Hz for men and 280 Hz for women, a high mid tone (H), reaching 160 Hz for men and 240 Hz for women, a low mid tone (L), reaching 130 Hz for men and 200 Hz for women and a low tone (L*), with 110 Hz for men and 180 Hz for women.

Men	H*= 160-200 H=130-160 L=120-130 L*=80-110
Female	H*=240-280 H= 200-240 L=180-200 L*=150-180

Table 20. The phrase tones for men and women in Hz

6.2.2.1. Declarative intonation

Declarative utterances generally showcase a rise in pitch in the first phrase, and then a gradual fall in the final phrase. The intonation contour is H H L*, with H assigned on the penultimate syllable of the first foot (from left to right) (Figures 32 and Table 21).

tb16-03685-nab-2014-06-02-CH6-07

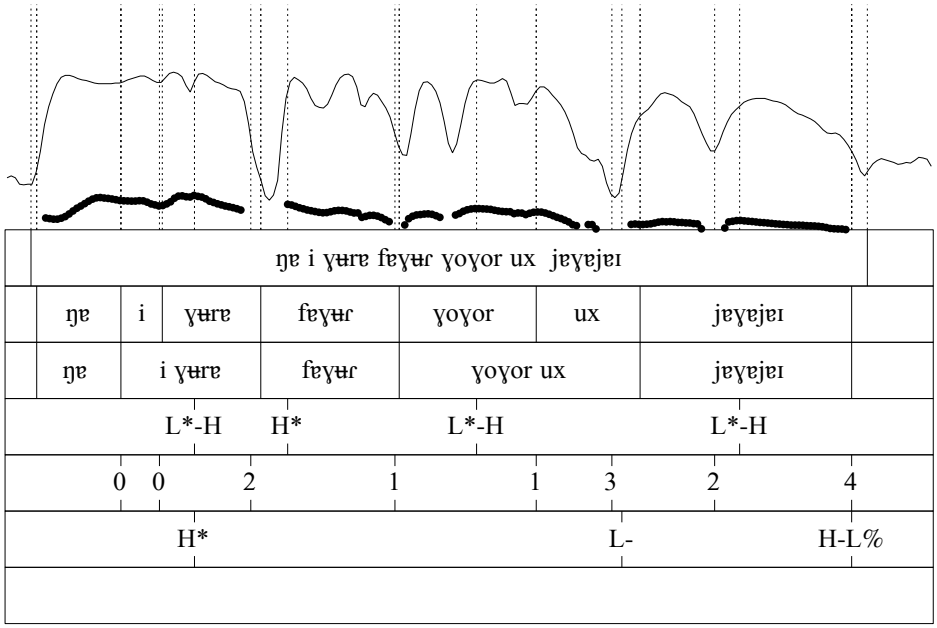


Figure 32. Spectrogram: I can go net fishing. (tb16-03685-nab-2014-06-02-CH6-07)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ŋɐ				
i=ɣʉrɐ, I=know	ɣ[ʉ]=85.46 dB		[ɣʉ]=154.8 Hz	H
fɐɣʉr, do	ɣ[ʉ]=85.09 dB		[fɐ]=134.3 Hz	H
ɣoɣor ux, use net	ɣ[o]r=84.05 dB	ɣ[o]=83.89 dB	[ɣor]=124.4 Hz	L

Table 21. Values: I can go net fishing. (tb16-03685-nab-2014-06-02-CH6-07)

Final phrases exhibit a deep descent in pitch particularly on the final syllable, as well as decreased loudness. Because of final vowel devoicing processes, final syllables are often virtually inaudible and can only be detected when the stress placement of the preceding syllable is taken into account (Figure 32).

6.2.2.2. Interrogative intonation

Interrogative sentences are marked in two ways: question words or yes or no questions. Yes/no questions presenting no morphological marking are usually marked by differences in their pitch contour in comparison to their declarative counterparts.

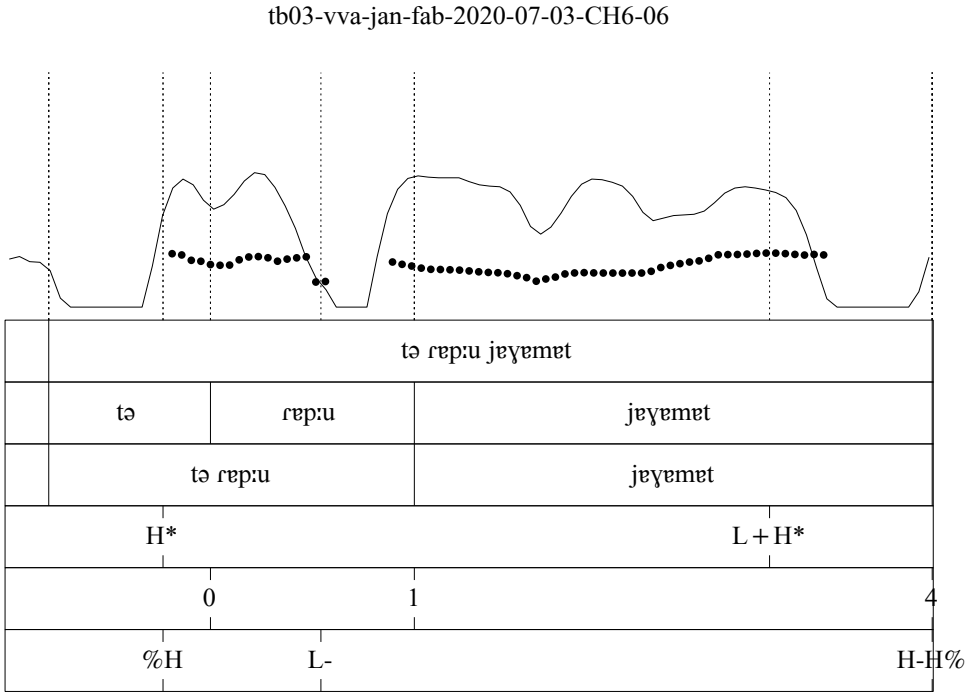


Figure 33. Spectrogram: Old person, isn't it? (tb03-vva-jan-fab-2020-07-03-CH6-06)

These unmarked interrogatives demonstrate a pitch rise on the final phrase contrary to the final fall of a declarative utterance. The intonation contour is H H, with the highest pitch assigned on the antepenultimate syllable of the first foot (from left to right) (Figure 33 and Table 22).

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
tə ɾɐpːu, NEG old	p[ː]u=78.3 dB	t[ə]=76.92 dB	[tə]=229.8 Hz	H
jɐɣɐmɐt aː, person FILLER	j[ɐ]=77.64 dB	ɣ[ɐ]=76.93 dB	[mɐt]=217.2 Hz	H

Table 22. Values: Old person, isn't it? (tb03-vva-jan-fab-2020-07-03-CH6-06)

Question word questions follow a declarative pattern of intonation (Figure 32). The onset, however, is higher in pitch than in a declarative utterance (Figure 32 vs 34 and Table 23). The intonation contour is H* H* L*, with the highest pitch assigned on the final syllable of the question word[6].

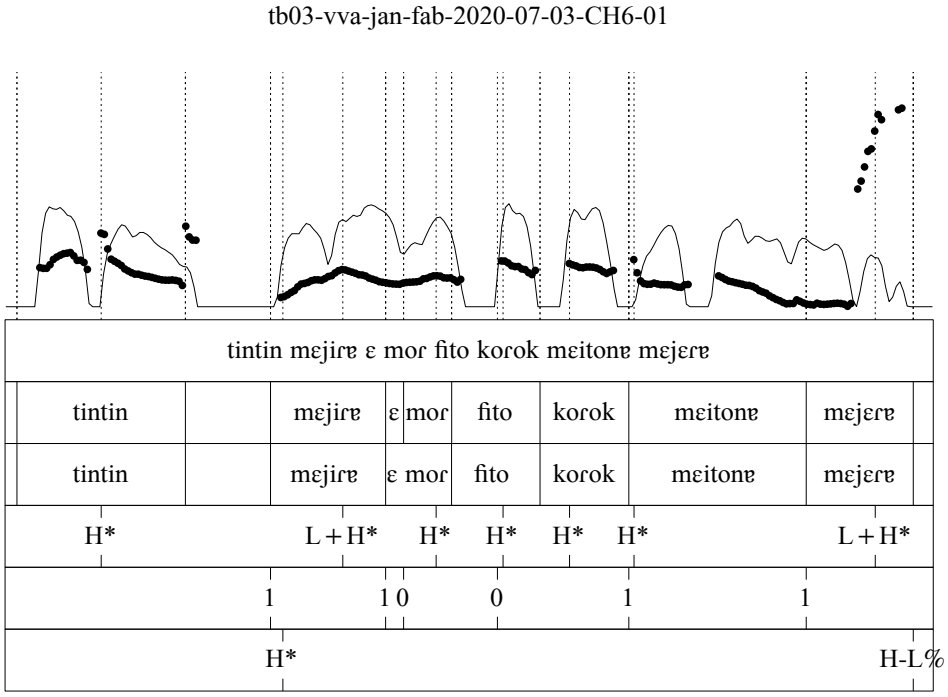


Figure 34. Spectrogram: Tintin what time is it over there now? (tb03-vva-jan-fab-2020-07-03-CH6-01)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
tintin, Tintin	t[i]n=71.1 dB		[tin]=259 Hz (second)	H*
mɛjiɾɐ, now	ɾ[ɐ]=70.44 dB	j[i]=69.28 dB	[ji]=181.2 Hz	H*
ɛ=moɾ, it=already	[ɛ]=69.5 dB		[mo]=168.3 Hz	H*

[6] The prosodic word *tintin* is not considered because it is a name.

fito, how many	t[o]=71.7 dB		[to]=199.7 Hz	H*
korok, o' clock	ɾ[o]k=70.78 dB		[ɾok]=193.9 Hz	H*
mɛ-itonɐ, with-over there	t[o]=68.59 dB	m[ɛ]=67.15	[mɛi]=198.3 Hz	H*
mɛjɛrɐ, now	m[ɛ]=63.19 dB		[mɛ]=109 Hz	L*

Table 23. Values: Tintin what time is it over there now? (tb03-vva-jan-fab-2020-07-03-CH6-01)

6.2.2.3. Greetings and positive opinion statements

Greetings present a rise on the second phrase but with a lower onset in pitch than a declarative and interrogative sentence (Figures 32, 34 vs 35). The pitch contour is L* L*, with the highest pitch on the penultimate syllable of the final foot (from left to right).

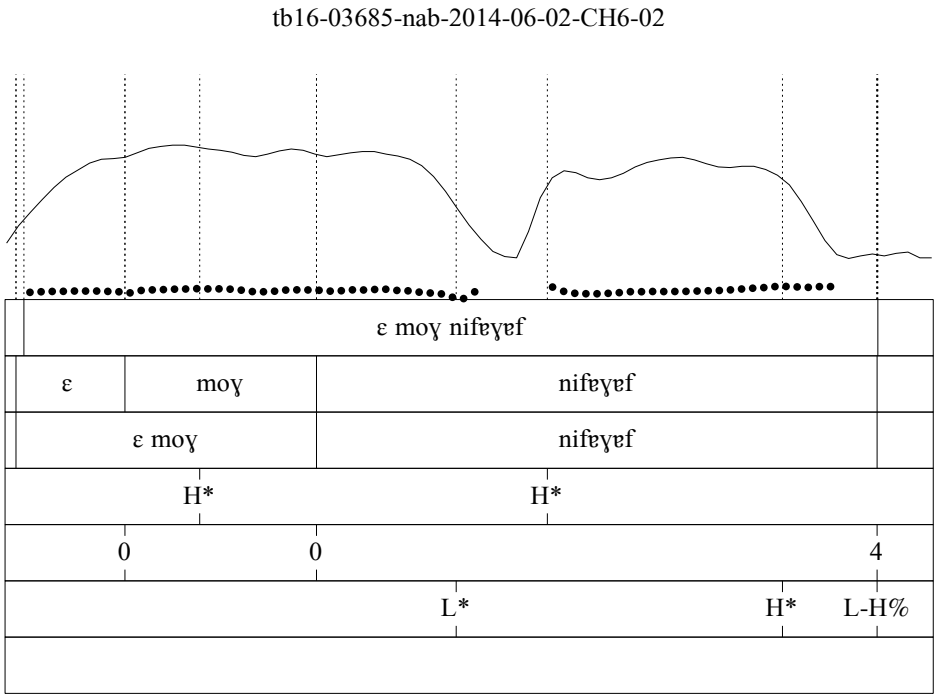


Figure 35. Spectrogram: Good evening. (tb16-03685-nab-2014-06-02-CH6-02)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent
ɛ=moɣ, it=good	m[o]ɣ=84.86 dB		[moɣ]=100.2 Hz	L*
ni-fɐɣɐf, at-evening	n[i]=83.45 dB	ɣ[ɐ]=82.19 dB	[fɐ]=104.2 Hz	L*

Table 24. Values: Good evening. (tb16-03685-nab-2014-06-02-CH6-02)

Positive opinion statements present a consistent high pitch, with the highest pitch on the final syllable of the final foot (from left to right). In this case (Figure 36 and Table 25), the L tone is because of *bʸɛ*, because.

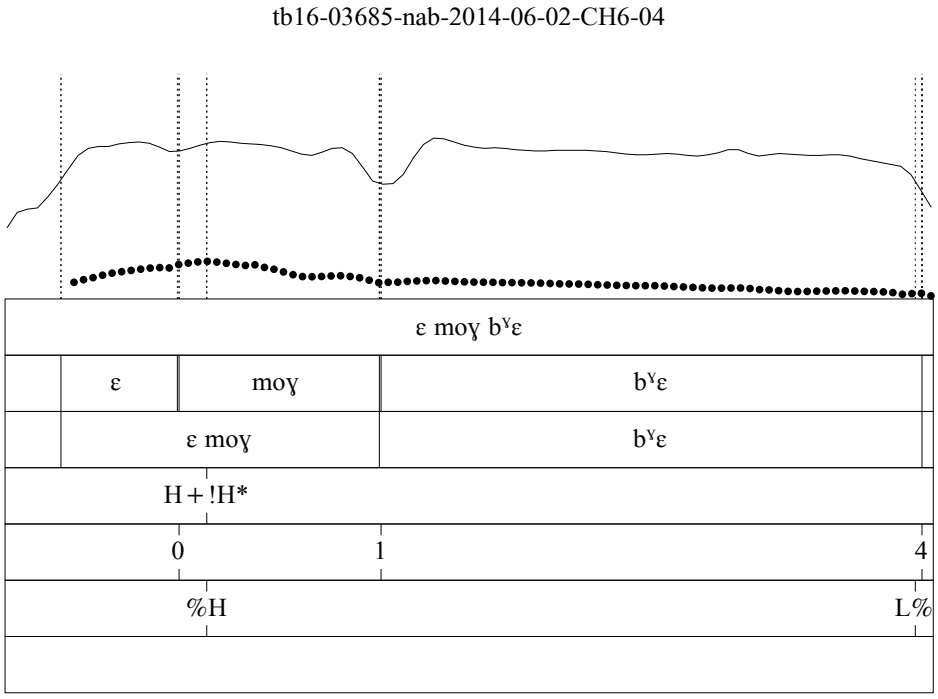


Figure 36. Spectrogram: It's good because… (tb16-03685-nab-2014-06-02-CH6-04)

Transcription and gloss	Primary stress	Secondary stress	Pitch	Phrase Accent

ɛ=moɣ, it=good	m[o]ɣ= 85.52 dB		[moɣ]= 162.2 Hz	H*

Table 25. Values: It's good because… (tb16-03685-nab-2014-06-02-CH6-04)

7. Discussion

7.1. Navigation

In terms of structure, the summary of the deposit includes information about the language, as well as the means for its documentation, through the collection of vocabulary, stories, poems and others (Figure 4). It is underlined that they have worked with over 50% of the adult population of the remaining Tobian speakers living in Palau at the time, thus demonstrating the credibility of their data (Black & Black, 2014). On a practical level, it is relatively easy to navigate through the collection and search for specific content, since the search column provides, as mentioned before, the option of searching by type, genre or participants. The summaries on the bundles and resources page provide more detailed information about the contents of each bundle, such as the location and time of the recording, as well as topics discussed (Figure 5). Thus, the deposit seems to be ethnographically driven, supported not only by its contents but also the depositors' academic interests. The collection is ideal for providing natural examples of language use with quality search mechanisms (Vinogradove, 2016: 137).

For example, Nixon Andrew's description of how he learned how to fish (Figure 9) includes not only ethnographic information, that is, Tobian fishing techniques, it also includes a monologue by Nixon containing fillers and pauses. Especially towards the end there is a mini interview with the curator that exhibits code-switching (8).

(8) tb16-04088-nab-2014-06-02-CH7-02[7]

ənli ɣɐɾ fitou^{w} tɐim ɛ b^{v}ɐ pːuwouwo

[7]View the transcriptions, translation, audio files and particular links to the collection here: https://docs.google.com/spreadsheets/d/1W0842Dt_W3HE1xSJAWLBzVAxRcnyJ0PunItDI6stLw4/edit?usp=sharing.

ənli ɣɐɾ fitouʷ tɐim ɛ-bʸɐ pːu-wou-wo

only only how.many times it-MOD go-with-out

pɐi

pɐ-i

dad-POSS.1SG

Only few times that I would go with my father.

Of special notice is the fact that when the curator asks a question in English, the speaker seems to switch to a mixture of English and Tobian, inserting English words, such as, *sometimes*, *only*, *time* and then he switches back to monolingual Tobian (9).

(9) tb16-04088-nab-2014-06-02-CH7-01

tɛi səmtaɪms aɪ

NEG sometimes I

tɛi pːipːi tɐim mʷɐ ibʸɐɣɐ bʸuwʉ ɣʉɣ

tɛi pːipːi tɐim mʷɐ i-bʸɐ-ɣɐ bʸu-wʉ ɣ-ʉɣ

NEG many times that I-MOD-EMP go-out sail-RDP

I don't sometimes I, I didn't go sailing that much.

Nevertheless, in order to navigate the particular deposit, it is necessary to have an above average level of English. Thus, in case the speakers were to be monolingual in Tobian, it would be hard to navigate through the collection. However, this is rarely the case with endangered languages, since it is usually a shift to a major language that leads to their endangerment (O'Shannessy, 2011: 78-79). In this case, because of the ecology of Tobian, speakers are probably trilingual in English and Palauan, the official languages of the Republic of Palau, and Tobian, or the emerging variety, Echangese (Black & Black, 2014).

7.2. Original and User-Generated Content

Austin (2010: 14) and Himmelmann (1998: 166) argue that language documentation is not about collecting raw data without analysing it. Analysis can make a documentation record accessible to multiple users, such as researchers of other fields and community members, particularly glossing and translating into languages of wider communication, along with the recording of metadata (Austin 2010: 18). Austin and Grenoble (2007: 22) also point out that linguistic analysis may shed light on speech genres, lexical forms, grammatical paradigms or sentence construction. They go as far as to underline that "without good analysis, recorded audio and video materials do not serve as data for any community of potential users" (Austin and Grenoble, 2007: 22).

Furthermore, "an understudied language often lacks generally accepted standards of description, such as a grammar or a dictionary, any kind of linguistic annotation is dependent on the subjectivity of the researcher" (Vinogradove, 2016: 130). Nevertheless, as Good (2011: 215) mentions, even the production of raw recordings presupposes certain choices, such as which video-recorder to use, which part of speech to record etc. Perhaps the biggest difference is that recording-techniques may belong to a variety of fields, in comparison to annotation which is a specifically linguistic activity, using linguistic conventions such as the IPA (*IPA Chart*, 2015).

For example, bundle 16 (Figure 9) includes a video, an ELAN (ELAN, 2020) file, a pdf with metadata and another one with an interlinearised transcription of the data. The transcription is in orthography, which facilitated the search of the vocabulary in the curators' dictionary and its annotation in linguistic terms, but this orthography is not standardised and there was no information in regards to the correspondence between letters and sounds. A disadvantage is that sometimes the terms in the dictionary are not accurately provided. For example, the word for to say is *ɛ=soɣ* (*ɛ=soɣ*, he/she=says) and to answer is *ɛsɐ soɣ* (*ɛ=sɐ soɣ*, he/she=TAM says). The particle *sɐ* seems to signify a TAM category unique to Tobian, with

the meaning of ‘as a result of the previous action’ (van den Berg, 2014: 25). This dissertation does not offer any specific advice on orthographic conventions, however, it does recommend considering van den Berg’s (2014: 45) recommendations.

Moreover, the two bundles (Figures 7 and 8) concerned with the story of clans and chickens were of particular interest. This story was narrated first in 2004 by Isauro Andrew, an older male speaker and in 2014 by Felicia Andrew, a younger female speaker. The content of the story describes a conversation between Peter and Santos and it is about a misunderstanding between the words for clan and chicken in Tobian. As mentioned by Seidel (2016: 28), from a literary perspective, it could be argued that these stories could be considered Tobian “literary” tradition and with each rendition of the text we have a new interpretation, or analysis of it.

> Like a manuscript that was preserved by chance and kept accessible for later generations (not for analysis under a specific theory but for its general interest and historical value) a language document should be recorded because of its potential to be analyzed at a later point in time. (Seidel, 2016: 42)

It could also be argued that accessing and analysing such manuscripts are both an observational and interpretative activity, especially for individuals who are not speakers or members of the community (Seidel, 2016: 39). In this case, this raw data (Himmelmann, 2006: 15) has led to primary data and finally to a description of the particular language and a better understanding of Micronesian prosody. For this reason, for the purposes of this study and for becoming accustomed with the language, transcriptions have been provided for all bundles included in this analysis. The ELAN (ELAN, 2020) files were then sent to the curators, who were encouraged to include them to their respective bundles, in the hope that future researchers would evaluate and improve them.

Finally, Conathan (2011: 489) notes that archives are sometimes sceptical about incorporating user-generated content. However, he underlines that such analyses, descriptions or additions could enhance the collection in creative ways, as long as they are additive and not replacing the original records. The depositor's role creates a different type of relationship with the collection, where they become mediators between active users and speaker communities (Garrett, 2014: 70).

7.3. Metadata

At the moment, metadata schemes focus on the research community, with speaker communities often searching for different types of information (Good, 2011: 231). That is, in this case, linguists would be interested to know the linguistic background of speakers, such as, which languages they use at home and which ones in their everyday lives, granted the multilingual context of Palau, while community members might be more interested in identifying their relatives. For instance, bundle 10 (Figure 10) contains descriptions of Rosania Victor and Stanley Magholahor's jobs.

Figure 37. Rosania Victor and Stanley Magholahor

This bundle contains a video recording, an ELAN (ELAN, 2020) file and a pdf with the interlinearised transcription of the file. The transcription is once again provided in the orthography proposed by the curators, while no translation added. The video quality is good, with the speakers' upper body visible to the audience. However, the interviewer is not visible on the screen and it would be more productive for them to be there too in order to avoid the observer's paradox and distract the speakers from the existence of a camera and a microphone (Figure 37).

The sociolinguistic data[8] in this video is interesting, with a lot of code-switching between Tobian, English and Palauan[9]:

(10) tb10-04115-rva-2013-11-05-CH7-02

ŋaː	i	fitɛɣ	bᵛɐ	ɾɣɐŋ	simɛɾ	ɹeɪndʒər	jɐ
ŋaː	i	fitɛɣ	bᵛɐ	ɾɣɐŋ	si-mɛɾ	ɹeɪndʒər	jɐ
I	I	work	CONJ	I	one-man	ranger	I
bᵛo	patɾol	tsʉ	ertʔ				
b-ᵛo	patrol	tsʉ	ertʔ				
go-out	patrol	on	boat				

I work as a ranger and I go out and patrol on the boat

Sometimes words are pronounced with a Palauan accent:

(11) tb10-04115-rva-2013-11-05-CH7-01

iː	i	fitɛɣ	ɣʉraː	i	fitɛɣiɐ	mesineri	a
iː	i	fitɛɣ	ɣʉraː	i	fitɛɣ-iɐ	mesineri	a
I	I	work	?	I	work-POSS	machinery	FILLER

boutʔ

[8] Find a sociolinguistic survey conducted by the curators in 2009 here: http://www.friendsoftobi.org/wordweek/surveynarrative022009.pdf

[9] tsʉ ertʔ is Palauan.

bout?

boat

I work in boat machinery.

The addition of sociolinguistic information would potentially have made it possible to understand the context and the language use of each speaker in more detail. For example, Stan code-switches extensively between Tobian and English. It was after consulting with Justin, who informed me about Stan's sociolinguistic background, that his choices were to some extent explained. Thus, the addition of more detailed metadata with sociolinguistic information would be suggested. For example, Conathan (2011: 237) argues that field notebooks are unique archival records because they contain biographical information about consultants, the location and date of work. This type of information is of high importance for linguists, particularly those researching "closely related varieties or the effects of social networks on language change" (Conathan, 2011: 248). However, the sensitivity of such data is recognized and it is understood that it is important to follow ethical practices and ensure their safety and privacy when required (Conathan, 2011: 248). Woodbury (2003: 29) also underlines the importance of metadata when studying dialect difference, social variation, genre and first language acquisition. Thus, information on the societal structure, activity and meaning is deemed significant and relevant.

Finally, as mentioned before, the idea of such deposits as collections to be curated (Woodbury, 2003: 29) enhances the need for detailed metadata and since language archives claim to assist and safeguard community knowledge, it could be argued that it is the researchers' responsibility to identify a strategy that would include participants in the archiving process (Garrett, 2014: 82). For example, it could be argued that in this particular context, by revisiting the collection, watching and listening to the recordings, providing comments and translations, the speakers participate in identifying the use of such data and in enhancing the collection.

7.4. Wider uses

As noted above, the curators underline that the collection is meant not only for speakers of Tobian, but also linguists (Black & Black, 2014). The content of the video and audio files could potentially be useful to anthropologists, focusing on comparative research, as well as ethnomusicologists and social scientists, since it includes songs and discussions on social practices. Particular bundles such as bundle 52 (Figure 11) and a new contribution from the sessions with Tintin and Justin in 2020 (Figure 12) are of particular interest because they contain recordings of naturally occurring conversations on social topics. It could be argued that these could become the most valuable sources of data for linguists in a variety of fields and especially for this study.

Bundle 52 (Figure 11) deals with a conversation between Jackie Victor and Paulina Theodore planning a community event. Culturally this recording exemplifies community practices and introduces vocabulary that could be useful for different audiences (future community members, linguists, anthropologists, even event planners), such as food for community events and arrangement of tasks, for example, that women bring the sides[10] and men bring the meat or fish (12), lists of honorary guests (13) and reasons for celebration (14). Linguistically, there is evidence of code-switching between Palauan, Tobian and English, but there is a possibility that it is only loan words. When talking to the interviewer there is some overlap and they interrupt each other.

(12) tb52-04241-jva-2014-05-28-CH7-04

ŋɛ	sɐ	ɣɐsomɐɾi	bᵛo	mɐɾ	tɛi	kɵtːo	tːɐɾij	mɛː	vri
ŋɛ	sɐ	ɣɐsomɐɾi	bᵛo	mɐɾ	tɛi	kɵtːo	tːɐɾij	mɛː	vri
CONJ	TAM	funny	CONJ	man	NEG	bring	meat	CONJ	woman

[10] Find the examples at: https://docs.google.com/spreadsheets/d/1W0842Dt_W3HE1xSJAWLBzVAxRcnyJ0PunItDI6stLw4/edit?usp=sharing.

kɐtoɣ mɐːŋ

kɐtoɣ mɐːŋ

bring eat

It's funny because men don't bring the meat but women bring the food

(13) tb52-04241-jva-2014-05-28-CH7-03

imʷɛɾɛ	tradɪsəna	liðer	mʷɛ	mʷɛɾɛɾɐ	deliket	mʷa	kavənoɪ
imʷɛɾɛ	tradɪsəna	liðer	mʷɛ	mʷɛɾɛ-ɾɐ	deliket	mʷa	kavənoɪ
this	traditional	leader	CONJ	this-DEM	delegate	CONJ	governor

We will have traditional leaders, the delegate and the governor

(14) tb52-04241-jva-2014-05-28-CH7-05

koʊ	wonoʊtsiwɐ	ɣɐ	tiwitu	mʷɐsiripːɛiːʲ	
koʊ	wonoʊ-tsi-wɐ	ɣɐ	tiwitu	mʷɐsiripːɛ-iːʲ	
these	kid-POSS.1PL.INCL-DEM	they	graduate	year-DEM	

mʷkɐ	jɛmɛntɨɾi	mʷɐ	ɣɐi	skuːl	mʷɐkɐ	tuːɾ	mʷɐ	pɪsisi
mʷkɐ	jɛmɛntɨɾi	mʷɐ	ɣɐi	skuːl	mʷɐkɐ	tuːɾ	mʷɐ	pɪsisi
these	elementar	CONJ	high	school	these	graduate	CONJ	PCC

Our kids who are graduating this year from elementary and from high school and from PCC (Palau Communit College)

The conversation between Tintin and Justin (Figure 16) is relevant to current events, with topics such as Covid-19 (test, quarantine, Guam-Palau-Saipan situations), school and dealing with the virus, face masks, and a prime example of a conversation between family members seeing each other after a while, with topics such as work, family, where they are, how they are doing, visiting Tobi island and retirement. Interesting observations include the fact that the recording shows how technology can be of help when in a pandemic but also when it is difficult to go to the field. Of course there are certain limitations but:

- Speakers get together to talk about things that interest them
- Not necessarily a constructed environment thus reducing the observer's paradox (the camera is already there instead of having a big microphone and a stranger in front of them)
- Important for finding information on their experiences in a pandemic which could be informative for future generations

In personal communication, both mentioned that they enjoyed the sessions because it encouraged them to think about Tobian and learn more about their own language. Tintin mentioned that it was nice to hear an elder speak the language. This type of Master-Apprentice context could potentially be useful for this particular community taking into account the high degree of migration and its cultural implications (Asang, 2000). Regarding Zoom (Zoom Video Communications Inc., 2016), the audio is not perfect and sometimes it is interrupted because of connectivity issues. What made annotating and transcribing difficult was the overlapping between the speakers as is natural in naturally occurring speech. On Zoom (Zoom Video Communications Inc., 2016), there is lack of immediate contact and body language, yet the speakers felt comfortable with the camera and the researcher.

7.5. Limitations and recommendations

In conclusion, this section reviews and evaluates the ELAR Tobian collection, deposited by Black and Black (2014). Overall, it is relatively easy to navigate through the deposit, since it is divided by genre and context, and a plethora of information is provided in each bundles' deposit. Nevertheless, it presents certain limitations. A recommendation would be a bilingual structure of deposits, with menus in both a major language of the world, as well as the endangered language in focus. This could potentially encourage community members, since they would witness their language used on an international website with an international audience. It could also be useful and productive for artists and social scientists to access and

identify sociolinguistic information in regard to particular bundles. Furthermore, it would be recommended for more information on the ELAN (ELAN, 2020) files to be included in the general description, since there are many possibilities for its use for language documentation and description. Without translations, analysis would not be possible, highlighting its importance especially when there are no other speakers.

Another limitation is that the collection includes recordings in a variety of qualities which may hinder phonological and phonetic analysis. For example, the recording from bundle 52 (Figure 11) was not used for the analysis of prosody. The audio quality was not good enough to extract accurate f0s, even though it contained natural language in a conversation which is ethnographically important. Furthermore, the recording from bundle 12 (Figure 7) contains a lot of background noise, while the speakers portrayed in Figure 37 were chewing betel nut making it difficult to distinguish the sounds.

As mentioned before, data from a language archive may encourage diachronic studies and identify points of language change (Hilpert, 2011: 436). Other benefits of the particular collection include the fact that it is ethnographically informed and documents not only the language but also its connection to the local culture (Figure 9). It contains naturally occuring conversations (Figure 11), although sometimes the beginning of recordings is a bit awkward, and on topics that are of personal interest to the speakers (Figure 10). A great advantage is the existence of both video and audio files, encouraging research on paralinguistic phenomena in different contexts and phonological research using spectrograms for the analysis of prosodic patterns, as in this case.

Finally, even though there seem to exist obstacles in terms of investigating linguistic structures for linguistic description and analysis, overall, it should be noted that this is a detailed and well-organised deposit of an endangered language with less than 150 speakers. The data included could be used to enhance the dictionary provided on the Friends of Tobi Island website

(*Tobian Language*, 2019), "support linguistic description and analysis, conduct historical and comparative research, provide input for language revitalization, and contribute to journalistic reporting and artistic media" (Conathan, 2011: 253).

8. Conclusion

This dissertation has applied raw data of Tobian, available on ELAR for the description of its prosodic patterns. As exemplified from the data and analyses, this dissertation argues that Tobian exemplifies a variety of moraic feet, with primary stress (intensity) dependent on the mora and pitch dependent on syllable structure. Due to time constraints, word limit and Covid-19 limitations, no investigation of secondary stress is presented, although it is hypothesized that it occurs on preceding moras in trimoraic feet as is characteristic of Micronesian languages (Kennedy, 2003: 35) and on final moras in disyllabic words, that is the same syllables that exhibit high pitch. As far as working with data from the particular collection is concerned, even though there are certain limitations, such as audio and video quality, lack of sociolinguistic metadata and translations, "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language" (Black and Black, 2014) is a detailed collection that could be used for various purposes and can be enhanced since the curators are open and active in working on it.

Future recommendations for research would be an ethnographic documentation of Sonsorolese and of the new variety, Echangese, in order to shed light on the processes of language change and potentially the origins of speculated harmony in the area. The data on the archive could be used for the creation of materials for the Tobian community, with the curators already having published children's stories and a pdf version of their online dictionary, available online (Friends of Tobi, 2020). For the linguistic community, the data have the potential of a diachronic study of the language in change, leading to the discovery of its exact

location within the family, while a sociolinguistic study could shed light on potential projects for the formalisation of the language in the multilingual Republic of Palau.

The particular research has provided a description of the prosodic patterns of Tobian and filled to some extent the gap in Micronesian prosody and in the Sonsorol to Truk dialect continuum. It has also contributed in better understanding the capabilities and limitations of a language archive and provided recommendations for improvement and further applications to future research. Furthermore, it has shed light on the role of archives when conducting research during a pandemic, as well as the implications of conducting remote fieldwork, which might exemplify the inequalities in access to technology but also how important it could be for diminishing the observer's paradox. Furthermore, since the analysed ELAN (ELAN, 2020) files and speaker sessions will be given to the curators, this research has enhanced the collection with an example of naturally occuring language, exemplifying what has become of our everyday life with technology, where catching up happens on face-time. Finally, the collection has been developed not only in terms of transcribing and glossing data for future generations of linguists but also of a recording of native speakers talking about their experiences during the Covid-19 pandemic in their respective countries for future Tobian generations.

9. References

Asang, I., (2000). Remaking Footprints: Palauan Migrants in Hawai'i. *The Contemporary Pacific*, vol. 12 (2). Honolulu: University of Hawai'i Press, pp. 371-384.

Austin, P. K. (2010). Current issues in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, vol 7. London: SOAS, pp. 12-33.

Austin, P. K. & Grenoble, L. (2007). Current trends in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description,* vol 4. London: SOAS, pp. 12-25.

Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., Pulsifer, P., Beaver, D. I., Chelliah, S., Dubinsky, S., Meier, R. P., Thieberger, N., Rice, K. and Woodbury, A. C. (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, vol. 56 (1), pp. 1–18. doi:10.1515/ling-2017-0032.

Berg, E. (2013). *Hatohobei and Sonsorol*. [audio]. Kaipuleohone: University of Hawaii at Manoa. Available at: https://scholarspace.manoa.hawaii.edu/handle/10125/32448 (Accessed 17 September 2020).

Black, P. W. and Black, B.W. (2014). *Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language*. London: SOAS, Endangered Languages Archive. Available at: http://elar.soas.ac.uk/deposit/0295 (Accessed 19 April 2020).

Blevins, J. and Harrison, S.P. (1999). Trimoraic Feet in Gilbertese. *Oceanic Linguistics*, vol. 38 (2), pp. 203-230.

Boersma, P. & Weenink, D. (2020). *Praat: doing phonetics by computer* [Computer program]. Version 6.1.16. Available at: http://www.praat.org/ (Accessed 6 June 2020).

Britain, D., and Matsumoto, K. (2015). Palauan English. In J. Williams, E. Schneider, P. Trudgill, & D. Schreier (Eds.), *Further Studies in the Lesser-Known Varieties of English*, *Studies in English Language.* Cambridge: Cambridge University Press, pp. 305-343.

doi:10.1017/CBO9781139108652.014.

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

Comrie, B., Haspelmath, M., & Bickel, B. (2008). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology & the Department of Linguistics of the University of Leipzig. Available at: https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf (Accessed 10 September 2020).

Conathan, L. (2011). Archiving and language documentation. In Austin, P. K., & Sallabank, J. (Eds.), *The cambridge handbook of endangered languages*. Available at: http://ebookcentral.proquest.com (Accessed 10 September 2020).

ELAN (Version 5.9) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Available at: https://archive.mpi.nl/tla/elan (Accessed 10 September 2020).

Endangered Languages Archive, SOAS University of London. Available at: https://www.soas.ac.uk/elar/about-elar/ (Accessed 20 April 2020).

Ellis, J. S. (2012). *Saipan Carolinian, one Chuukic language blended from many.* PhD dissertation, University of Hawai' i, Mānoa.

Garrett, E. (2014). Participant-driven language archiving. In David Nathan & Peter K. Austin (eds) *Language Documentation and Description*, vol 12: Special Issue on Language Documentation and Archiving. London: SOAS: pp. 68-84.

Good, J. (2011). Data and language documentation. In Austin, P. K., & Sallabank, J. (Eds.). *The cambridge handbook of endangered languages*. Available at: http://ebookcentral.proquest.com (Accessed 10 September 2020).

Grant, A. (2017). The Western Micronesian Sprachbund. In R. Hickey (ed.) *The Cambridge*

Handbook of Areal Linguistics. Cambridge: Cambridge University Press, pp. 852-877.

Henke, R.E. and Berez-Kroeker, A.L. (2016). A Brief History of Archiving in Language Documentation, with an Annotated Bibliography. In M.A. Shepard, G. Holton and R. Henke (eds) *Language Documentation and Conservation,* vol. 10, Series: Emergent Use and Conceptualization of Language Archives*,* pp. 411-457.

Hilpert, M. (2011). Dynamic visualizations of language change: motion charts on the basis of bivariate and multivariate data from diachronic corpora, *International Journal of Corpus Linguistics,* vol. 16 (4), pp. 435–61.

Himmelmann, N. P. (2006). Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation* (*Trends in Linguistics. Studies and Monographs*, 178). Berlin: Mouton de Gruyter, pp. 1-30.

Himmelmann, N. P. (1998). Documentary and descriptive linguistics, *Linguistics*, vol. 36 (1), pp. 161-196. doi: https://doi.org/10.1515/ling.1998.36.1.161

Holton, G. (2012). Language archives: They're not just for linguists any more. In F. Seifart, G. Haig, N. P. Himmelmann, D. Jung, A. Margetts, and P. Trilsbeek (eds.), *Language Documentation & Conservation Special Publication No. 3: Potentials of Language Documentation: Methods, Analyses, and Utilization*, pp. 105–110.

Hughes, K. (2020). *The Synchronic and Diachronic Phonology of Nauruan: Towards a Definitive Classification of an Understudied Micronesian Language*. CUNY Academic Works. Available at: https://academicworks.cuny.edu/gc_etds/3599 (Accessed 14 September 2020).

Huvila, I. (2008). Participatory archive: towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Arch Sci* 8, pp. 15–36. doi:10.1007/s10502-008-9071-0.

IPA Chart. (2015). International Phonetic Association. Available at: http://www.internationalphoneticassociation.org/content/ipa-chart, under a Creative Commons Attribution-Sharealike 3.0 Unported License.

Jackson, F. (1983). *The internal and external relationships of the Trukic languages of Micronesia*. PhD dissertation, University of Hawai'i, Mānoa.

Jurafsky, D., Shriberg, E., and Biasca, D. (1997). Switchboard-DAMS Labeling Project Coders Manual. *Technical Report* 97-02, University of Colorado Institute of Cognitive Science.

Kennedy, R. (2018). *Phonology: A Coursebook*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107110793.

Kennedy, R. (2003). *Confluence in Phonology: Evidence from Micronesian Reduplication.* Phd Dissertation. Department of Linguistics, The University of Arizona.

Ladefoged, P. (2003). Phonetic Fieldwork. In M. J. Solé, D. Recasens, and J. Romero (eds.), *15th International Congress of Phonetic Sciences*, Barcelona, Spain, August 3-9, ISBN 1-876346-48-5; ICPhS Archive. Available at: http://www.internationalphoneticassociation.org/icphs/icphs2003 (Accessed 10 September 2020).

Matsumoto, K. and Britain, D. (2000). *Hegemonic Diglossia and Pickled Radish: Symbolic Domination and Resistance in the Trilingual Republic of Palau.* Available at: https://www.researchgate.net/publication/268390875 (Accessed 14 August 2019).

Matsumoto, K. and Britain, D. (2006). Palau: Language Situation. In Brown, K. (ed.), *Encyclopedia of Language and Linguistics*, vol. 9, 2nd ed., Elsevier, pp.129-137.

Mosel, U. (2006). Grammaticography: The art and craft of writing grammars. In Ameka, Felix, Alan Dench and Nicholas Evans (eds.). *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter.

O'Shannessy, C. (2011). Language contact and change in endangered languages. In P. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press, pp. 78-99. doi:10.1017/CBO9780511975981.005

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* vol. 90, pp. 2956-2970.

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

Regh, K. (1993). Proto-Micronesian Prosody. *Oceanic Linguistics Publications, Tonality in Austronesian Languages*, 24. Honolulu: University of Hawai'i Press, pp. 25-46.

Research Ethics at SOAS. (2019). SOAS University of London. Available at: https://www.soas.ac.uk/research/ethics/ (Accessed 17 September 2020).

Seidel, F. (2016). Documentary linguistics: A language philology of the 21st century. In Peter K. Austin (ed.) *Language Documentation and Description*, vol 13. London: SOAS, pp.23-63.

Silverman, K., E., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C., Price, P., Hirschberg, J. (1992). *ToBI: A Standard Scheme for Labeling Prosody. International Conference Spoken Language Processing*. Banff, Canada: 867–869.

Sohn, H. and Bender, B.W. (1973). A Ulithian Grammar. *Pacific Linguistics Series* C, no. 27. The Australian National University: Department of Linguistics Research School of Pacific Studies.

Taborosi, D., Ierago, L., Aquino, J., Pedro, L., Theodore, I., Tirso, S. and Tirso, P. (2018). *Sonsorolese Language Ramari Dongosaro*. Sonsorol Island. Available at: http://www.sonsorol.com/language/ (Accessed 29 December 2019).

Take-down Policy. (2018). Endangered Languages Archive. Available at: https://www.soas.ac.uk/elar/depositing-with-elar/our-policies/file131941.pdf (Accessed 20 April 2020).

Tobian Language. (2019). Friends of Tobi. Available at: http://www.friendsoftobi.org/tobilanguage.htm (Accessed 20 April 2020).

Vinogradov, I. (2016). Linguistic corpora of understudied languages: do they make sense? *Corpus de lenguas poco estudiadas: ¿Tiene sentido?.* Káñina, Rev. Artes y Letras, Univ. Costa Rica XL (1): 127-141, 2016 / EISSN:2215-2636.

uh (2020). Friends of Tobi. Available at: http://www.friendsoftobi.org/tobilanguage.htm (Accessed 7 September 2020).

van den Berg, E. (2014). *Linguistic Sketch of Hatohobeian and Sonsorolese: a study of phonology and morphology.* B.A. Thesis. Vrije Universiteit Amsterdam, VU Amsterdam.

Woodbury, A.C. (2003). Defining documentary linguistics. In Peter K. Austin (ed.) *Language Documentation and Description*, vol. 1. London: SOAS, pp. 35-51.

Woodbury, A.C. (2014). Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. In D. Nathan & P.K. Austin (eds) *Language Documentation and Description,* vol. 12: Special Issue on Language Documentation and Archiving. London: SOAS, pp. 19-36.

Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France. 7-10.

Zoom Video Communications Inc . (2016). *Security guide*. Zoom Video Communications Inc. Available at: https://d24cgw3uvb9a9h.cloudfront.net/static/81625/doc/Zoom-Security-White-Paper.pdf (Accessed 10 September 2020).

10. Appendix

10.1. Consent Forms

Consent Form for MA Dissertation in Language Documentation and Description

Please complete this form after you have read the Abstract and listened to an equivalent explanation about the research.

Project Title: Prosodic Patterns in Ramari Hatohobei

Abstract: With only one grammar describing the languages of Sonsorol-Tobi and only its phonetics, this dissertation will focus on describing prosodic patterns in Ramari Hatohobei, or Tobian, a severely endangered Micronesian language. The primary aim is to contribute to the description of Ramari Hatohobei based on data from the ELAR collection, "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language". Another aim is to identify the extent to which such data could be useful for linguistic description and in particular to the field of phonology and phonetics. The theoretical framework proposed is a combination of linear, functional and acoustic models of phonemic analysis using spectrograms extracted from Praat from conversations, descriptions and stories. Consultations with the curators and speakers in order to investigate particular hypotheses are attempted. Particular tasks include assisting with the translation of recordings from the ELAR collection, narrating stories in Ramari Hatohobei and providing relevant comments on the data and analyses presented. Self-reflective research will also shed light on the benefits, disadvantages and difficulties of working with data from a language archive, through the collection of diary entries. Due to my personal interest in documenting Sonsorolese, a closely related language, this dissertation could potentially become an axis in distinguishing the different prosodic patterns between the two languages.

Researcher Name: Vasiliki Vita

	Please tick the appropriate boxes	**Yes**	**No**

	I agree to take part in the project and understand that taking part involves answering a questionnaire with personal information, assisting with the translation of recordings from the ELAR collection, "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language", narrating stories in Ramari Hatohobei and providing relevant comments on the data and analyses presented to me.	X	
	I agree that my interviews are recorded using audio and video.	X	
	I understand that I can refuse to answer questions.	X	
	I understand that my taking part is voluntary; I can withdraw from the study at any time by notifying the researcher/s involved and I do not have to give any reasons for why I no longer want to take part.	X	
	I understand that my withdrawal or refusal to take part will not affect my relationship with Vasiliki Vita and SOAS, University of London.	X	
	I understand that that personal information collected about me that can identify me, such as my name or where I live, might be shared beyond the scope of this particular research.	X	

	I understand information I provide will be stored securely by Vasiliki Vita and that I will have access to it.	X	
	I understand that the information I provide will be used for research purposes.	X	
	I would like to be named in publications, reports, web pages, and other research outputs.		X
	I agree to waive copyright and other intellectual property rights in the material I contribute to the project.	X	

Contact Information

Telephone No: +30 6988257124

Email Address: 675802@soas.ac.uk / vasilikivita9@gmail.com

Postal Address: Skepasmeno 7, Velventos, Kozani, 50400, Greece

Alternative contact:

Lutz Marten

+44 (0)20 7898 4653

Room 426 Russell Square: College Buildings

SOAS University of London

Thornhaugh Street, Russell Square, London WC1H 0XG

Research Participant Declaration

Felicia Andrew	Felicia Andrew	June 19, 2020
Name of Participant [printed]	Signature	Date

I have accurately read out the information sheet to the potential participant and to the best of my ability, ensured that that participant understands what they are freely consenting.

Vasiliki Vita	Vita	June 19, 2020
Name of Researcher [printed]	Signature	Date

Please ensure a copy of this document is retained safely for future reference.

Consent Form for MA Dissertation in Language Documentation and Description

Please complete this form after you have read the Abstract and listened to an equivalent explanation about the research.

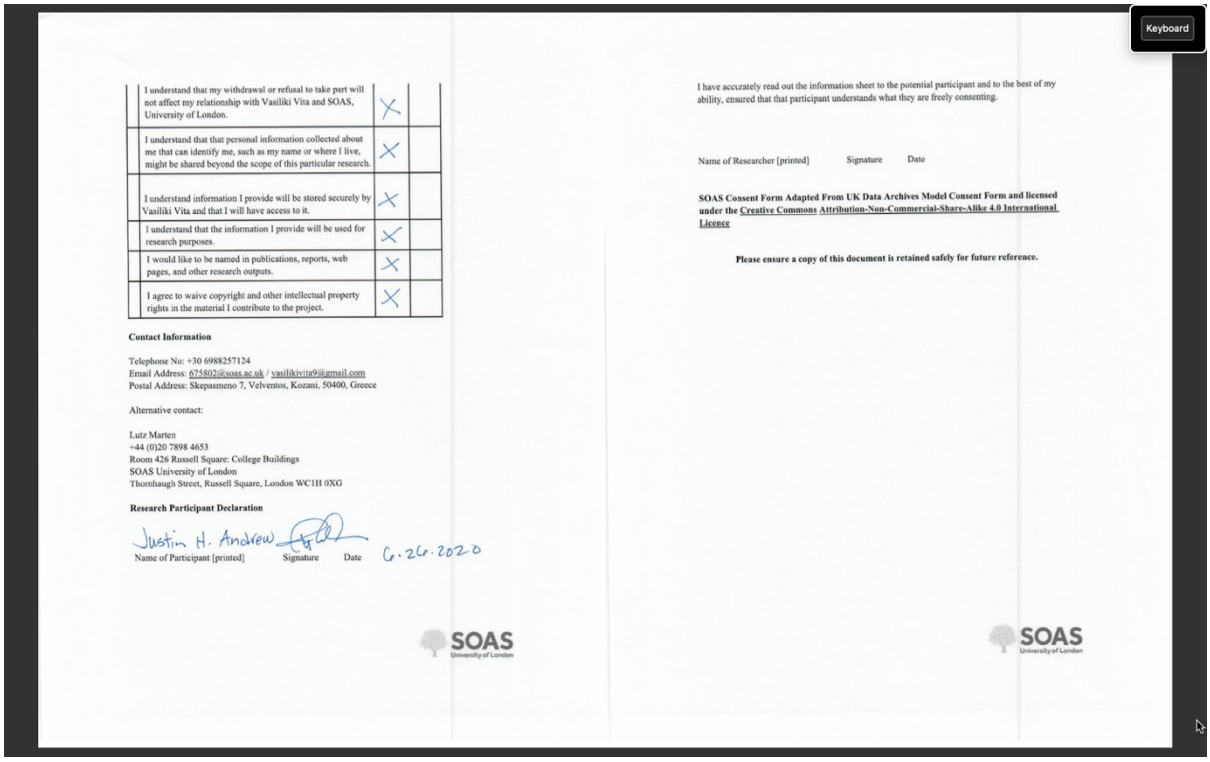
Project Title: Prosodic Patterns in Ramari Hatohobei

Abstract: With only one grammar describing the languages of Sonsorol-Tobi and only its phonetics, this dissertation will focus on describing prosodic patterns in Ramari Hatohobei, or Tobian, a severely endangered Micronesian language. The primary aim is to contribute to the description of Ramari Hatohobei based on data from the ELAR collection, "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language". Another aim is to identify the extent to which such data could be useful for linguistic description and in particular to the field of phonology and phonetics. The theoretical framework proposed is a combination of linear, functional and acoustic models of phonemic analysis using spectrograms extracted from Praat from conversations, descriptions and stories. Consultations with the curators and speakers in order to investigate particular hypotheses are attempted. Particular tasks include assisting with the translation of recordings from the ELAR collection, narrating stories in Ramari Hatohobei and providing relevant comments on the data and analyses presented. Self-reflective research will also shed light on the benefits, disadvantages and difficulties of working with data from a language archive, through the collection of diary entries. Due to my personal interest in documenting Sonsorolese, a closely related language, this dissertation could potentially become an axis in distinguishing the different prosodic patterns between the two languages.

Researcher Name: Vasiliki Vita

	Please tick the appropriate boxes	**Yes**	**No**
	I agree to take part in the project and understand that taking part involves answering a questionnaire with personal information, assisting with the translation of recordings from the ELAR collection, "Documenting Ramari Hatohobei, the Tobian language, a severely endangered Micronesian language", narrating stories in Ramari Hatohobei and providing relevant comments on the data and analyses presented to me.	X	
	I agree that my interviews are recorded using audio and video.	X	
	I understand that I can refuse to answer questions.	X	
	I understand that my taking part is voluntary; I can withdraw from the study at any time by notifying the researcher/s involved and I do not have to give any reasons for why I no longer want to take part.	X	

SOAS
University of London



I understand that my withdrawal or refusal to take part will not affect my relationship with Vasiliki Vita and SOAS, University of London.	X	
I understand that that personal information collected about me that can identify me, such as my name or where I live, might be shared beyond the scope of this particular research.	X	
I understand information I provide will be stored securely by Vasiliki Vita and that I will have access to it.	X	
I understand that the information I provide will be used for research purposes.	X	
I would like to be named in publications, reports, web pages, and other research outputs.	X	
I agree to waive copyright and other intellectual property rights in the material I contribute to the project.	X	

Contact Information

Telephone No: +30 6988257124
Email Address: 675802@soas.ac.uk / vasilikivita9@gmail.com
Postal Address: Skepasmeno 7, Velventos, Kozani, 50400, Greece

Alternative contact:

Lutz Marten
+44 (0)20 7898 4653
Room 426 Russell Square: College Buildings
SOAS University of London
Thornhaugh Street, Russell Square, London WC1H 0XG

Research Participant Declaration

Justin H. Andrew — 6.26.2020
Name of Participant [printed] Signature Date

SOAS University of London

I have accurately read out the information sheet to the potential participant and to the best of my ability, ensured that that participant understands what they are freely consenting.

Name of Researcher [printed] Signature Date

SOAS Consent Form Adapted From UK Data Archives Model Consent Form and licensed under the Creative Commons Attribution-Non-Commercial-Share-Alike 4.0 International Licence

Please ensure a copy of this document is retained safely for future reference.

SOAS University of London

10.2. R Codes

```
my_data <- For_STATS

my_data <- read_xlsx("~/Desktop/For_STATS.xlsx")

ggplot(my_data, aes(x = Moras)) +

  geom_bar() + ylab("Amount of words")
```

The For_STATS.xlsx file is based on data found here:

https://docs.google.com/spreadsheets/d/1XYiYJJFsEh353-U9-h8FlG_5Ch0-cwyuIHwTKgXlOoU/edit?usp=sharing